Deliverable 1.3 Updated

Panorama of ICT landscape in EU and US: ICT, Policies, regulations, programmes and networks in the EU and US with Updates for Latest US Announcements on 5G Initiative

Authors: Haydn Thompson, *THHINK – UK*
Daniela Ramos-Hernandez, *THHINK - UK*

ICT Policy, Research and Innovation
for a Smart Society

September 2016

www.picasso-project.eu

PICASSO has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 687874.

Project Deliverable

Project Number:	Project Acronym:	Project Title:
687874	PICASSO	ICT Policy, Research and Innovation for a Smart Society: towards new avenues in EU-US ICT collaboration

Instrument:	Thematic Priority:
COORDINATION AND SUPPORT ACTION	H2020 ICT

Title:
D1.3 Panorama of ICT landscape in EU and US: ICT, Policies, regulations, programmes and networks in the EU and US with Updates for Latest US Announcements on 5G Initiative

Contractual Delivery Date:	Actual Delivery Date:
May, 31st 2016	May, 27th 2016, Update September 29th 2016

Start date of project:	Duration:
January 1st, 2016	30 months

Organization name of lead contractor for this deliverable:	Document version:
THHINK	V1.4

Dissemination level (Project co-funded by the European Commission within the Horizon 2020 Programme)		
PU	Public	X
PP	Restricted to other programme participants (including the Commission)	
RE	Restricted to a group defined by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission)	

Authors (Organizations):

Haydn Thompson (THHINK)

Daniela Ramos-Hernandez (THHINK)

Reviewers (Organization):

5G Expert Group, CPS/IoT Expert Group, Big Data Expert Group, Policy Expert Group and NIST.

Abstract: This report provides a panorama of ICT policies, regulations, programmes and networks in smart cities, smart transportation and smart energy and also an overview of industry-driven programs, priorities, networks, major projects in EU and US in the areas of 5G Networks, Big Data, Internet of Things and Cyber-Physical Systems. Additionally, these areas are also actively being pursued at the world level and relevant major programmes around the world have also been identified. Key work with respect to regulations and standards is highlighted for each of the domains. It is clear that there are many opportunities for joint collaborations between the EU and US. It would be possible to collaborate on research and policy, on regulations and on standards.

There are key opportunities in the areas of smart cities and IoT/CPS which are rapidly developing areas and where there are common research, regulatory and standardisation needs. There are also great opportunities in the areas of smart energy and smart transportation, however, here there is existing regulation and legislation which needs to be harmonised. For the underlying technologies, which are the basic building blocks of future applications: 5G, Big Data and IoT/CPS, there are many opportuntiies to work together which would allow bilateral access to EU and US markets and would allow technology and products to be sold on the world stage increasing the competitiveness of EU and US companies in existing and developing markets.

This report is an update to the original report from May 2016 as the US Government announced a major new initiative on 5G in July 2016.

Keywords: Smart Cities, Smart Energy, Smart Transportation, 5G, Big Data, IoT, CPS

Disclaimer

Change History

Haydn Thompson 12/5/2016	Draft Circulated
Haydn Thompson 24/5/2016	V1.0 Updates to include 5G Expert Group and IoT/CPS Expert Group Comments
Haydn Thompson 25/5/2016	V1.1 Suggested collaboration opportunities tables for internal use removed from published deliverable
Haydn Thompson 27/5/2016	V1.2 Submitted to Commission
Haydn Thompson 31/5/2016	V1.3 Addition of information on document update plan and contact details for input
Haydn Thompson 29/9/2016	V1.4 Addition of information on 5G PPP projects and update on US 5G Initiative announced in July 2016. Also addition of project lists supplied by PICASSO Expert Groups in Appendices.

Table of Contents

List of Figures

Acronyms and Definitions

Acronyms	Defined as
ACER	Agency for the Cooperation of Energy Regulators
ADEME	Environment and Energy Management Agency
ADS-B	Automatic Dependent Surveillance - Broadcast
AID	Automatic Incident Detection
AIOTI	Alliance for the Internet of Things
AMI	Advanced Metering Infrastructure
ANPR	Automatic Number-Plate Recognition
ANR	National Research Agency
ANSI	American National Standards Institute
APAC	Asia Pacific
APEC	Asia Pacific Economic Cooperation
ARES	Aerial Reconfigurable Embedded System
ARIB	Association of Radio Industries and Businesses
ARMD	Aeronautics Research Mission Directorate
ARPA-E	Advanced Research Projects Agency–Energy
ARRA	American Recovery and Reinvestment Act
ARTEMIS-IA	Advanced Research and Technology for Embedded Intelligence and Systems Industry Association
ASC	Amsterdam Smart City
ASTRAEA	Autonomous Systems Technology Related Airborne Evaluation & Assessment

ATP	Automatic Train Protection system
AVI/AVL	Automatic Vehicle Identification and Location
BMBF	Federal Ministry of Education and Research (Germany)
BND	Federal Intelligence Service (Germany)
BSI	British Standards Institute
BUB	Bottom-UP Broadband
BuNGee	Beyond Next Generation project
BUTLER	uBiquitous, secUre inTernet-of-things with Location and contEx-awaReness project
CAA	Civil Aviation Authority
CAGR	Compound Annual Growth Rate
CAM	City Analysis Methodology
CCIT	California Center for Innovative Transportation
CEDS	Cyber-security for Energy Delivery Systems
CEER	Council of European Energy Regulators
CEF	Connecting Europe Facility
CEN	European Committee for Standardization (Comité Européen de Normalisation)
CENELEC	European Committee for Electrotechnical Standardization
CFM	Community Foundry Model
CIP PSP	Competitiveness and Innovation framework Programme, Policy Support Programme
CISE	Common Information Sharing Environment
CNAP	Cyber-security National Action Plan
C-ITS	Cooperative Intelligent Transport Systems
CML	Compass Modelling Language
COFRET	Carbon Footprint of Freight Transport
COMPASS	Comprehensive Modelling for Advanced System of Systems
CPS	Cyber-Physical Systems
CPSoS	Cyber-Physical Systems of Systems
CPS PWG	Cyber-Physical Systems Public Working Group
C-RAN	Cloud Radio Access Network
CRE	National Energy Regulator (France)
CRE	Comisión Reguladora de Energia (Mexico)
CRM	Citizen Relations Management
CROWD	Connectivity management for eneRgy Optimised Wireless Dense networks
CRYSTAL	CRitical sYSTem engineering AcceLeration project
CSA	Coordination and Support Actions
CSAAC	Cyber-Security Advanced Analytics Cloud
CSIA	Cyber-Security and Information Assurance
C2C	Car to Car
C2C-CC	CAR 2 CAR Communication Consortium
C2I	Car to Infrastructure
DANSE	Designing for Adaptation and evolutionN in System of systems Engineering project
DARPA	Defense Advanced Research Projects Agency
DCC	Dublin City Council
DCE	Digital City Exchange
DECC	Department of Energy and Climate Change
DHS	Department of Homeland Security
DISA	Defense Information Systems Agency
DITMAC	Defense Insider Threat Management and Analysis Center
DLR	German Aerospace Centre
DoD	Department of Defence

DOE	Department of Energy
DOT	Department of Transportation
DSM	Digital Single Market
EASA	European Aviation Safety Agency
EISA	Energy Independence and Security Act
EC	European Commission
ECHR	European Convention on Human Rights
ECSEL	Electronic Components and Systems for European Leadership
ECSEL-JU	Electronic Components and Systems for European Leadership programme
EEA	European Economic Area
EERA	European Energy Research Alliance
EFTA	European Free Trade Association
EIP	European Innovation Partnership
EIP – SCC	European Innovation Partnership on Smart Cities and Communities
EIT	European Institute of Innovation and Technology
EMC^2	Embedded Multi-Core systems for Mixed Criticality applications in dynamic and changeable real-time environments project
EMSA	European Maritime Safety Agency
ENSG	Electricity Networks Strategy Group
EPSRC	Engineering and Physical Sciences Research Council
ERA	European Railway Agency
ERTMS	European Railway Traffic Management System
ESOs	European Standards Organizations
ETCS	European Train Control System
ETI	Energy Technology Institute
ETNA	European Transport Network Alliance
ETP	European Technology Platforms
ETSI	European Telecommunications Standards Institute
EU	European Union
EV	Electric Vehicle
EV4SCC	Electric Vehicle for Smart Cities and Communities
FAA	Federal Aviation Administration
FCC	Federal Communications Commission
FERC	Federal Energy Regulatory Commission
FHWA	Federal Highway Administration
FP7	7^{th} Framework Programme
FTC	Federal Trade Commission
GIS	Geographic Information System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSM	Global System for Mobile
HEFCE	Higher Education Funding Council for England
HetNets	Heterogeneous Networks
HGV	Heavy Goods Vehicle
HIPAA	Health Information Portability and Accountability Act
HNDU	Heat Networks Delivery Unit
HPC	High Performance Computing
ICT	Information and communications technology
IEC	International Electrotechnical Commission
IEDs	Improvised Explosive Devices
IEEE	Institute of Electrical and Electronics Engineers
IERC	IoT European Research Cluster
IGERT	Integrative Graduate Education and Research Traineeship

IIC	Industrial Internet Consortium
IMO	International Maritime Organization
IOS	Interoperability Specification
IoT	Internet of Things and
IoTSF	Internet of Things Security Foundation
IREEN	ICT Roadmap for Energy Efficient Neighbourhoods
ISGAN	International Smart Grid Action Network
ISO	International Organization for Standardization
ISTAR	Intelligence, Surveillance, Target Acquisition and Reconnaissance
IT	Information Technology
ITA	International Trade Administration
ITDP	Institute for Transportation and Development Policy
ITL	Information Technology Laboratory
ITS	Intelligent Transport Systems
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
I4MS	Innovation for Manufacturing SMEs Initiative
JMS	Joint Management System
JPO	Joint Program Office
JTC	Joint Technical Committee
KIC	Knowledge and Innovation Community
LA	Latin America
LCICG	Low Carbon Innovation Coordination Group
LCNF	Low Carbon Networks Fund
LeGSB	Local eGovernment Standards Body
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
MEA	Middle East and Africa
MIF	Maritime Industries Forum
MIIT	Ministry of Industry and Information Technology
MIMO	Multiple Input, Multiple Output
MOST	Ministry of Science and Technology
MOU	Memorandum of Understanding
MSIP	Ministry of Science, ICT (information and communications technology) and Future Planning
M2M	Machine to Machine
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
NATO	North Atlantic Treaty Organization
NCATS	National Center for Advancing Translational Sciences
NCCOE	National Cyber-security Center of Excellence
NCCP	NIST Cloud Computing Program
NCI	National Cancer Institute
NDRC	National Development and Reform Commission
NFV	Network Functions Virtualization
NGMN 5G	Next Generation Mobile Networks 5G
NGO	Non-Governmental Organization
NIBIB	National Institute of Biomedical Imaging and Bioengineering
NiCE	Networking Intelligent Cities for Energy Efficiency
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NBD-PWG	NIST Big Data Public Working Group
NITRD	Networking and Information Technology Research and Development
NRC	Nokia Research Center

NSF	National Science Foundation
NSTC	National Science and Technology Council
OBSSR	Office of Behavioral and Social Sciences Research
OCF	Open Connectivity Foundation standard
ODI	Open Data Institute
OECD	Organisation for Economic Cooperation and Development
OEM	Original Equipment Manufacturer
OGD	Open Government Data
OIC	Open Interconnect Consortiums
OMG	Object Management Group
ON.Lab	Open Networking Laboratory
ONF	Open Networking Foundation
ONRC	Open Networking Research Center
ORION	On Road Integrated Optimisation and Navigation
OSNP	Open Sensor Network Platform
PDG	Public Data Group
PEB	Positive Energy Blocks
PFN	French Nuclear Platform
POSSE	Promotion of Open Specifications and Standards in Europe
PPP	Public-Private Partnership
PSCM	Public Service Concept Model
RANaaS	RAN-as-a-Service
RBC	Radio Block Centre
RDT&E	Defense Research, Development, Test, and Enhancement
RESS	Renewable Energy self-sufficiency
RFI	Request for Information
RFID	Radio Frequency Identification
RTOs	Research and Technology Organisations
RTP	Reference Technology Platform
RUS	Rural Utilities Service
RPAS	Remotely Piloted Air System
SAE	Smart Everything Everywhere initiative
SC	Sub-Committees
SCF	Smart City Framework
SDN	Software-Defined Networking
SERA	Single European Railway Area
SESAR	Single European Sky ATM Research programme
SETIS	Strategic Energy Technologies Information System
SGA	Smart Growth America
SGCC	Smart Grid Interoperability Panel (SGIP) Cyber-security Committee
SGDP	Smart Grid Demonstration Program
SGIG	Smart Grid Investment Grant
SGIP	Smart Grid Interoperability Panel
SMART	Sustainable Mobility & Accessibility Research & Transformation
SMEs	Small and Medium-sized Enterprises
SysML	Systems Modeling Language
SoS	System of systems
SRA	Strategic Research Agenda
SRRA	Strategic Rail Research Agenda
SSH	Smart Systems and Heat
SSI	Smart Systems Integration
SSTI	State Smart Transportation Initiative
STTP	Strategic Transport Technology Plan
Tbps	Terabit per second

TCAUP	Taubman College of Architecture and Urban Planning
TEN-T	Trans-European transport networks
TEPCO	Tokyo Electric Power Corporation
TERN	Trans European Road Network
TOD	Transit-Oriented Development
TRIP	Transport Research and Innovation Portal
TRKC	Transport Research Knowledge Centre
TRL	Technology Readiness Level
TSB	Technology Strategy Board
TWh	Terawatt hour
TX	Transformer program
UAS	Unmanned Aircraft Systems
UAV	Unmanned Aerial Vehicle
UC	University of California
UMTRI	University of Michigan Transportation Research Institute
US	United States of America
USDA-NIFA	U.S. Department of Agriculture-National Institute of Food and Agriculture
UTC	Urban Traffic Control
U-TacS	UAV Tactical Systems
VPPs	Virtual Power Plants
VSAT	Very-Small-Aperture Terminal
VSL	Variable Speed Limits
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
WeGO	World e-Governments
WG	Working Groups
WIVW	Wuerzburg Institute of Traffic Sciences
WNCG	Wireless Networking and Communications Group
WWRF	Wireless World Research Forum
W3C	World Wide Web Consortium
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
5GIC	5G Innovation Centre
5GMF	Fifth Generation Mobile Communications Promotion Forum
5G VIA	5G Vertical Industry Accelerator

1 Executive Summary

This deliverable provides a "Panorama of ICT landscape in EU and US: ICT, Policies, regulations, programmes and networks in the EU and US". Additionally, key activities in the rest of the world are also presented in order to put the work being performed within the EU and US in context. The intention of this document is to set the scene for the PICASSO project which has the intention to reinforce EU-US ICT collaboration in pre-competitive research and innovation related to key societal challenges: smart cities, smart energy and smart transportation, and in key enabling technologies: 5G Networks, Big Data, Internet of Things and Cyber-Physical Systems. These areas were chosen because they are key enablers for tackling the societal challenges mentioned above. Additionally, they are also an enabler for higher productivity and better service. Notably large ICT research and innovation efforts exist in both the EU and the US in these areas and they are also priority topics in various funding programmes in the EU and the US. As a consequence networks and critical mass in these subjects is present on both sides of the Atlantic which affords the opportunity for collaboration. The document provides:

- An overview of ICT policies, regulations, programmes and networks in the selected application domains (societal challenges): smart cities, smart transportation and smart energy.
- An overview of industry-driven programs, priorities, networks, major projects in EU and US related to 5G Networks, Big Data, Internet of Things and Cyber-Physical Systems
- An overview of ICT regulations in the selected domains.

The key aim of the document is to identify gaps and opportunities, the key challenges in the selected domains and open problems, and the needs for supporting policy measures and strategic EU-US initiatives (both policy and research related). Here potential areas where collaboration may be possible between the EU and US have been formulated as input for discussion within the PICASSO Expert Groups. These are not included in this document as they will be discussed and refined in future work. In total 15 potential areas have been identified where it may be possible to collaborate on research and policy, 16 areas where there is an opportunity to work together on regulations and 9 areas where it would be beneficial to work together on standards.

The most furtile areas for collabration are in smart cities and IoT/CPS where there are a range of research, regulatory and standardisation needs. However, there are also great opportunities in the areas of smart energy and smart transportation. In terms of the underlying technologies it is clear that a common approach and joint work on addressing 5G, Big Data and IoT/CPS would not only allow bilateral access to EU and US markets but would also allow technology and products to be sold on the world stage increasing the competitiveness of EU and US companies in existing and developing markets.

2 Introduction

Increased urbanisation is a key challenge for the future. It is predicted that by 2050, the world population will reach nine billion with a fundamental shift in demographic towards a more elderly population. The expectation is that 60% of the population will be older than 50 and that 75% of the population will live in cities. This will create great challenges to provide energy supply, logistics, health care, security, food and water. The growing rise in the use of ICT to provide interconnectivity, information and optimisation of services is leading to many “Smart” solutions being proposed. In this report we give a panorama of a number of the key projects which have been initiated in the EU, US, and around the world, to tackle these societal issues and also the underlying technologies. The key areas addressed in this report are:

Smart Cities: The area of smart cities includes many potential topics. In this domain ICT supports the efficient use of space, infrastructure and other resources in cities, e.g. integrated smart transportation concepts, lighting, smart garbage collection, optimising use of water and energy, monitoring for safety and the well-being of the inhabitants. It can also include governance, education and the monitoring of citizens health.

Smart Energy: ICT is exploited in many areas within energy supply and is used to provide availability of services, and management to reduce consumption and CO2 emissions. It is used for the stable operation of grids using secure communication systems with automatic load balancing and rejection of attacks, for incorporation of renewables within the main supply and also increasingly for interacting with grids for demand-side management. Smart energy usage is also being enabled within smart factories and smart homes to provide improved energy efficiency in buildings and production.

Smart Transportation: ICT is being used in smart transportation to provide an optimised use of infrastructure to increase capacity and also to improve the safety of road transport. This is being achieved through sophisticated traffic management systems that are relying on increased connectivity between cars and between cars and infrastructure. In the longer term the introduction of increased autonomy will lead to fundamental changes to traffic operation.

The key technologies addressed are:

5G: 5th generation wireless systems will be the next major phase of mobile telecommunications standards going beyond current 4G/IMT-Advanced standards offering data rates of 1 gigabits per second for several hundreds of thousands of simultaneous connections with extremely low latency of less than 1 ms. This coupled with reductions in energy consumption will enable massive sensor deployments in future.

Big Data: Big Data analytics covers the processing, analysis and use of very large data sets to extract value for data to provide trend detection, decision making, greater efficiency, cost reduction, risk reduction, crime and disease prevention, or to provide services to citizens.

IoT/CPS: The Internet of Things (IoT) is used to describe the network of physical objects, e.g. devices, vehicles, buildings, embedded electronics, software and sensors and network connectivity. Here objects can collect and exchange data allowing objects to be sensed and controlled remotely across network infrastructure to improve efficiency, accuracy and provide economic benefit. When feedback is introduced combining sensors with computing and actuators then Cyber-Physical Systems are created allowing control of smart cities, smart grids, intelligent transportation systems.

The social and economic challenges are common across the world and there are opportunities for the EU and the US to work together on these global challenges for mutual benefit, not only in allowing solutions from EU and US providers to be sold within each other's economic areas but also on a world-wide scale. In addition to economic benefits, there will be benefits to society and the end-users. The expectation is that by working together competitiveness on both sides will be improved leading to better products and services. Joint research and innovation will lead to a faster development of better solutions and will enable societal challenges to be addressed more efficiently. Alliances of industrial enterprises and academic institutions across Europe and the US and public-private partnerships will speed up the transfer of existing knowledge to applications and the development of new technologies for the benefit of all citizens.

Economic competition and differences in approaches and policies between the EU and the US have contributed to a number of projects and technologies being pursued on both sides of the Atlantic. The aim of the panorama report is to highlight the key initiatives that are ongoing as the basis for discussion for potential alignment of activities with a view towards pre-competitive research and innovation. Through collaboration it will be possible to enable faster development, provide opportunities to market new technologies, and better understand the cultures and markets that exist. Collaborative efforts will lead to addressing the global challenges more efficiently and will create market opportunities in the US and world-wide for European companies, who will also have advantages from new open innovation opportunities at a greater scale. In the other direction, EU citizens will profit from open competition.

2.1 Objectives and Scope of the Deliverable

PICASSO WP1 (EU and US Industrial drivers, societal needs and ICT collaboration landscape) aims to set the scene for the project execution. In order to achieve this, this deliverable as part of task 1.2 in particular provides a panorama of the EU and US ICT landscape focusing on ICT policies and regulations, programs, priorities, networks and projects.

The task objectives are to provide an:

- Overview of ICT policies, regulations, programmes and networks in the selected application domains (societal challenges): smart cities, smart transportation and smart energy.
- Overview of industry-driven programs, priorities, networks, major projects in EU and US related the expert group topics.
- Overview of ICT regulations in the selected domains.

These areas are consolidated in this report (D1.3) to provide a Panorama of ICT landscape in EU and US: ICT, Policies, regulations, programmes and networks in EU and US. The deliverable is organised in three main sections addressing the societal challenges, the thematic domains and identified opportunities for collaboration.

2.2 Panorama of ICT Landscape in the EU and US

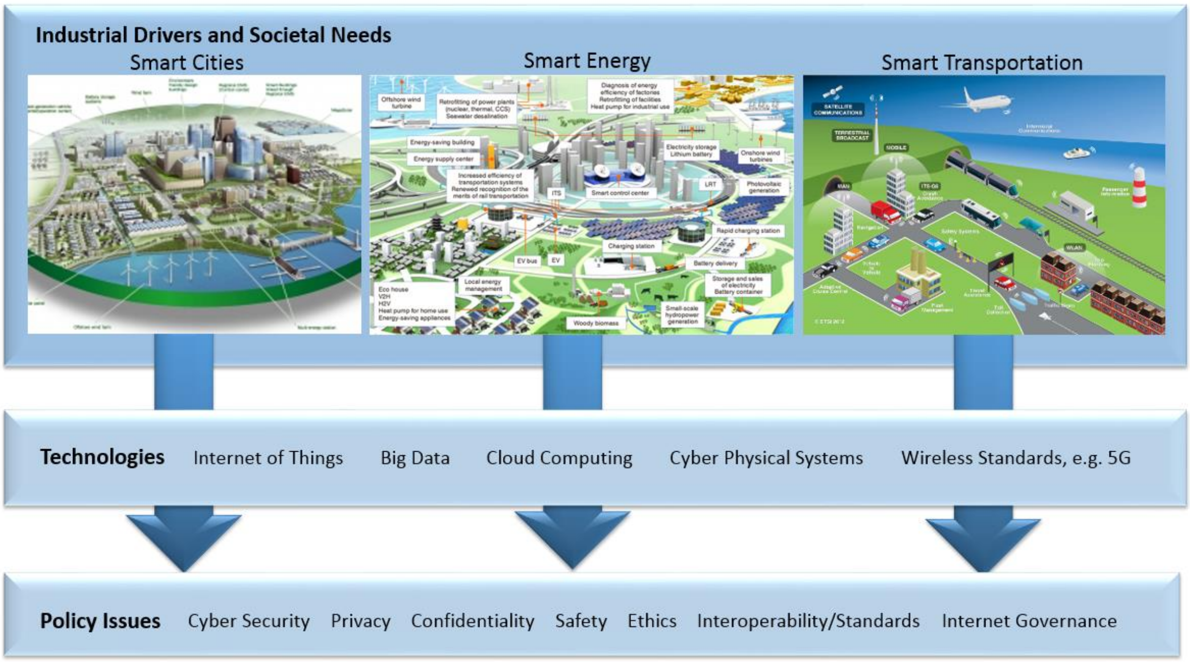


Figure 1. PICASSO Panorama

In this section the panorama of ICT landscape in EU and US is described. It should be noted that the area is huge and thousands of references have been found across the domains. It is not possible to include all of these references or describe all of the work that is going on in the area. Thus the report confines itself to the key programmes, projects or initiatives that are being undertaken within the EU and US. Additionally, where considered appropriate, references to other key work going on in the rest of the world is included to set the overall context of the work. The panorama covers the ICT, policies, regulations, programmes and networks in three societal challenges: Smart Cities, Smart Energy and Smart Transportation as depicted in Figure 1.

These societal needs as well as the industrial drivers will drive the future of key technologies such as Internet of Things, Big Data, Cloud Computing, Cyber-Physical Systems and Wireless Standards such as 5G. The successful deployment of these technologies is also dependent upon policy issues and introduction of standards. Here there needs to be harmonisation across Europe and also across the world. There are needs to consider cyber-security, privacy, confidentiality, safety, ethics, interoperability/standards, and internet governance. In many cases these are currently barriers to adoption of new solutions and so they need to be addressed at a fundamental level on a world-wide scale.

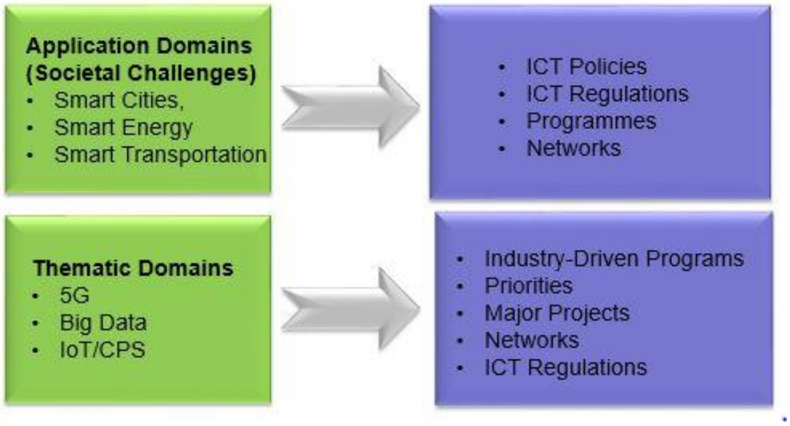


Figure 2. Panorama of ICT Landscape

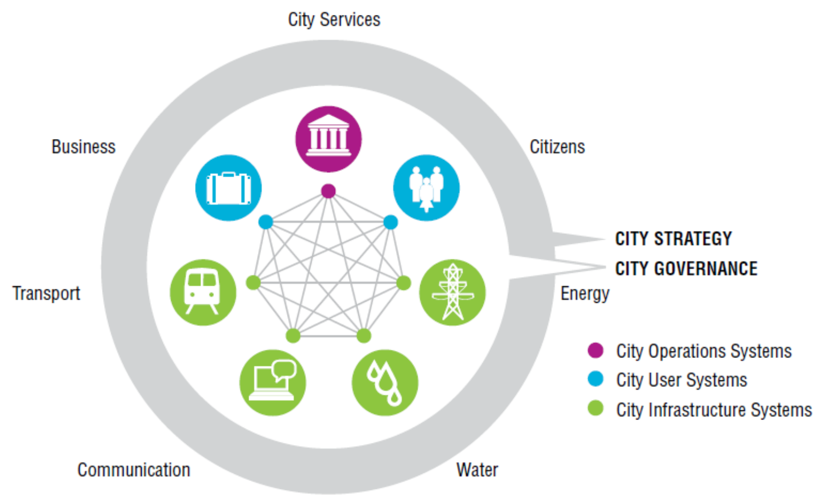
The structure of this report follows that shown in Figure 2. The application domains (societal challenges) of Smart Cities, Smart Energy and Smart Transportation are first considered, and for each of this, the main programmes, projects, networks and important ICT policies and ICT regulations within EU and US are highlighted. It should be noted that the topic of Smart Cities has a very wide remit and also covers Smart Energy and Smart Transportation to some extent. Here care has been taken to try and make a distinction between projects/programmes directly addressing Smart Cities and those that are focused on Smart Energy and Smart Transportation.

Following this, at the technological level, the main industry-driven programs, priorities, projects, networks and ICT regulations both in Europe and in the US for each of the selected thematic domains (5G, Big Data, IoT/CPS) are considered. This corresponds to the 3 main technical working groups set up within PICASSO. The intention of this is to provide input into the 3 main working groups, defining a baseline and starting point for discussions. In addition a 4th working group will consider the policy issues that need to be addressed providing input papers in a number of areas, e.g. privacy.

3 Smart Cities

3.1 European Smart City Drivers and Policy Activities

Increased urbanisation combined with increased instrumentation and interconnection is leading to cities having greater control and intelligence. Smartness allows cities to be sustainable and to provide public safety, health and education services and a good quality of life for their citizens. 100 years ago less than 20 cities had populations of more than 1 million people. Now there are 450 [1] and as they get larger they gain greater economic, political and technological power becoming the hubs of a globally integrated, services-based society. Operationally, cities are based on a number of core systems composed of different networks, infrastructures and environments related to their key functions: city services, citizens, business, transport, communication, water and energy. Service delivery is provided by the city authority which coordinates delivery across different agencies, allocating public funds and conducting physical planning activities. Cities also offer citizens and business the ability to move things around through their transport systems and to share ideas and information through their communication systems. Two core utilities necessary for all economic and social activity are also provided – water and energy. EC assessments have shown that resident engagement is one of the most critical conditions for the operating effectiveness of all smart city programmes/initiatives. Residents are both end-users and are at the front-end of energy use and energy services [2].



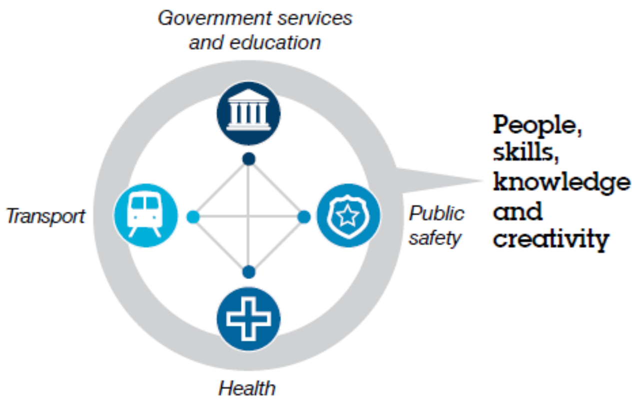
Source: IBM Center for Economic Development analysis.

Figure 3. Cities Systems and their Interrelationship with City Strategy and Governance

The various systems are interconnected to form a “system of systems” as shown in Figure 3. There is a need to balance the demands of each through service coordination, considering health, the needs of business, efficiency of transport, water leakage/quality and energy efficiency/security.

New technologies provide much greater scope for instrumentation, interconnection and intelligence of a city’s core systems. However, it is necessary to carefully prepare and plan how to implement change and take into account the interrelationships between systems. For example, transport, business and energy systems are closely interrelated as they are all users of energy. Likewise a substantial amount of electricity is used in pumping and treating water. In Malta, for example, a new smart utility system will inform citizens and business about their use of both energy and water, enabling them to make better decisions about resource consumption. A reliable and efficient transport system reduces congestion, reduces carbon dioxide emissions improving health, and reduces stress and accidents. Less congestion also frees up more time for people to work or to relax.

The application of advanced information technology, analytics and systems thinking is also likely to attract, create, enable and retain citizens' skills, knowledge and creativity within cities [3]. With urbanisation cities contain an increasingly large number of the world's highly skilled, educated, creative and entrepreneurial population, giving rise to highly concentrated and diverse pools of knowledge and knowledge-creation networks. This coupled with the ability of cities to support large-scale business and investment networks is likely to see a shift towards cities for economic development. Thus strategic design of public services, building better transportation systems, supporting creative innovation and technological research and development, can be used to provide a supportive environment for delivering a higher quality of life, making a city more attractive to a knowledge-based population (See Figure 4).



Source: IBM Global Center for Economic Development.

Figure 4. Interrelated Connections between a City's Core Systems

3.1.1 Scope of Smart Cities

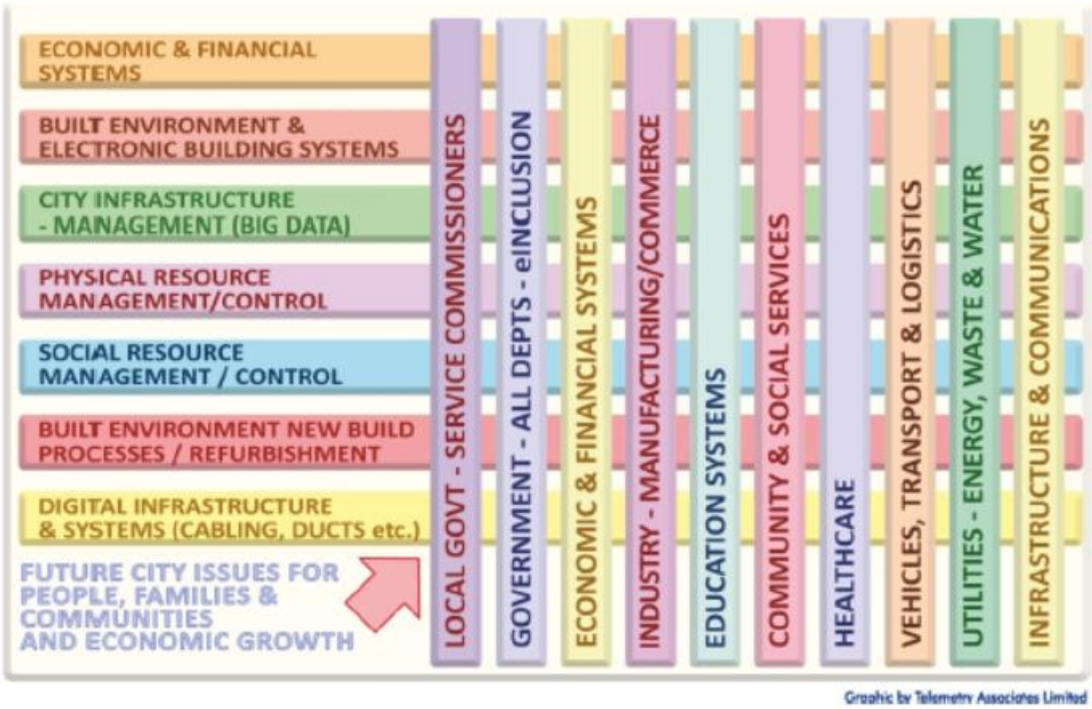
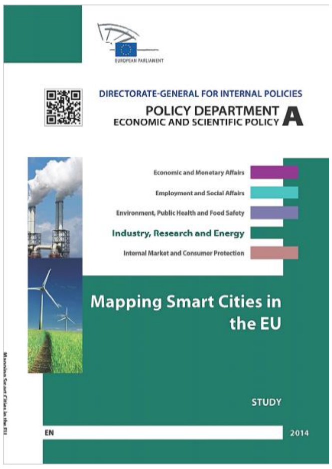


Figure 5. Areas that are Encompassed by Smart Cities

A challenge when considering the domain of Smart Cities is the shear breadth of different areas that are being affected by the use of ICT. In Figure 5 one can see that “smartness” can be utilised in a number of areas affecting government, economic and financial systems, building management, manufacturing, education, community and social services, healthcare, transportation, utilities (energy, waste water, rubbish collection, etc.), infrastructure and communications. In this report it is not possible to consider all of these domains as there are literally 1000’s of references covering all of these areas. The panorama report thus concentrates on the domains which are of particular interest to PICASSO.

3.1.2 Smart Cities within Europe

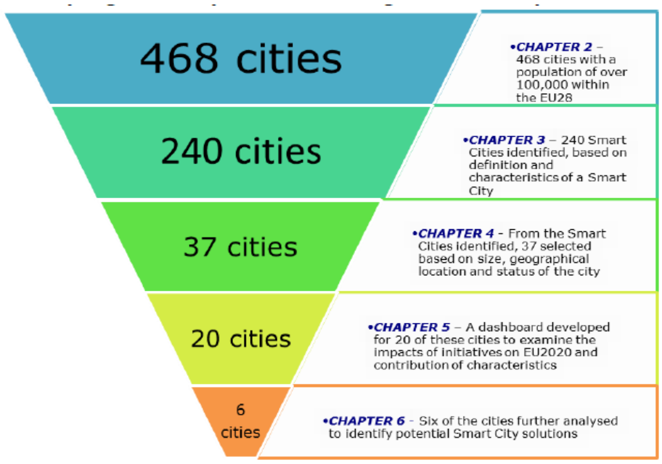


Recommendation	Intended for
Understanding Smart Cities: research and evaluation	
Detailed panel of longitudinal case studies with city-level funding and outcome data	DGCNECT, DG JRC
Standardised evaluation and assessment methods to measure success at internal, city and European level for impact assessment and benchmarking	The European Commission (EC) and Impact Assessment Board (IAB)
Develop methods and structures for a needs assay of the city's performance against relevant targets and presentation scorecards	Collective effort led by existing Smart City clusters[1]
Designing Smart City initiatives and strategies	
Mandate specialised impact assessment guidelines for Smart City strategies and initiatives to include: SMART objectives, issues of timing and uncertainty, and assessment of experimental variation	Funding bodies,[2] IAB, Smart City clusters
Promote local modularity for early-stage initiatives	Funding bodies, Smart City clusters; additional specific funding from EC, local government stakeholders
Facilitate exit and change of participation during the latter stages of an initiative	Funding bodies, Smart City clusters, local government stakeholders
Structural conditionality in funding for Smart City initiatives	Funding bodies
Specific design procedure for structuring Smart City initiative components	IAB, Smart City clusters, local government stakeholders (as monitoring hosts)
Smart City governance	
European-level Smart City platform with brokerage or intermediary functions	EC
Privileged or low-cost access to existing infrastructures	Local government stakeholders, infrastructure operators, national regulatory agencies
Mandatory multi-stakeholder governance with lay users represented and on integrated project teams	Funding bodies and government authorities and participants
Encourage industry-led public–private partnership consortia	Funding bodies and government authorities and participants

Recommendation	Intended for
Supporting the development of Smart Cities	
Use demand-side measures to stimulate demand for city-based 'Smart solutions'	Member State and local government procurement agencies, Horizon 2020, service users, standards bodies, national regulatory agencies
Selective use of regulatory forbearance and/or pro-competitive sourcing	Procurement agencies, national regulatory agencies, European Parliament
From Smart Cities to a Smarter Europe: replication, scaling and ecosystem seeding	
Periodic assessment of scalability potential and identification of instruments and activities to optimise pan-European dissemination of good practices and solutions	EC (platform), IAB (guidelines), local authority participants
Include Smart Cities as a future internet public–private partnership (PPP) use case or involve Smart City stakeholders in large-scale pilots, standards bodies, etc.	Future Internet Public–Private Partnership (FI-PPP), Horizon 2020, EC (supporting standards body engagement with additional specific funding)
Expand support for Smart Cities and Communities – European Innovation Partnership	EC
Additional resources for Smart City translation and transfer	EC, Member States
Create and encourage Smart City-specific new intellectual property ownership rights and contract forms	EC, Council, Parliament; possible WIPO

Figure 6. Mapping Smart Cities in EU

In [4] (See Figure 6) a study was performed within Europe to identify the key characteristics of smart cities and what the main areas of interest were. The report also makes recommendations on assessment of initiatives, how to design initiatives and strategies, on governance within smart cities, on how to support smart cities and how to replicate best practice across Europe and on creating a supporting ecosystem.



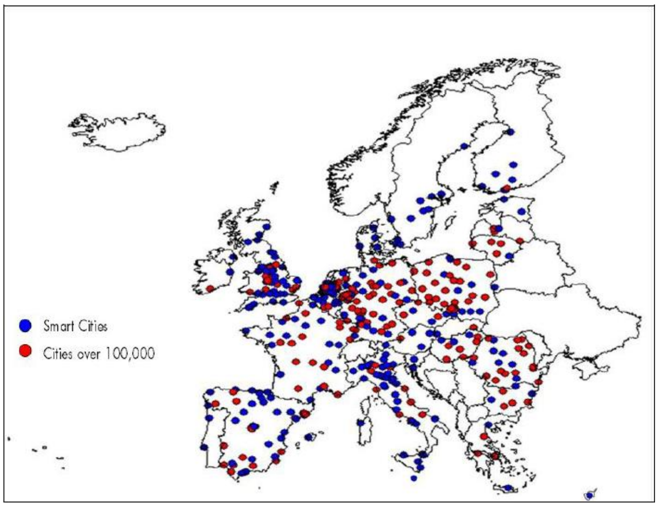
KEY FINDINGS

- There are many definitions of Smart Cities. Some focus on ICT as a technology driver and enabler, while broader definitions include socio-economic, governance and multi-stakeholder aspects such as the use of social participation to enhance sustainability, quality of life and urban welfare.
- In any case, a Smart City is quintessentially enabled by the use of technologies (especially ICT) to improve competitiveness and ensure a more sustainable future by symbiotic linkage of networks of people, businesses, technologies, infrastructures, consumption, energy and spaces.
- In this study, **a Smart City is a city seeking to address public issues via ICT-based solutions on the basis of a multi-stakeholder, municipally based partnership. These solutions are developed and refined through Smart City initiatives**, either as discrete projects or (more usually) as a network of overlapping activities.
- More concretely, the strategies and initiatives of a Smart City must include at least one of the following characteristics (objectives and/or modes of operation): Smart Governance, Smart People, Smart Living, Smart Mobility, Smart Economy and Smart Environment. These characteristics constitute the ends for which stakeholders participate in a Smart City initiative (e.g. to solve an environmental issue).
- The means by which those ends are achieved include a range of components: technologies; material, financial, organisational and knowledge inputs; processes; and norms or standards. These components may already be present or may be created specifically for use in Smart City initiatives.
- Components therefore provide the building blocks of Smart City initiatives and comprise three types: technological, human and institutional.

Figure 7. Key Findings

The report performs an analysis of 468 cities. These are considered with respect to roll out of different smart technologies and their use. From this there are a number of key findings as shown in Figure 7. These indicate that there is different emphasis across Europe in the areas that are being addressed within Smart Cities. In some areas, smart transport, and emissions are key drivers. In other areas smart governance is a key driver. It should also be noted that a city can be large or small and also be smart, indeed the majority of smart cities tend to be smaller, however larger cities display the use of more smart technologies as one might expect.

Finally, the report considers 6 smart cities in particular which display the highest number of smart features. These were Amsterdam (the Netherlands), Barcelona (Spain), Copenhagen (Denmark), Helsinki (Finland), Manchester (UK) and Vienna (Austria). Notably Amsterdam started work on its smart city initiative in 2009 and has 45 on-going projects in three geographic areas: Nieuw West, Zuidoost, and Ijburg. Barcelona launched the first smart city project in 2010. Other cities are also covered in the report, e.g. Tallinn and Santander, but these display less smart features.



	Smart Neighbourhoods	Testbed micro infrastructures	Intelligent traffic systems	Resource management systems	Participation platforms
	10	7	11	14	8
Smart Environment	+++	+++	++	+++	*
Smart Mobility	++	++	+++		*
Smart Governance				++	+++
Smart Economy	++	++		++	++
Smart Living	++			+	*
Smart People	++			+	++

50 Smart City Projects

Figure 8. Smart Cities in Europe

At a country level looking across Europe, the UK, Spain and Italy have the largest number of Smart Cities with more than 30 each as shown in Figure 8. Notably of the larger countries Germany and France have fewer Smart Cities. In general smaller countries have absolute lower numbers of Smart Cities. It can also be seen in Figure 8 that considering the top 50 smart cities the projects are concentrated in the areas of smart environment (12 projects), smart mobility (8 projects) and smart economy (8 projects).

3.1.3 European Commission Policy Actions and Strategy

The European Commission has been extremely active in trying to promote the concept of smart cities [5] and also in encouraging development of new technologies to enable smart features to be introduced into leading cities. Here the Commission has produced a number of strategic documents [6], communications [7], and also developed research strategies [8].

3.2 Key European Initiatives on Smart Cities

3.2.1 European Initiative on Smart Cities

The European Initiative on Smart Cities has been set up to support cities and regions. The key goal is to help cities take ambitious and pioneering measures to reduce greenhouse gas emissions by 40% through sustainable use and production of energy by the year 2020. This requires a systemic approach and organisational innovation to bring together measures on energy efficiency, adopt low carbon technologies and encourage smart management of supply and demand. The initiative particularly targets buildings, local energy networks and transport and builds upon existing EU and national policies and programmes, such as CIVITAS, CONCERTO and Intelligent Energy Europe.

This is also linked to the Strategic Energy Technologies Information System (see section on SETIS in Smart Energy section) which has an industrial initiative in the areas of Solar and Electricity Grid, as well as the EU public-private partnership for Buildings and Green Cars established under the European Economic Plan for Recovery. Notably this initiative is mobilising the Covenant of Mayors which involves local government from more than 4500 cities and has produced a strategic roadmap of activities covering building, heating and cooling, electricity and transport as shown in Figure 9.

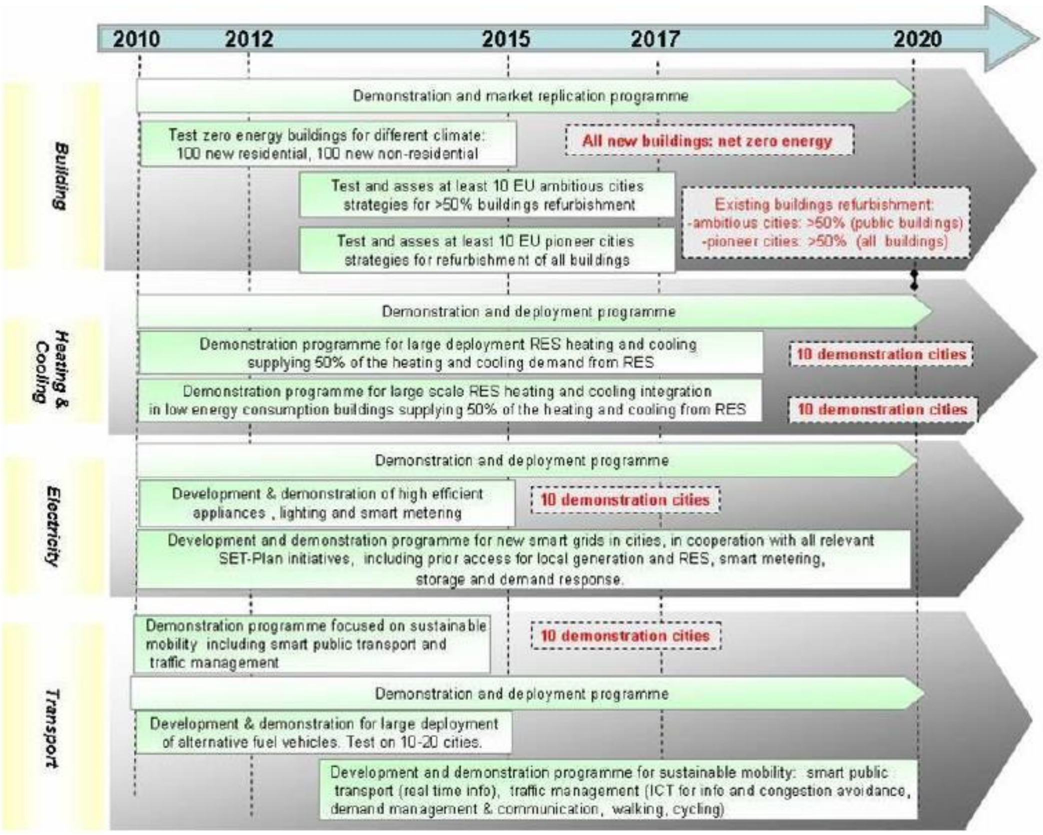
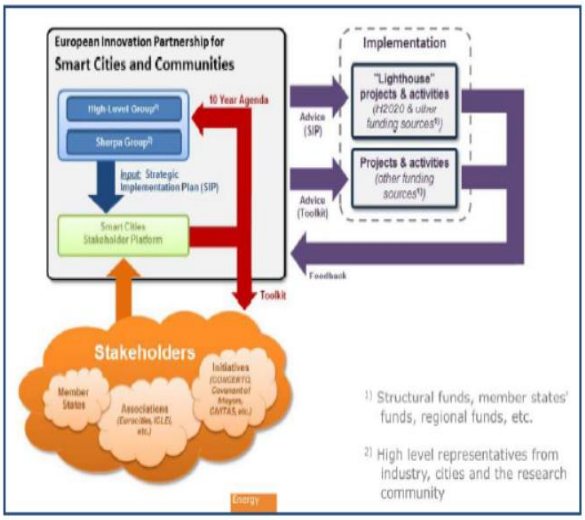


Figure 9. Smart Cities Technology Roadmap [9]

3.2.2 European Innovation Partnership for Smart Cities and Communities



The High Level Group

A "High Level Group" as the first pillar of the EIP-SCC, consists of CEOs from research-intensive industries, city mayors, regulatory authorities and public financing institutions. It was established to support the implementation of the EIP-SCC. It is responsible (together with a "Sherpa Group") for the Strategic Implementation Plan (SIP), which helps define how concepts to promoting Smart Cities are put into practice. It also looks at how the European Commission can support these measures during the next Research Framework Programme – Horizon 2020.

EU Smart Cities Stakeholder Platform[1]

The EU Smart Cities Stakeholder Platform is the second governance body of the Smart Cities and Communities European Innovation Partnership (EIP-SCC).It was initiated by the European Commission with the dual aim of

- identifying and spreading relevant information on technology solutions and needs required by practitioners; and
- providing information for policy support to the High Level Group and the European Commission.

It is both a web-based and physical Platform open to anyone who registers on it. Backbone is the contributions by stakeholders in a bottom-up way, owned by the stakeholders. The Platform will bring city authorities, industry, NGOs and civil society together. It will accompany the implementation of the lighthouse projects and monitor overall implementation of the Innovation partnership. It will organise activities so that experience and knowledge from lighthouse projects will be shared.

Figure 10. European Innovation Partnership for Smart Cities and Communities

The European Innovation Partnership for Smart Cities and Communities (EIP-SCC), as shown in Figure 10, is not a single initiative but actually part of a much broader effort by the EC to foster a new approach to EU research and innovation [10][11][12][13][14][15]. In total five European Innovation Partnerships have been launched. The European Innovation Partnership for Smart Cities and Communities is targeted at investing in sustainable development in as many cities as possible. The key word is "partnership" and the concept is to create equal partnerships between cities and companies based on synergies between ICT, energy and mobility. As highlighted by Günther H. Oettinger, then the EU Commissioner for Energy in November 2013, "This will lead to projects that make a real difference in our everyday lives."

Neelie Kroes, European Commissioner for the Digital Agenda in May 2013, stated that "By being geared towards addressing the energy and climate targets that we have set for ourselves for 2020, this Partnership will focus on our cities becoming more sustainable places to live and do business in. This translates into a greater quality of life for the individual citizen. It also provides tremendous market opportunities for local innovative businesses as well as our industries in the ICT, energy and mobility sectors."

The partnership is organised into a High Level Group made up of CEOs from large companies, city mayors, regulatory bodies and public financing institutions, and the Smart Cities Stakeholder Platform. This platform is used to engage with stakeholders who can sign up to the platform allowing development of an ecosystem in a bottom up fashion.

Supporting this is a need to create "Lighthouse Projects" which act as a model project with a signal effect for numerous follow-up projects in other cities. €81 Million of EU funds have been earmarked for these covering two sectors relevant to smart cities: transport and energy.

A Strategic Action Plan has also been created to promote Smart City concepts on a wider scale. This is shown in Figure 11.

Figure 11. European Innovation Partnership on Smart Cities and Communities Operational Implementation Plan (Sherpa Group Feb 2014)

The plan covers activities in sustainable mobility, sustainable districts and the built environment and also in integrated infrastructures and processes. Here a holistic approach is being adopted covering citizens and decision makers, knowledge and data gathering, funding and procurement.

3.2.3 Small Giants

In order to address smaller cities the Small Giants Initiative is targeted at small and medium-sized European cities that have less than 250,000 inhabitants. This has been set up by the EIP to strengthen the uptake of smart city solutions by facilitating networking, project building, fostering access to finance and ensuring up-scaling of solutions. Small Giants is project-driven and considers the fact that such cities have budget constraints and limited human resources [16].

3.3 National Initiatives

In addition to European-wide initiatives, there are a large number of national initiatives across Europe as indicated in section 3.1.2. It is not possible to include information on all initiatives in this document and thus a number of key initiatives are highlighted which are of interest.

3.3.1 UK

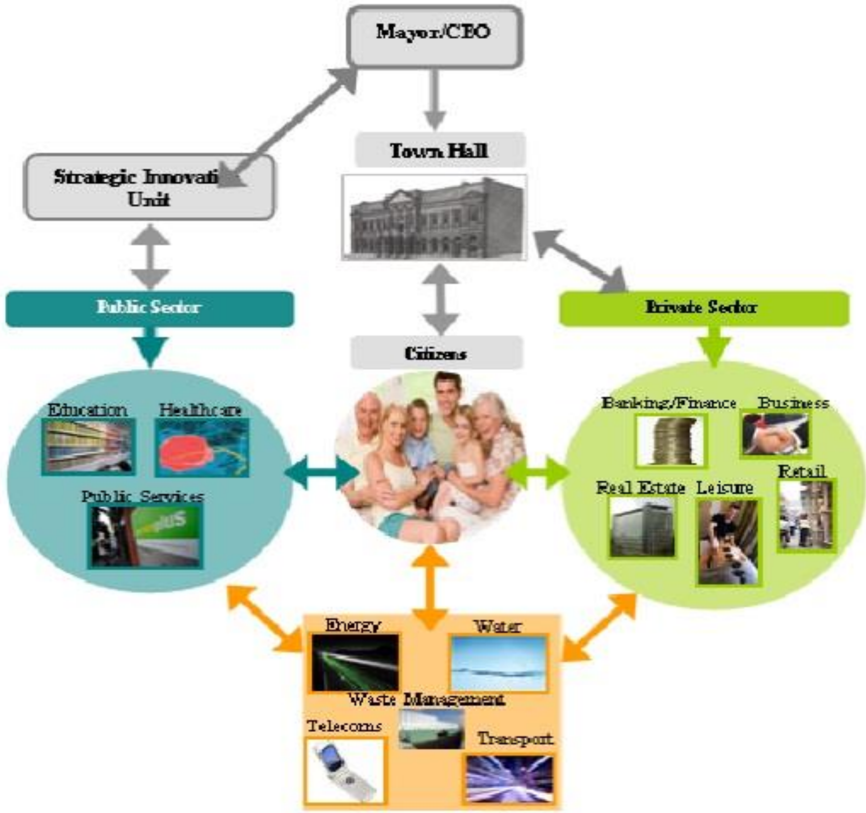


Figure 12. Smart City Model [17]

Smart Cities is a particularly active area in the UK with many initiatives. The UK does not want to only address the issues driving smart cities within the UK but also sees the area as a huge opportunity for UK businesses to develop new products and services that can be sold worldwide. The smart cities industry is valued at more than $400 billion globally by 2020, with the UK expected to gain a 10% share ($40 billion) [17] [18] [19] [20]. The area is thus being strongly driven by the UK Government via the Business, Innovation and Skills Department and over 100 opportunities for UK businesses have been identified. The UK Government has set up policies and programmes in five main areas:

- encouraging and empowering city authorities to develop the vision and leadership to provide solutions to their own problems;
- promoting open data and the capacity of organisations to improve access to open data, to share and to use it, including the development of open standards;
- programmes to develop underpinning technologies and to demonstrate their efficacy;
- Departmental programmes to encourage the adoption of new approaches and technologies, to transform both the service systems and consumer behaviour;
- participating actively in EU programmes.

A report [20] has been produced to highlight the economic opportunities from smart cities and also the impact on citizens (See Figure 12). Here the provision of real-time information will allow people to make more informed choices, e.g. for journey planning on buses and trains and for identifying car parking spaces. Supporting this a Smart Cities Forum has been established, chaired by Universities and by the Science and Cities Minister David Willetts and Cities Ministers with representatives from cities, business, and scientists.

The Technology Strategy Board (TSB) has also set out a strategy for future cities [21] which focuses on a more integrated and systemic approach to services. Two key initiatives have been launched to support UK businesses:

3.3.1.1 Future Cities Demonstrator Programme

The Future Cities Demonstrator Programme is a £33m TSB project which aims to demonstrate at scale the added value of integrating city systems [22][23]. 30 UK cities submitted proposals and Glasgow City Council won the funding with a project to integrate services across health, transport, energy and public safety. The city allows UK businesses to test new solutions that can be replicated around the globe. Additionally, Bristol, London and Peterborough also each received £3m and the four cities will work together to provide data into the Future Cities Laboratory and to the TSB DALLAS (Digital Assisted Living Lifestyles at Scale) project, already underway in Glasgow. This project is addressing smart applications of technology for transport and energy and it has recently announced Catapults in relation to Transport Systems and the Connected Digital Economy.

3.3.1.2 Future Cities Catapult UK

The other major initiative in the UK is the Future Cities Catapult. This has been set up to act as a global centre of excellence in urban innovation bringing together business, universities and city-administrations to develop solutions for future cities. Within the Catapult there are a number of activities. A Cities-Lab has been set up to integrate live data feeds with advanced modelling, simulation, analysis and visualisation with the aim to create new, commercialisable solutions to city challenges. The idea behind this is to allow cities, SMEs and businesses to plug in their data and ideas and use this as a test bed for improved solutions and new business models. There is also a "futures and best practice programme" to identify what future sustainable and future-proofed cities will look like and to showcase best practice. The idea behind this is to identify factors for success, risks and opportunities which can be used as input to decision makers and innovators. There is also work on addressing the barriers to city-integration, such as uneven procurement rules, inadequate templates for shared IP and lack of financing. It has been noted that the deployment of integrated city systems depends both on financial as well as technological innovation. A challenge here is to identify new financial models to create potential revenue streams and turn these into infrastructure investment. To address this, the UK's financial sector is also engaged in the Catapult.

3.3.1.3 e-infrastructure Leadership Council

The UK government has also set up an e-infrastructure Leadership Council as a recommendation of the Tildesley Report "A strategic vision for the UK e-infrastructure" in 2011 [24]. The Leadership Council targets collaborative academic and industrial precompetitive research and development. Under this the Digital Economy Programme has been set up. This looks at how, in sustainable societies of the future, people will be able to make informed sustainable choices. Examples of this are the Urban Prototyping London project which is addressing how digital technologies, art, and design can be used as new tools by citizens to improve their urban environment. Here questions of citizen empowerment and engagement are being considered along with identification of the real barriers to entry for creating start-ups. A £5.9M five-year multi-disciplinary research programme, Digital City Exchange (DCE), has been set up to find innovative solutions to optimise the use and planning of cities and a 5 year EPSRC programme grant has been given to develop a method of designing and engineering UK cities called Liveable Cities. This will be achieved via the development of a City Analysis

Methodology (CAM) that will measure how cities operate and perform in terms of their people, environment and governance, taking account of wellbeing and resource security.

3.3.1.4 UK - HyperCatCity

Innovate UK has also provided £8 million of investment for the HyperCat project [25][26] to create common, secure standards and protocols to unlock the potential of the 'Internet of Things'. Under this HyperCatCity brings together London, Bristol and Milton Keynes working on better and more effective ways to deliver services.

3.3.2 Spain

The leading smart cities project in Spain is Smart Santander [27]. Smart Santander is a city-scale experimental research facility that supports typical applications and services for a smart city. The facility is large, open and flexible to enable horizontal and vertical federation with other experimental facilities. A key aim is to stimulate development of new applications including experimental advanced research on IoT technologies and provide realistic user acceptability tests for assessment. The project has a vision not only to deploy 12,000 sensors in Santander but also to deploy 20,000 sensors in Belgrade, Guildford and Lübeck.

The population of Santander is 180,000. 10,000 sensors have been installed around downtown Santander, in a 6 square kilometre (2.3 square miles) area hidden inside small grey boxes attached to street lamps, poles, building walls and buried beneath parking lots. The sensors measure light, pressure, temperature, humidity, and the movements of cars and people [28]. Every 2 minutes data is sent to the Muñoz's laboratory at the University of Cantabria where there is a central computer that builds a picture of the city. Buses, taxis and police transmit their position, mileage and speed, as well as environment data, such as ozone or nitric oxide pollution levels. Citizens can also download an app and become human sensors themselves. Using the data information can be provided on traffic jams, and exceedance of EU air quality and noise levels. The street lamps have also been made smart and are able to adjust their brightness according to weather conditions (or when no-one is on the street) and warn when a bulb has failed. Sensors are also being used to control the amount of watering done in parks to conserve water.

Citizens can access data via a "Pulse of the City" app which gives information on when the next bus is going to arrive at a bus stop, the program of events for the city's concert hall, tourist information and special offers at supermarkets. Citizens can also take photos of potholes and send it to city hall to request repair. The time for repair is made public which puts the onus on the city to repair the problem as soon as possible. The city mayor is making data available to encourage programmers to create more apps to make Santander even smarter.

3.3.2.1 Barcelona

Barcelona has established a number of projects that can be considered "smart city" applications within its "CityOS" strategy [29]. Sensor technology has been implemented in the irrigation system in Parc del Centre de Poblenou, where real-time data is transmitted to gardening crews about the level of water required for the plants. Barcelona has also designed a new bus network based on data analysis of the most common traffic flows in Barcelona utilising smart traffic lights to optimise the number of green lights. If there is an emergency the approximate route of the emergency vehicle is entered into the traffic light system and the traffic lights are set to green as the vehicle approaches using GPS data and traffic management software to minimise delay [30].

3.3.3 The Netherlands

Amsterdam Smart City (ASC) is a partnership between companies, governments, knowledge institutions and the people of Amsterdam to turn the city into a Smart City [31]. The aim is to use social and technological infrastructures and solutions to facilitate and accelerate sustainable economic growth. ASC believes in a habitable city where it is pleasant to both live and work. In six years ASC has grown into a platform with over 100 partners, which are involved in more than 90 innovative projects [32]. There are many different ideas that could be applied and the Amsterdam Smart City platform challenges parties to submit and execute innovative solutions to urban issues. ASC also addresses the possibilities to strengthen previous activities. This advances the development of new markets for innovative solutions. Where possible, these solutions are replicated elsewhere in the city. ASC works on a number of focus areas, in which projects, ideas and new business models are being developed: smart mobility, smart living, smart society, smart areas, smart economy, Big & open data, and Infrastructure (water/roads/energy/ICT).

3.3.4 Estonia - Tallinn

The Estonian capital Tallinn is one of the world's most technologically advanced cities. Notably it was the birthplace of Skype [33]. Tallinn residents are used to exploiting technology, with high levels of automation and use of connection to the Internet. Estonians are required to carry chip-embedded identification cards, without which they do not officially exist. The cards are used for voting, prescriptions and many other transactions. For example:

- Riding the bus is free with a smart card allowing the transit authorities to track your movements.
- Mailing a package requires a sender to use a cell phone to request a code from the electronic post office downtown. The code opens a locker where the package is posted.
- City parking utilises a text message with the car's number and code for the parking lot. The fee appears on the monthly cell phone bill, which is paid electronically.

Private sector Internet and communications technology is a priority in Estonia and a “Silicon Valley” called Technopolis has been set up near Tallinn airport that demonstrates the latest high-tech services and gadgets to visitors. This reliance on Internet technologies also means that Estonia has also had to deal with cyber-attacks. The country hosts routine cyber-attack drills in which government and private sector participants learn to spot and stop phishing attempts and viruses. NATO has set up a Cooperative Cyber Defence Centre of Excellence in Tallinn. International lawyers in Estonia have also recently completed the "Tallinn Manual," the world's first comprehensive guide on legally tackling cyber-warfare.

3.3.5 Sweden - Stockholm

Stockholm’s smart city technology is underpinned by the city owned Stokab dark fibre system which provides a universal fibre optic network across Stockholm [34]. A Green IT strategy is being pursued to reduce the environmental impact of Stockholm using IT to control energy efficient buildings (minimising heating costs), traffic monitoring (minimising the time spent on the road) and development of e-services (minimising paper usage). The e-Stockholm platform provides a number of e-services, including political announcements, parking space booking and snow clearance. Combining information with GPS data allows route planning in the city. This is further being developed through GPS analytics, allowing residents to plan their route through the city.

3.4 US Smart City Drivers and Policy

In the US the White House Smart Cities Initiative [35] plans to invest over $160 million in federal research. The aim is to set up more than 25 new technology collaborations to help local communities tackle key challenges such as reducing traffic congestion, fighting crime, fostering economic growth, managing the effects of a changing climate, and improving the delivery of city services. The President's 2017 budget proposal includes a focus on Cyber-Physical Systems and Smart Cities.

3.4.1 Smart and Connected Communities Framework

The NITRD Cyber-Physical Systems (CPS) Senior Steering Group (SSG) has set up a framework to help coordinate Federal Smart Cities activities and agency investments with outside collaborations. The aim is to guide foundational research and accelerate the transition into scalable and replicable Smart City approaches. The framework [36] outlines a vision and an approach agencies can use to move forward together in pursuing a vision of smart and connected communities. Activities within the framework include the entire pipeline from research and development to deployment of new tech-driven services and infrastructure in cities. The success of the framework requires federal interagency coordination, cooperation with external stakeholders, and next steps planning for Federal action.

3.4.1.1 National Institutes of Health (NIH)

One such action is the coordination, with the National Institutes of Health, of a programme to enhance collaboration between researchers, citizens, local cities and municipalities to evaluate the health-related benefits of networked sensors, infrastructure, and computing systems. This particularly addresses health-related Cyber-Physical Systems in smart cities considering security, privacy, health disparities and human factors. Key aims are to develop interoperable solutions and gain consensus on standards, and evidence that the technologies are safe, effective and sustainable.

3.4.1.2 National Institute of Standards and Technology (NIST)

A number of initiatives are being supported by NIST. The Global City Teams Challenge, brings together teams of cities and innovators working in partnership to use Internet of Things technologies for improving the safety, sustainability, livability, and workability of communities worldwide. NIST is also running open, consensus-based public working groups to develop a comprehensive framework for the design, evaluation, and operation of complex Cyber-Physical Systems, including smart city technologies at scale. Via the Smart Grid program and public-private Smart Grid Interoperability Panel, NIST is working with the private sector on smart grid interoperability and security standards. Security is a key concern and here NIST is working with the National

Cyber-security Center of Excellence (NCCOE) to provide real-world cyber-security solutions based on commercially available technologies for smart city applications. This is for a range of applications including energy, transportation, and finance. The NIST Big Data program also has a public working group and is driving standards effort to produce a reference architecture that is vendor-neutral, technology- and infrastructure-agnostic. Finally, the NIST Cloud Computing Program (NCCP) is providing a cloud architecture and metrics to enable secure storage, transmission and processing of data and services through collaborations with public working groups, industry and international standards efforts. The overall aim is to promote international standards for interoperable smart city solutions.

3.4.1.3 National Science Foundation (NSF)

The NSF is bringing together academic researchers, industrial and non-profit partners, and local cities, municipalities and regions to integrate data sources and networked computing systems with people, physical devices and infrastructure. Applications include health and wellness, energy efficiency, building automation, transportation, etc. The aim of NSF is to support foundational research that supports design and management of Smart and Connected Communities.

3.4.1.4 National Aeronautics and Space Administration (NASA)

NASA also has a role to play and here research is being performed on collaborative, planning, and scheduling applications to enhance multi-modal traffic flow in Smart Cities and also UAV operations. NASA are also developing and sharing verification and validation tools to enable smart city developers to assure high integrity, robust, and interoperable complex systems.

3.5 Smart City Activities within the US

There are a number of smart city initiatives in the US. In the following sections a number of key activities are highlighted.

3.5.1 Boston

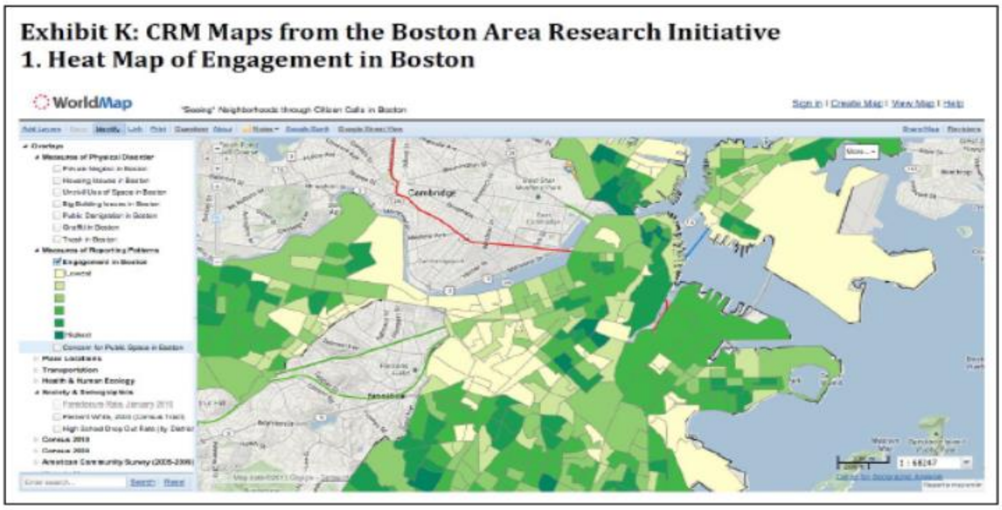


Figure 13. Boston “Heat Map” of Engagement [37]

One of the leading cities in the US with respect to implementation of smart cities ideas is Boston. Boston launched its first CRM system (Citizen Relations Management) in 2008 [38]. There are five city apps with a wide range of uses, from allowing citizens to report neighborhood problems to the government, to helping commuters find on-street parking in the Innovation District. Here there has been very good engagement with community. A heat map from an area in Boston showing the levels of engagement is shown in Figure 13. Here the darker the green, the higher the engagement [37].

The city has introduced Time to Destination message signs and Smart Parking Sensors to better manage traffic. Additionally, it is using smart technology to encourage children to walk to school and collects data from its Hubway rental bikes to reduce road congestion. The city has also introduced solar powered benches that can charge gadgets and monitor air quality and sound levels.

3.5.2 San Francisco

San Francisco is a global leader in smart-city projects providing a large number of free WiFi hotspots which cover downtown. Citizens are encouraged to participate in energy conservation through provision of web access to precise, near real-time energy use data and advice on how they can save energy. The city also has over 100 electric vehicle charging stations to promote the use of hybrid and electric cars.

3.5.3 New York

Figure 14. New York – Hudsons Yard (Artists Impression)

New York is a very densely populated city and it has initiated many smart and sustainable initiatives. City 24/7 [39] is an interactive platform that integrates information from government programs, local businesses and citizens to provide knowledge to anyone, anywhere, anytime, on any device. The city is also building the largest city-wide WiFi network [40] in the US by turning old phone booths into WiFi hotspots. The $20 billion Hudson Yard Project [41] (See Figure 14) is developing a 28-acre commercial and residential area on Manhattan's west side. Hudson Yards will digitally track traffic, energy consumption and air quality. It is being designed to provide the highest quality of life for those living, working, and visiting the area.

3.5.4 Seattle

Seattle is one of the greenest cities in the US. The city's electric meters are being upgraded to give more accurate readings of electricity consumption to allow management. Seattle has partnered with Microsoft to launch its High Performance Building Program [42] that allows real-time tracking of energy efficiency.

3.5.5 San Jose

In partnership with Intel San Jose has initiated a project, "Smart Cities USA", which is tracking real-time data on air quality, noise pollution, and traffic flow [43]. This is Intel's first smart city implementation in the United States with expectations to help the city to grow economically, create 25,000 clean-tech jobs, create environmental sustainability, and enhance life for residents. The data collected is expected to help citizens make decisions, e.g. if the air-quality is bad they may elect to use public transit, a bicycle, or carpool.

3.5.6 Washington, D.C.

The capital is one of the top cities in the U.S. for transit use and e-governance. Washington D.C. has been a pioneer in the adoption of new technology, including the launch of a private cloud in 2010 and the early use of mashups to become a GIS model city [44].

3.5.7 Chicago

Chicago has provided 851 open data sets and the Chicago's Digital Excellence Smart Communities Program [45] is working with local communities to close the digital divide for the elderly and lower-income residents of the city [46].

3.6 Rest of World

Around the world, there are many smart initiatives. It is not possible to list them all here but a number are of interest with linkages going beyond national borders.

3.6.1 IEEE Smart Cities Initiative

At a worldwide level the IEEE has launched the Smart Cities Initiative [47] which spans multiple IEEE Societies. The aim is to provide a trusted voice for the engineering, computing and technology communities around the world in the area of smart cities. Strategic and practical guidance can be obtained on management of essential services, the smooth operation of critical infrastructure and on providing a clean, economic and safe environment for inhabitants to live, work and play. Selected municipalities can join the active community of 10 cities that the IEEE aims to engage in developed and developing countries through 2016.

3.6.2 Brazil

Rio de Janeiro set up a smart city operation centre in 2010, using real data feeds from all data sources including cameras, sensors, actuators and vehicle GPS's. A key focus was on adopting smart technologies for hosting the World Cup in 2014 and the Summer Olympics in 2016.

3.6.3 China

Hong Kong is ranked 13th in the world in terms of the number of Internet of Things firms [48]. An example of smart city development in Hong Kong is the Octopus which is an electronic system that is widely used in public transport, retailing, online payments, parking facilities, self-services, access control systems, recreational facilities and schools, and public services [49]. There is also an EU-China Smart and Green City Cooperation. In this a comparative study has been performed into smart cities in Europe and China [50].

3.6.4 Australia

There are a number of smart city initiatives in Australia [51][52] notably in Adelaide, Brisbane and Melbourne.

3.6.4.1 Adelaide

The Connect Adelaide initiative [53] is providing free high-speed Wi-Fi across public spaces to support real-time transport information, enhanced CCTV coverage, interactive digital maps and live city event broadcasts. This is supported by a number of council smartphone apps, Splash Adelaide, Adelaide Street Eats and Adelaide Report It.

3.6.4.2 Brisbane

CitySmart [54] was created by Brisbane City Council to help make the city the nation's most sustainable city. CitySmart is currently delivering $25 million worth of projects, with a further $290 million in projects in development, all designed to contribute to Brisbane's economic growth and reduce environmental impact. This includes a district cooling energy system to provide cheaper/more efficient air conditioning. Queensland Watt Savers gave more than 300 SMEs easy-to-use tools and expertise to reduce energy consumption and related expenses. EzyGreen, a residential energy reduction program, engaged 61,000 Brisbane households to save over $10 million in annual energy costs.

3.6.4.3 Melbourne

Melbourne has many knowledge industries, a background in technical innovation, an emerging new media sector, and a focus on urban design and the quality of the urban environment [55][56]. The city is developing a number of initiatives based on open data and on providing wired and wireless access to urban spaces.

3.6.5 Japan

In Japan the Ministry of Economy, Trade and Industry has a program underway with companies, such as Panasonic, Hitachi, Toshiba and others, to develop smart city services that can be tested in four domestic pilot cities, and sold internationally. Japanese companies are actively participating in projects in the United States, France, Spain, India and China.

3.6.6 South Korea

The South Korean city of Songdo which is currently being built is expected to be one of the smartest cities on the planet. It will have sensors and cameras at every corner (monitoring temperature, traffic, electricity) that are all interconnected and linked to a central computer that will process information in real-time to optimise the management of the city. The city is attracting startups and companies like Cisco as it is a clean slate for developing technologies. A living lab was set up in November 2015. The lab is called the "IoT (Internet of Things) Cube" and gathers real-time data, and performs analytics [57].

3.6.5 Iceland

In Reykjavik, the city council implemented a Better Reykjavik website [58] where citizens can submit ideas. These are discussed and if there is sufficient political backing they are implemented. Interestingly, 60% of citizens have used the platform with the city spending €1.9 million on more than 200 projects based on ideas from citizens. Since 2008, the Citizens Foundation has used Your Priorities to promote online, democratic debate in Iceland and worldwide. The open source platform is available free of charge to any group, city or country around the world interested in using it to source ideas from citizens. Another prominent use of the platform has been in Estonia where 50,000 citizens used it to submit more than 2,000 proposals to government. Fifteen were taken forward to parliament, and seven have since become Estonian law.

3.7 Needs for Regulation

The very wide scope of smart cities which covers not only interactions with citizens and use of their data but also control of the energy, waste, transportation systems and social interactions with government, education and e-health leads to many areas were regulation may be required. Regulation with respect to energy and transportation systems is covered in the relevant sections in this report. Regulation with respect to gathering and use of data is considered in depth under the topic of Big Data. Safety is of underlying importance to citizens and is considered in the areas of energy, transport and IoT/Cyber-Physical Systems.

In this section, however, it is worthy to note that there is critical overarching need for regulation in the area of privacy and in allowing sharing of data to provide services. Smart cities rely on a myriad of sensors that constantly register and process private data from individuals. Some planned initiatives have failed because citizens have been worried about privacy. Privacy concerns must be handled sensitively, since smart cities rely on gathering large amounts of data that could be compromising or embarrassing if made public [59].

In some cases citizens reject surveillance which is necessary to bring convenience, efficiencies and energy savings to smart cities. There are now efforts under way to establish a legal framework in Europe on gathering data with consent, to reconcile the value of services with privacy, develop ways to use anonymous data and in providing protection through encryption. Full anonymisation is impossible so processing in encrypted domains may be the way forward [60].

Another key barrier to creation of services is the ability to access and utilise data. For data sharing there is a need for openness. This is discussed in detail in the section on Big Data and Open Data.

3.8 Smart City Standards

To help city government and developers provide new services there is a need for standardisation at many levels. This includes in defining reference models, planning guidelines, guides for sustainable development and standards for interoperability of data and communications. Again more detailed information is provided in specific sections on energy, transport, 5G, Big Data and IoT/CPS. In the following section a number of standards which specifically address smart cities are considered.

3.8.1 UK Standardisation Activities

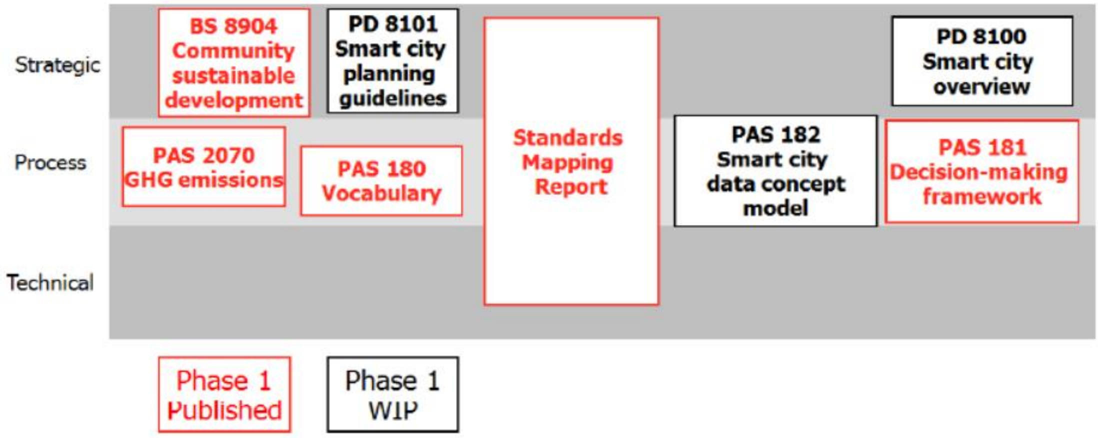


Figure 15. Standards

As an example of all the areas that need to be considered see Figure 15 which shows areas where standards are under development in the UK. The British Standards Institute (BSI) has established a Smart Cities Advisory Group, comprising representatives from cities, business, academia and NGOs. This has identified a number of areas where standards could reduce barriers to smart city implementation. Key aims are to share good practice on development and implementation of new service models, identify common solutions to technical problems and set out the preconditions for interoperability of data. A vision is being set out for the benefits of smart cities and the approaches that can be taken to improve city performance. This provides a smart city overview to communicate the benefits of smart cities to key decision makers. It is also necessary to create common terminology for smart cities to allow sharing of concepts (PAS 180) and provide smart city planning guidelines to set out how major new residential, retail and business developments can support the wider plans of that city to become smarter. Underlying this there is a need for guidelines for economic assessment and funding of smart city initiatives as well as highlighting potential business models and means of procurement. A standard (PAS 181) has also been produced to provide a decision-making framework for smart city leaders, setting out how to deliver a smart city project. Other documents describe data sharing, mapping across different international standards bodies, contributing to ISO standards on sustainable community development and metrics. Standards are being developed for:

- Good practice in provision of digital services, including sharing of open data, protection of privacy and inclusiveness of services
- Evaluating smart city performance, building on the current ISO programme to provide a means of evaluating the effectiveness of smart city products and services
- Procurement of smart city services, building on the initial economic assessment and funding model
- Practical approaches to collaboration between partners in delivery of smart city programmes
- Interoperability of systems, including a framework description of smart city systems building on the mapping work “Preparing the way for smart cities” performed by the BSI [61]

A BSI standards strategy for smart cities was also commissioned by the Department of Business Innovation and Skills to outline UK Department guidelines, metrics, management processes and technical specifications. The smart city framework (SCF) was published in 2014 [62]. This is a practical “how to” guide intended for use by leaders, at all levels and from all sectors, of smart city programmes.

The Institute for Transportation and Development Policy (ITDP) has also developed the Transit-Oriented Development (TOD) Standard, which was published in June 2013. This provides a common understanding of what constitutes urban development best practice from international experience in promoting sustainable urban transport. Key aims are to minimise the use of personal cars and reduce greenhouse gas emissions.

3.8.2 ETSI

Many technical activities of ETSI [63] standards are linked directly to the concept of “smart cities”, e.g. mobility, transportation, M2M, energy efficiency, security, etc. ETSI organised 2 events in 2013 gathering multiple players together involved in Smart Cities. The purpose of these events was to define the driving expectations from an ICT standards organization with regards to smart cities. In October 2013 the ETSI board agreed on a roadmap for Smart Cities to define a High Level Architecture for smart city from the ICT perspective, perform an initial in-house ETSI standards inventory of existing standards that may be applied in a smart city and identify National Standard Organisations outside of ETSI and plan for outreach.

3.8.3 ITU-T

ITU-T has established a new Focus Group on Smart Sustainable Cities to assess the standardisation requirements of cities aiming to boost their social, economic and environmental sustainability through the integration of ICTs in their infrastructures and operations.

3.8.4 China

In China, several national standardisation committees and consortia have started standardisation work on Smart Cities, including the China National IT Standardisation TC (NITS), the China National CT Standardization TC, the China National Intelligent Transportation System Standardization TC, the China National TC on Digital Technique of Intelligent Building and Residence Community of Standardization Administration and the China Strategic Alliance of Smart City Industrial Technology Innovation.

4 Smart Energy

4.1 European Smart Energy Drivers and Policy

The European power grid is highly complex spanning 29 countries over three continents with more than 10 million kilometres of power lines. It transfers over 3,500 TWh of electrical energy annually and this is expected to grow to 4,300 TWh by 2050. A significant proportion of this extra demand will be met by renewable sources which are expected to increase to 50-80% of the total generation capacity from the current value of approximately 14%. The use of electricity is also becoming more concentrated. Presently, 70% of the EU population lives in urban areas, and this figure is likely to increase over the next few decades.

Cities are the main centres for economic, social and cultural activities in Europe and create around 80% of the EU's gross domestic product. However, a consequence of this is that urban areas consume 70% of energy, and account for 75% of the EU's greenhouse gas emissions. Most of the energy consumption occurs within buildings, transportation systems, water supply and treatment, and sewage management. Cities are thus an area where most energy savings could be made and have an important role to play in achieving national and European green-growth strategies. Here innovations are needed in energy, transport and Information and Communications Technology (ICT). There is great interest in making buildings much more energy efficient [64] and a number of strategic priorities have been defined [65].

Notably across the five different priority areas of Europe 2020, environmental issues and green solutions are a key concern for the majority of European Smart City initiatives [66] with nearly 50% of the initiatives addressing environmental problems. This includes energy efficiency in buildings and smarter city transportation options. In Europe many initiatives in energy are transnational involving multiple cities, e.g. Networking Intelligent Cities for Energy Efficiency (NiCE). The NiCE initiative aims to decrease the direct carbon footprint of ICT by 30% per city, contributing to the Europe 2020 energy efficiency and CO2 targets [67].

4.2 European Smart Energy Initiatives

4.2.1 European Energy Research Alliance (EERA)

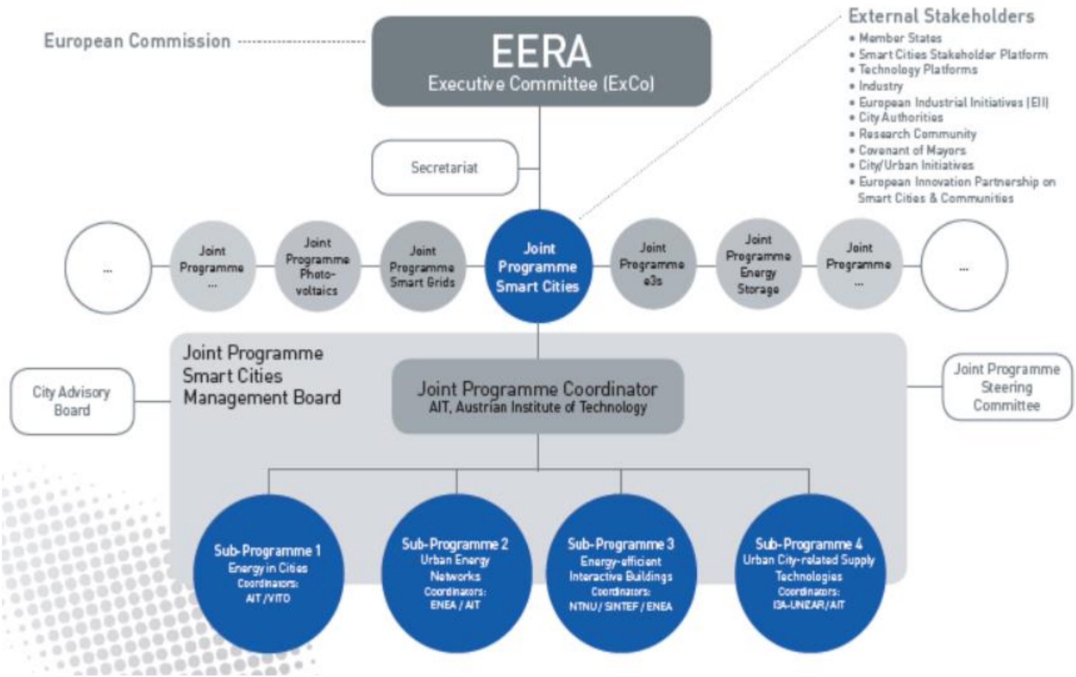


Figure 16. EERA Joint Programme on Smart Cities

The European Strategic Energy Technology Plan (SET Plan) has identified the need for development of energy technologies to combat climate change and the need for securing energy supply at the European and global level in the long term. Successful achievement of Europe's 2020 and 2050 targets for greenhouse gas emissions requires novel technologies designed to increase energy efficiency and enable large-scale integration of renewable energy. To support this, the European Energy Research Alliance (EERA) was set up by leading European research institutes. The aim of EERA is to streamline and coordinate national and European research activities and pool resources to maximise synergies and optimise European energy research capabilities and infrastructures. EERA brings together national facilities in Europe into a virtual centre of excellence and initiates joint research programmes. Currently, 3000 researchers from over 150 public research centres and universities are involved. The topics being addressed include photovoltaics, wind energy, fuel cells, hydrogen, smart grids and smart cities.

The Joint Programme on Smart Cities (See Figure 16), for instance, is developing new scientific methods, concepts and tools to support European cities in their transformation into smart cities. The key focus is on large-scale integration of renewable energy and enhanced energy efficiency using smart energy management at the city level. This requires an integrated systems view as well as innovative, intelligent approaches to the design and operation of urban energy systems. Four key areas are being addressed:

1. **Energy in Cities** is taking an integrated approach towards urban energy planning and transformation processes by understanding the energy performance characteristics and energy flows in urban areas and by providing new planning methods and best practice examples. Decision support and simulation tools will be developed to model economic scenarios and energy flows and thus support experts and authorities in developing urban energy roadmaps.
2. **Urban Energy Networks** concentrates on the intelligent planning, design and operation of thermal and electrical networks in cities through the use of smart energy grids which are able to communicate with each other to balance thermal and electrical loads depending on supply and demand. This requires comprehensive sensor networks feeding energy-related data into a multifunctional ICT platform.
3. **Energy-efficient Interactive Buildings**. Buildings account for around 40 per cent of European primary energy demand. In the work energy-efficient buildings will make use of energy conservation measures and on-site renewables to reduce their energy demand. For this innovative, competitive holistic concepts, tools and demonstration cases are being developed to coordinate the exchange of energy with thermal and electrical grids while providing a comfortable healthy indoor environment to their users.
4. **Urban City-related Supply Technologies** is investigating how on-site renewable supply technologies can be integrated into the urban infrastructure such as heat pumps, solar thermal, photovoltaics, energy storage units, etc.

4.2.2 SETIS (Strategic Energy Technologies Information System)

Figure 17. Strategic Energy Technologies Information System

The SETIS Initiative [68] aims to support cities and regions to taking ambitious and pioneering measures for the sustainable use and production of energy with the target to reduce greenhouse gas emissions by 40% by 2020. In particular, measures on buildings, local energy networks and transport are outlined (See Figure 17).

4.2.3 ICT Roadmap for Energy Efficient Neighbourhoods

Figure 18. Energy Efficient Neighbourhoods

The ICT Roadmap for Energy Efficient Neighbourhoods (IREEN) [69] (See Figure 18), was launched in September 2011 to encourage the development of energy-positive buildings, and eventually whole districts. Coordinated by Manchester City Council the IREEN roadmap was developed over a 2 year period taking input from over 200 experts from the ICT, energy and construction sectors across Europe. A number of recommendations are made. These include the needs for flexible and pay-per-use pricing schemes; "modular and configurable monitoring control systems" to manage neighbourhood energy services, such as LED public lighting; and using chargeable electric vehicles as storage for microgrids. The use of web and mobile applications is also recommended to engage with the local community with provision of e-learning and gamification to raise awareness of energy usage and encourage participation.

4.2.4 KIC InnoEnergy

KIC InnoEnergy is a Knowledge and Innovation Community (KIC) [70] created by the European Institute of Innovation and Technology (EIT). KIC InnoEnergy has been set up as a commercial company dedicated to promoting innovation, entrepreneurship and education in the sustainable energy field. It brings together 30 shareholders including industry, research centres and universities. The targets of KIC InnoEnergy are to reduce costs in the energy value chain, increase security and reduce CO2 and other greenhouse gas emissions.

4.2.5 Positive Energy Blocks

The aim of the Positive Energy Block initiative [71] is to build 100 Positive Energy Blocks (PEB) across the EU. There will be at least one PEB per EU Member State with 50% of these located in smaller cities (maximum 250.000 inhabitants, so called "Small Giants"). A Positive Energy Block is a group of connected buildings in a neighbourhood that produces more energy than it uses on a yearly basis. Here the contributions of heating, cooling, ventilation and lighting are considered. The buildings within a "block" must be a mix of housing, offices and commercial properties to demonstrate typical living environments and a block can also be made up of both new and old buildings (with suitable retrofit).

4.3 National Smart Energy Initiatives

4.3.1 Germany

Within Germany sustainable growth and energy supply has been a major public concern for the past 10 years driven by the federal government's 2010 plan to phase out nuclear power. This will require a fundamental transformation in the way existing infrastructure is operated and also in the exploitation of renewables. Many German municipalities/regions have Renewable Energy self-sufficiency (RESS) goals and the efficient use of energy is also being promoted with support for innovative research technologies and pilot projects.

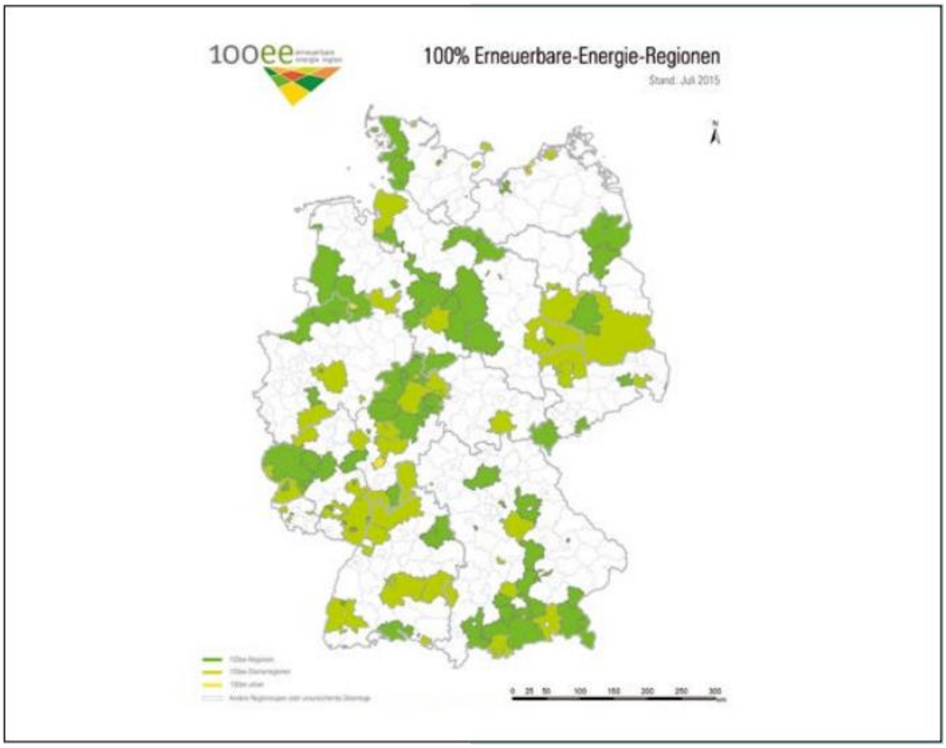


Figure 19. German Cities that are Becoming 100% Self-Sufficient in Renewable Energy

The map in Figure 19 shows the German cities that have been most active in their efforts to become 100% self-sufficient in renewable energy.

4.3.2 France

France derives about 75% of its electricity from nuclear energy, due to a long-standing policy based on energy security introduced after the “oil shock” in 1974 [72]. In mid-2010 the International Energy Agency adopted a strategic policy to make France a provider of low-cost, low-carbon base-load power for the whole of Europe. France is now the world's largest net exporter of electricity due to the very low cost of generation (nearly the lowest in Europe) and earns nearly €3 billion per year from this. Additionally France exports reactors, fuel products and services around the world. Nearly 17% of France's electricity is produced from recycled nuclear fuel. The CO2 emissions per capita from electricity generation is extremely low since over 90% of electricity is from nuclear or hydro sources. In 2013 French electricity prices for medium-size industrials were about 90% that of the EU-27 average, and those for medium-size households (at less than 8 c/kWh) were less than half of the EU-27 average. The national energy regulator (CRE) sets the price at which EdF’s electricity is sold to competing distributors.

President Francois Hollande called for a national debate on energy in 2013. 170,000 people took part in 1000 regional debates and 1200 submissions were received over the Internet. As a result in October 2014 an “Energy Transition for Green Growth” bill was passed by the National Assembly which aims to reduce the share of nuclear energy to 50% by 2025. The bill also sets long-term targets to reduce greenhouse gas emissions by 40% by 2030 compared with 1990 levels and by 75% by 2050. The bill also calls for energy consumption to be halved by 2050 compared with 2012 levels with reductions in fossil fuel consumption of 30% by 2030 and an increase of renewables to 32% by 2030. The bill also sets long-term targets for France's carbon tax: €14.50 per tonne CO2 currently, to €22 in 2016, then to €56 in 2020, rising to €100 in 2030. Areva, EdF and CEA announced the formation of the tripartite French Nuclear Platform (PFN) to improve the joint effectiveness of the three bodies and produce a shared vision of a medium- and long-term goal for the industry. In 2015 RTE started work on a new 1200 MWe HVDC connection to Turin in Italy supplying 10.5 TWh, costing about €1 billion which will be the longest subterranean high-voltage power line when it goes into service in 2019.

4.3.2.1 SmartGrids France

In 2012, nine French business and research clusters teamed up to create SmartGrids France. This aims to exchange experience to create a uniform grid architecture and standards across French SmartGrid initiatives. Several pilot projects have already been deployed in France, e.g. via funding programs managed by the ADEME (Environment and Energy Management Agency) and the ANR (National Research Agency). For example, in southern France, the Nice Grid project, supported by state-owned ERDF, has tested the integration of solar PV and energy storage with the national grid using networked smart meters to monitor usage. The project aims to develop a smart grid that integrates local produced photovoltaic electricity with a 1 MW battery energy storage system supplied by Alstom and Saft, and smart meters in residents' homes. In Toulouse, the Sogrid project will trial a next generation integrated chip to enable the transmission of digital information on the grid. One of the aims of the project is to develop a new international standard. The pilot forms part of a larger vision to create a smart electricity network that can integrate electric vehicles, renewable energy and provide demand management [73].

4.3.3 UK

4.3.3.1 The Energy Technology Institute (ETI) (UK)

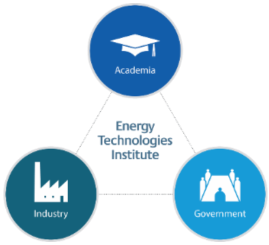


Figure 20. Energy Technology Institute

The ETI (See Figure 20) is a public private research partnership, involving six companies whose funding of projects is matched by the public sector, in particular EPSRC. Its Smart Systems and Heat (SSH) Programme aims to design and test a commercially viable Smart Energy System in the UK, facilitating improved heat management and low carbon energy services across the country. This involves the investigation of mass-market consumer behaviour and requirements to understand the likely future demand for heat and energy usage. The design methodology phase of the programme is under way, and work has begun on Enabling Component Technologies (identifying gaps in the potential range of smart systems technologies), Energy System Design Tool Development (assessing the impact of a Smart Energy System in a geographical area), Data Management and Architecture (fulfilling information and service requirements of a smart energy system), Value Management and Delivery (identifying how value can be delivered across the value chain) and a Consumer Behaviour Study (providing insight into consumer requirements for heat and energy both now and in the future). ETI has also started engagement work with local authorities, who will provide the demonstration locations for the second phase of the programme: a mass-market field trial of up to 10,000 homes to ensure any system design can be replicated geographically across the UK.

4.3.3.2 Low Carbon Economy

Work in the UK on energy is led by the low carbon economy. The Department of Energy and Climate Change (DECC) has defined the ambitious target of an 80% cut in greenhouse gas emissions by 2050. Currently the UK spends £32 billion a year on heating which accounts for around a third of UK's greenhouse gas emissions [74]. The UK is on track to meet its first carbon budgets and also the EU target for renewable energy. A Green Deal programme has been introduced to drive energy efficiency across the country. Every home will be supplied with a smart meter helping consumers to understand their energy consumption and make savings. At the same time this should reduce supplier costs, enable new services and facilitate demand-side management. Smart metering is seen as a key enabler of the future Smart Grid, as well as facilitating the deployment of renewables and electric vehicles. An Impact Assessment has been performed by DECC on the deployment of smart electricity and gas meters in domestic premises and in smaller non-domestic premises in Great Britain [75].

4.3.3.3 Low Carbon Pioneer Cities Heat Networks Initiative

The Low Carbon Pioneer Cities Heat Networks project began in March 2013 supporting five of England's core cities to move towards the deployment of low carbon heat networks, Leeds City Region, Greater Manchester (city region), Newcastle, Nottingham and Sheffield. Heat networks can be powered using a variety of fuels, including lower carbon sources such as biomass and energy from waste. Feasibility work was supported to investigate the potential for heat networks in each area with help from the Department of Energy and Climate Change (DECC). This led to the Heat Networks Delivery Unit (HNDU) scheme [76].

4.3.3.4 Low Carbon Innovation Coordination Group (LCICG)

The LCICG brings together the major public sector backed organisations that are supporting low carbon innovation in the UK. The LCICG aims to maximise the impact of UK public sector funding for low carbon technologies, in order to deliver affordable, secure, low carbon energy for the UK, encourage UK economic growth and increase the UK's capabilities, knowledge and skills. The LCICG have developed a Strategic Framework – Coordinating Low Carbon Technology Innovation [77]. The Strategic Framework sets out the LCICG's planned approach to collaboration and the prioritisation of future innovation support programmes. Building on the Technology Innovation Needs Assessment project (TINA) it highlights the key innovation needs up until 2020. LCICG have also highlighted the importance of energy storage systems to enable electricity generated at a time of low demand to be stored and used at a later time when electricity demand is high. This is being supported by a funding programme the DECC Energy Storage Innovation Competition [78].

4.3.3.5 Low Carbon Networks Fund (LCNF)

Ofgem [79] has made £500 million available to network operators over 5 years (2010-2015) to trial new technologies and approaches for efficiently connecting renewable generation, meeting the needs of small-scale and intermittent generation, addressing an increase in the use of electric vehicles, heat pumps, smart domestic appliances and other low-carbon technologies, using smart meter data to improve network performance and reduce costs and for incentivising customers to reduce their carbon footprint and cut bills, by managing their energy demand.

4.4 US Smart Energy Drivers and Policy

In the US there are a number of government initiatives and policies including investment grants, totalling $3.4 billion, dedicated to Smart Grid projects [80]. This includes funding to promote energy-saving choices for consumers, increasing efficiency, and fostering the growth of renewable energy sources such as wind and solar. The grants were part of the Reinvestment and Recovery Act that was a response to the global economic crisis. The grants follow an industry matching model, meaning that every private investment made is matched by federal grants. Development is driven by private companies. A key challenge in the US is that the grid infrastructure is largely outdated some parts being over 100 years old.

Category	$ Million
Integrated/Crosscutting	2,150
AMI	818
Distribution	254
Transmission	148
Customer Systems	32
Manufacturing	26
Total	3,429

18 million smart meters
1.2 million in-home display units
206,000 smart transformers
177,000 load control devices
170,000 smart thermostats
877 networked phasor measurement units
671 automated substations
100 PEV charging stations

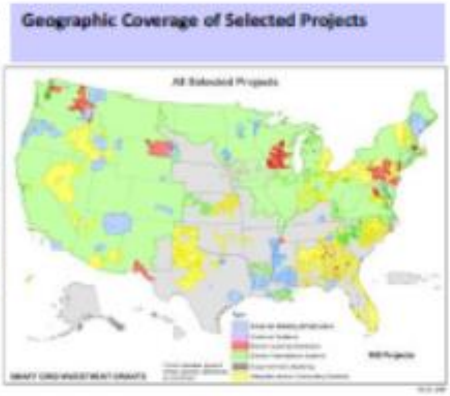


Figure 21. US Energy Programmes

Congress and the Administration have outlined a vision for a future Smart Grid and have also put forward policy to enable it to be built. The Energy Independence and Security Act of 2007 (EISA) [81] made it the policy of the United States to modernize the nation's electricity transmission and distribution system to create a smart electric grid. This coupled with the "The American Recovery and Reinvestment Act of 2009 (ARRA)" [82] is accelerating the development of Smart Grid technologies with $4.5 billion investment for electricity delivery and energy reliability activities (See Figure 21). This includes programmes to modernize the electric grid and implement demonstration programmes. This is supported by President Obama who has put forward a vision of a clean energy economy in a State of the Union Address and the Administration's commitment in the "Blueprint for a Secure Energy Future." The White House also released a report in 2011 by the Cabinet level National Science and Technology Council (NSTC) highlighting "A Policy Framework for the 21st Century Grid: Enabling Our Secure Energy Future."

4.4.1 Blueprint for a Secure Energy Future

The White House, "Blueprint for a Secure Energy Future" [83] outlines a three-part strategy:

- Develop and Secure America's Energy Supplies: We need to deploy American assets, innovation, and technology so that we can safely and responsibly develop more energy here at home and be a leader in the global energy economy.
- Provide Consumers With Choices to Reduce Costs and Save Energy: Volatile gasoline prices reinforce the need for innovation that will make it easier and more affordable for consumers to buy more advanced and fuel-efficient vehicles, use alternative means of transportation, weatherise their homes and workplaces, and in doing so, save money and protect the environment. These measures help families' pocketbooks, reduce our dependence on finite energy sources and help create jobs here in the United States.
- Innovate our Way to a Clean Energy Future: Leading the world in clean energy is critical to strengthening the American economy and winning the future. We can get there by creating markets for innovative clean technologies that are ready to deploy, and by funding cutting edge research to

produce the next generation of technologies. And as new, better, and more efficient technologies hit the market, the Federal government needs to put words into action and lead by example.

Standards for the Smart Grid have been identified as being critical in the EISA and 2011 NSTC report. Standards are needed to make sure that investment remains valuable in the future and help with innovation, highlight best practice, support consumer choice and open global markets for smart grid technologies and create economies of scale to reduce cost.

4.4.2 A Policy Framework for the 21st Century Grid: Enabling Our Secure Energy Future

The report “A Policy Framework for the 21st Century Grid: Enabling Our Secure Energy Future” [84] outlines policy recommendations that build upon the Energy Independence and Security Act of 2007 and the Obama Administration's smart grid investments. The report was prepared by the Subcommittee on Smart Grid of the National Science and Technology Council, Committee on Technology. It provides a policy framework that promotes cost-effective investment, fosters innovation to spur the development of new products and services, empowers consumers to make informed decisions with better energy information, and secures the grid against cyber-attacks. The framework has four pillars:

1. Enabling cost-effective smart grid investments
2. Unlocking the potential for innovation in the electric sector
3. Empowering consumers and enabling them to make informed decisions, and
4. Securing the grid.

Each pillar supports a set of policy recommendations that focus on how to facilitate a smarter and more secure grid [85].

4.5 US Smart Energy Initiatives

4.5.1 Department of Energy Smart Grid

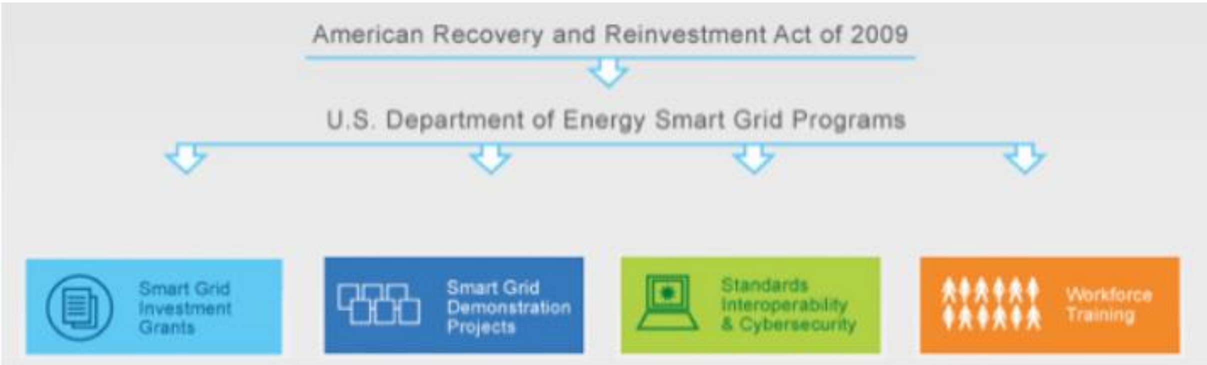


Figure 22. DoE Smart Grid

The American Recovery and Reinvestment Act of 2009 (Recovery Act) provided the U.S. Department of Energy with $4.5 billion to modernise the electric power grid and to implement Title XIII of the Energy Independence and Security Act (EISA) of 2007. The two largest initiatives within the US are the Smart Grid Investment Grant (SGIG) Program and the Smart Grid Demonstration Program (SGDP) (See Figure 22 and Figure 23). The DOE Office of Electricity Delivery and Energy Reliability (OE) is responsible for managing these five-year programs [85].

Figure 23. Smart Grid Initiatives

The DOE's Office of Electricity Delivery and Energy Reliability holds regional stakeholder meetings, to stimulate peer-to-peer dialogue on smart grid deployments, share lessons learned, and help replicate successes. The DOE is expanding cooperative relationships with the National Association of Regulatory Utility Commissioners and the National Association of State Utility Consumer Advocates, to provide technical assistance and share information on consumer empowerment from Recovery Act projects. A Smart Grid Innovation Hub has also been set up. Other initiatives include grid research and design investments by the Advanced Research Projects Agency–Energy (ARPA-E), a Home Energy Education Challenge, consumer behaviour studies funded by the

Recovery Act; and investment in smart grid technologies by the Department of Agriculture's Rural Utilities Service (RUS) [86]. The Department of Energy Office of Electricity Delivery and Energy Reliability is also launching a new Smart Grid Integration Challenge for Cities in 2016 to recognise US cities as "smart city leaders" in implementing sensing, data sharing, and data analytics to achieve energy consumption reduction targets set by individual cities. Each winning city will serve as a model for replication in other cities.

4.6 Smart Energy Initiatives in the Rest of the World

A worldwide study by DNV KEMA Energy & Sustainability [87] commissioned by the Netbeheer Netherlands made a global inventory of smart grid initiatives. This highlights lessons learned which can be used in the Netherlands. For instance, in America there is a strong focus on peak load reduction technology and dynamic pricing tariff pilots, whilst in Europe more emphasis is placed on improving energy efficiency and reducing emissions through the use of more decentralised means of production. In the Asia-Pacific region drivers vary country to country – from modernizing and improving grid reliability in China, to techniques for load management in Australia and New Zealand.

Notably there is a wide variety of technologies and services being demonstrated. This is affected by business drivers that are different in different regions and countries. Generally, smart meters and IT systems are being applied in most demonstration projects and demand response is being applied as a smart grid service to reduce peak loads rather than to 'fill' load gaps. To improve integration with renewable energy sources and apply smart devices and electric vehicles agent-based algorithms are being demonstrated. These algorithms are also being used for congestion management, critical operations, and load shedding. The PowerMatcher concept is being applied in several European countries and projects. Other decentralised controllers are being demonstrated for Virtual Power Plants (VPPs) by distribution network operators. For flexibility many demonstration projects also include energy storage. In some cases thermal storage devices have given storage capacity up to weeks.

Information Technology is a crucial part of a smart grid but introduces risk of communication failure and unwanted system errors. Interoperability and clear interface specifications are needed to allow integration. Other lessons learned are that there is a need for clear scoping and cost-benefit analyses prior to project commencement to prevent large project overruns. There is also a need to manage expectations to provide promised benefits to energy consumers and stakeholders. To deal with technical and financial risks it is important to perform demonstration projects at the correct scale.

Other countries investing heavily in smart grid infrastructure are Canada, Mexico, Brazil, many of the member states of the EU, Japan, South Korea, Australia, India, and China.

4.6.1 IEEE Smart GRID

The IEEE Smart Grid Initiative [88] brings together IEEE's broad array of technical societies and organizations to encourage the successful rollout of technologically advanced, environment-friendly and secure smart-grid networks around the world.

4.6.2 China

China is committed to reducing its carbon intensity for its GDP by 40 to 45 percent by 2020 relative to 2005. Here there is investment in renewable power transforming the nation's energy landscape [89]. Equipment makers, communication device players, and integrated solutions providers from around the world see this as a major commercial opportunity. The Chinese government has announced plans to construct power grids in the north-west province of Xinjiang to allow interconnection with the country's eastern provinces, Pakistan and other Asian countries. An Ultra High Voltage grid will be constructed to transmit power to the eastern coastal developed areas.

The State Grid Corporation of China has allocated US$31 billion to the Xinjiang Electric Power Company to perform grid integration of renewable energy and installation of power transmission lines by 2020. China also plans to construct a Global energy network by 2050 to meet global power demand using smart energy technology. This was announced by the Chinese President, Xi Jinping, to the UN Sustainable Development Summit. The aim is to create grid interconnectivity with neighbouring countries such as Russia, Mongolia, Kazakhstan, Pakistan, Myanmar, Laos, Nepal and Thailand. Since 1992 China has relied heavily on electricity it purchases from Russia.

China's annual predicted investment in smart grid development and related infrastructure from 2016 to 2030 is estimated to reach $128 billion and the Chinese National Development and Reform Commission has committed to producing 20% of the country's energy needs from renewable and nuclear energy by 2030 [90].

4.6.3 Canada

In Canada, the state government of Ontario has a Smart Grid Fund [91]. The Smart Grid Fund supports projects that test, develop and bring to market the next generation of energy grid solutions. The aim is to help consumers and businesses manage energy costs, improve conservation efforts, and integrate new beneficial technologies like electric vehicles and storage.

4.6.4 Japan

The failure of the Daiichi nuclear power plant in Fukushima following the Great East earthquake in 2011 prompted the government to reform the power system to ensure a future stable supply [92]. Market liberalisation is also being introduced to provide lower costs to customers. This was launched in 2016 and will be complete by 2020. A driver to install smart grid infrastructure within the next five years is the Tokyo Olympic Games in 2020. The country's largest utility Tokyo Electric Power Corporation (TEPCO) has stated that it will have 27 million smart meters in the field by 2020.

4.6.5 South Korea

The South Korean government has imposed emission cuts on power plants. Here investment is being provided for clean and renewable power sources. Smart grids are seen as the way to integrated renewable sources so there is investment in smart meters, software and hardware, smart T&D equipment, communication and wireless network infrastructure and sensors. The market for smart grid equipment in South Korea is expected to grow by more than 20% during the five years to 2019 and home-grown smart grid equipment is being developed to support this with also a view to the export market [93]. As an example In June 2015, Kepco announced a $9.1 million contract to create a microgrid in Canada.

4.6.6 Brazil

Brazil suffers from frequent power outages [94]. This is driving interest in smart grid technology and the Mining and Energy Ministry is pushing for diversification in the energy mix to reduce reliance on hydro-power. Brazil's smart grid market is expected to grow by 21.17% over the period 2014-2019, according to market research company Tech Navio.

4.6.7 India

India's government has committed billions of dollars for smart grid infrastructure with cumulative spending forecasted to be US$21 billion over the period 2015-2025 [95]. They have also created a national smart grid organisation that will invest $210 million into planning, monitoring and implementing grid modernization up until 2017. There is a need in India to stop rampant electricity theft which is estimated to cost $16.2 billion a year.

4.6.8 Australia

In 2009 the Australian Government recognised the importance of investing in commercial-scale trials of promising smart grid technologies. $100 million was put into the Smart Grid, Smart City Program with further funding of $390 million by other contributors. The Program ran from 2010 to 2013 and determined the benefits of smart grid deployment [96].

4.6.9 Mexico

Mexico plans a 30.2 million smart meter deployment between 2015-2025 [97]. Mexico is the second largest consumer of smart grid technology behind Brazil in the Latin America. Regulatory momentum for smart grid infrastructure investment is being driven by a smart grid roadmap developed by the regulator Comisión Reguladora de Energia (CRE) which has a mandate to modernise the country’s power grid. A $10.9 billion smart grid infrastructure investment is planned with deployments across a number of market segments including smart metering, distribution automation, battery storage, home energy management, information technology and wide area measurement. This is supported by the PIDIREGAS program, which uses private company finance to carry out public works projects.

4.7 Needs for Regulation

The smart grid market is led by regulation. In many countries reductions in emissions, consumer choice and energy security are driving adoption of smart grid technologies. Grid regulation varies considerably per country, particularly across Europe which has a big impact on possibilities for smart grid investments. Advantages of innovation funding is therefore difficult to measure. As an example although Virtual Power Plants (VPPs) have been shown to have a high socio-economic value within the energy market it is difficult to copy results from field trials in one country to another country due to different regulations and ancillary services.

A key factor that influences consumer benefit is pricing structure used which can increase the value of smart grids (and smart appliances). Experience shows that tiered pricing rates are more effective at reducing load than a time-of-use pricing (off-peak/on-peak) schemes [98]. Tiered rate structures allow prices to reflect system capacity with super-peak price levels being applied when large load reduction is needed. More dynamic price levels are required but this has to be balanced against the increase in uncertainty for customers. Several pilots in the US, has shown that the use of “Critical Peak” prices is an effective technique to trigger load reduction. The use of rebates was found to be less effective indicating that "punishments" are more effective than "rewards".

4.7.1 European Regulation on Smart Grids

In order to get investors to commit there is a need for a regulatory framework for Smart Grids. Within Europe the Electricity Directive and the Energy Services Directive provide a mix of obligations and incentives to Member States to establish such a framework [99]. Regulation is being targeted to encourage network operators to earn revenue not from additional sales but from efficiency gains and lower peak investment needs. There are 3 key groups as outlined below:

4.7.1.1 Smart Grids Task Force

The Smart Grids Task Force [100] was set up by the European Commission in 2009 to advise on issues related to smart grid deployment and development. There are five Expert Groups who focus on specific areas and shape EU smart grid policies.

- Expert Group 1 – Smart grid standards. In 2015 this expert group performed an investigation on the interoperability, standards and functionalities applied in the large scale roll-out of smart metering in EU Member States.
- Expert Group 2 – This group addresses regulatory recommendations for privacy, data protection and cyber-security in the smart grid environment
- Expert Group 3 – This group gives regulatory recommendations for smart grid deployment
- Expert Group 4 – This group considered smart grid infrastructure deployment
- Expert Group 5 – This group considers implementation of smart grid industrial policy

4.7.1.2 Agency for the Cooperation of Energy Regulators (ACER)

The Agency for the Cooperation of Energy Regulators (ACER) [101] is a European Union Agency, based in Ljubljana, Slovenia that was set up to progress the completion of the internal electricity and gas energy markets. It is an independent body that fosters cooperation among European energy regulators to ensure market integration and the harmonisation of regulatory frameworks within the framework of the EU's energy policy objectives.

4.7.1.3 CEER

The Council of European Energy Regulators (CEER) [102] is the voice of Europe's national regulators for electricity and gas at the European and international level. CEER, is a non-profit association that allows national regulators to cooperate and exchange best practice, creating a single, competitive, efficient and sustainable EU internal energy market that works in the public interest.

4.7.1.4 Regulations in UK

Smart grids are being driven in the UK by energy and climate goals. Time varying electricity tariffs are Smart grid policy in the UK**.** Building a smarter grid is an incremental process of applying information and communications technologies to the electricity system, enabling more dynamic real time flows of information on the network and more interaction between suppliers and consumers. DECC published a vision document in December 2009 [103], and the Electricity Networks Strategy Group (ENSG) published a Smart Grid Routemap [104]. Smart electricity and gas meters are being rolled out to all UK homes by 2020 and Ofgem is providing £500 million through the Low Carbon Networks Fund [105] to support smart grid trials. A further £2.8 million has been given to 8 smaller smart grid demonstration projects. A framework for smart grid standards, focused on cyber-security issues has also been put in place [106].

In the UK the Smart Grid Forum was created by the Department of Energy and Climate Change (DECC) and the industry regulator, Ofgem, to bring together representatives from electricity network companies, consumer groups, energy suppliers and wider industry. The aim of the forum is to address technical, commercial and regulatory issues associated with developing smart grids in order to support the UK's transition to a secure, safe, low carbon, affordable energy system. The forum discusses new issues and questions that arise from smart grids with respect to the existing regulatory and commercial framework. Specific challenges are the increase in complexity, the need to balance supply and demand at local levels, the introduction of intelligent control networks and utilisation of energy storage. Investment is needed and existing regulatory and commercial frameworks need to evolve to incentivise this. The Ofgem Low Carbon Network Fund and DECC/Ofgem-chaired Smart Grid Forum have initiated some work, however, there is a need to remove regulatory barriers to storage and demand side response, deliver clearer price signals to allow more flexibility from consumers, e.g. time of use tariffs, and to catalyse innovation, so that new solutions can emerge [107]. The government is also assessing whether more fundamental changes are required to deliver a future smart system, including in the operation of the market and existing institutional arrangements [108].

4.7.1.5 Regulations in Germany

Regulation within Germany is leading to new roles, e.g. metering point operator and metering service provider, with new contracts and pricing models. Protection clauses are needed as well as changes in the regulation to allow for Balance Responsible Parties to take advantage of offering flexibility to consumers.

4.8 Smart Grid Standards

4.8.1 Smart Grid Standards in Europe

Within Europe standards in the sector are voluntary and are developed by industry and market actors following principles such as consensus, openness, transparency and non-discrimination. In the smart grid area there is a need for standards for interoperability and safety. The aim of standards is to reduce costs and to allow different vendors to integrate their equipment together. For the electricity sector standards are set by three European Standards Organizations (ESOs):

- the European Committee for Standardisation, CEN
- the European Committee for Electrotechnical Standardisation, CENELEC, and
- the European Telecommunications Standards Institute, ETSI and can be used to support EU legislation and policies.

The European Commission and EFTA issued the Smart Grid Mandate M/490 [109] in March 2011. This was accepted by the, CEN, CENELEC and ETSI in June 2011. The mandate highlights the need for speedy action, the need to accommodate a huge number of stakeholders and to work at an international level to develop a framework to enable the three ESOs to perform continuous standard enhancement and development in the smart grid field [110] [111].

4.8.2 Smart Grid Standards in US

Recognizing that standards play a critical role in the development of smart grids, EISA called for NIST and FERC to facilitate the development and adoption of interoperability standards (U.S. Congress 2007) [84]. The NIST led smart grid interoperability process, aims to develop flexible, uniform, and technology neutral standards that can enable innovation, improve consumer choice, and yield economies of scale.

4.8.2.1 NIST

NIST has been given the primary responsibility to coordinate development of a framework that includes protocols and model standards for information management to achieve interoperability of Smart Grid devices and systems. NIST has developed a three-phase plan to accelerate the identification and consensus on Smart Grid standards, establish a robust Smart Grid Interoperability Panel (SGIP) that sustains the development of the many additional standards that will be needed, and to create a conformity testing and certification infrastructure. The progress made in Phases II and III of the plan are described in [112]. This report also presents a reference model, standards, gaps, and action plans. The processes established by the SGIP are now guiding standardization efforts across more than 20 standards-setting organizations. NIST’s standards for the Smart Grid [113] are targeted at industry utilities, vendors, academia, regulators, integrators and developers, and other Smart Grid stakeholders.

NIST and the International Trade Administration (ITA) have partnered with the Department of Energy to establish the International Smart Grid Action Network (ISGAN). This is a multi-national collaboration of 23 countries and the European Union. ISGAN is designed to complement the Global Smart Grid Federation which is a global stakeholder organization which serves as an "association of associations" to bring together leaders from Smart Grid stakeholder organizations around the world.

4.9 Cyber-Security

The deployment of networks of computers, intelligent electronic devices, software, and communication technologies presents greater infrastructure protection challenges than those of the traditional infrastructure. A smarter grid includes more devices and connections that lead to vulnerabilities which can be used for intrusions, error-caused disruptions, malicious attacks, destruction, and other threats. Smart grid technologies thus need open standards, specifications and requirements to assure interoperability but there is also need to protect consumer privacy and provide security to ensure resilience. As the electric grid network is key to the operation of a country cyber-security is a key topic on both sides of the Atlantic.

4.9.1 European Initiatives on Cyber-Security

Cyber-security is a highly important requirement for future smart grids. ICT infrastructures are an underpinning platform for critical infrastructure across Europe without which some services (e.g. in electricity transmission and distribution) could come to an abrupt halt. To tackle this challenge in 2010 the European Commission convened a multi-stakeholder and multidisciplinary group of experts to discuss and work on relevant matters regarding the security and resilience of communication networks and information systems for Smart Grids [114].

There is a drive to provide a high common level of network and information security across Europe. As a large system of distributed and interconnected systems, the smart grid offers an exceptionally large attack surface. Every asset of the smart grid (i.e., home gateways, smart meters, substations, control room) presents a potential target for a cyber-attack. A key concern is that an attack on a critical node may jeopardise grid security and lead a cascade effect and whole system blackout. The smart grid cyber-security challenge is thus to protect the ever-growing number of smart grid assets and their communication channels from fast-growing and continuously evolving cyber threats.

4.9.1.1 Cyber-Security Standards

Standards to develop smart grid cyber-security are already available, however, enhancements are needed to reflect the evolution of the smart grid, its technologies, and threats. A key challenge is to maintain these standards over time at an appropriate pace which requires considerable effort. In Europe Alstom Grid, Intel, and McAfee have joined their expertise to deliver a view on smart grid cyber-security in a white paper [115].

This initiative highlights challenges in migrating to the modern grid and different approaches to building it while addressing cyber-security risks.

4.9.2 US Initiatives on Cyber-Security

The U.S. White House has also expressed concern about cyber-security and protecting critical infrastructures. Cyber-security needs to address threats that are unique to electric grid technology which include long life expectancy of energy control systems, low-latency communications needed for real-time control, and differing requirements and regulatory frameworks among grid stakeholders. Although much has been done already in terms of processes for sharing information about risks and threats to the electricity sector the Administration has proposed specific cyber-security legislation. The Federal Government is seeking to ensure that grid operators have access to actionable threat information and provide support for research, development, and demonstration of cyber-security systems. Stakeholders (government agencies, industry, and utilities) are being encouraged to cooperate to identify and address cyber risks and open standards and guidelines for cyber-security will be developed through public-private cooperation. The aim is to identify and prioritise relevant cyber risks—including malware, compromised devices, insider threats, and hijacked systems—and develop standards and guidelines that enable the design of effective plans for mitigating those risks.

The overall goal is to develop policy and regulatory frameworks that ensure that effective and feasible security is appropriately implemented and that all stakeholders contribute to the security and reliability of the grid. A problem at present is that stakeholders have varying levels of awareness and understanding of current threats and specific vulnerabilities. It is necessary for Federal and industry partners to provide timely and actionable cyber threat and vulnerability information to state regulators, industry participants, and electric utilities. Already a number of threat warning bodies exist, e.g. Electricity Sector—Information Sharing and Analysis Center, the United States Computer Emergency Readiness Team, and the National Electric Sector Cyber-security Organization [116].

4.9.2.1 US Cyber-security Standard for Smart Grid Systems

Smart grid cyber-security must address not only deliberate attacks (from disgruntled employees, industrial espionage, and terrorists), but also inadvertent compromises due to user error, equipment failures and natural disasters. The NIST Information Technology Laboratory (ITL), Computer Security Division leads the Smart Grid Interoperability Panel (SGIP) Cyber-security Committee (SGCC), which was set up in response the Energy Independence and Security Act of 2007 to address cyber-security in the areas of Advanced Metering Infrastructure (AMI), cloud computing, supply chain, and privacy recommendations related to emerging standards. The primary goal of the SGCC is to develop a cyber-security risk management strategy for the Smart Grid to enable secure interoperability of solutions across different domains and components. It provides foundational cyber-security guidance, reviews of standards and requirements, outreach, and fosters collaborations in cyber-security within the smart grid. A key outcome from this are the NISTIR 7628 Guidelines for Cyber-Security (Volumes 1, 2, and 3). This publication is widely used by utilities, vendors, and regulators, and is also cited internationally. A first draft of (revised) NISTIR 7628 Guidelines for Smart Grid Cyber-Security, Revision 1 for SGCC review and comment has been published for comment [117].

The DOE/OE Cyber-security for Energy Delivery Systems (CEDS) programme is also working toward resilient energy delivery systems that are able to survive a cyber incident.

5 Smart Transportation

5.1 European Smart Transportation Drivers and Policy Activities

Transport is an essential component of the European economy accounting for about 7 % of GDP and for over 5 % of total employment in the EU.

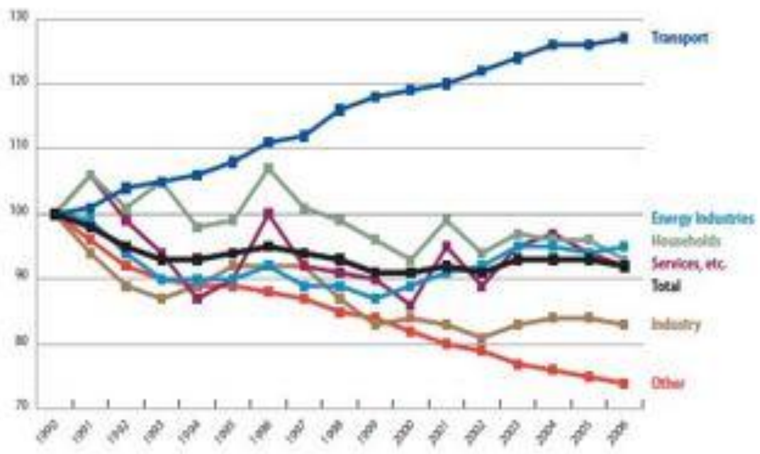


Figure 24. Contributions to CO2 Emissions

Sustainability is a key issue in Europe and there has been a dramatic increase in both freight (35%) and passenger transport (20%) between 1995 and 2006 and this is set to rise. Along with this increase in traffic there has been an increase in emissions. According to data from the European Environment Agency [119], transport accounted for close to a quarter (23.8%) of total greenhouse gas emissions and slightly more than a quarter (27.9%) of total CO_2 emissions in the EU-27 in 2006. No other sector has the growth rate of greenhouse gas emissions as high as in transport (See Figure 24). As the transport sector relies on fossil fuels for 97% of its needs, the fight against climate change in this sector is also synergistic with energy security of supply. The expectation is that traffic will increase further in all sectors and the infrastructure needs to support continuing increase in demand. Freight transport follows trade activity and in recent years this has grown more than GDP. Passenger transport, excepting aviation, has undergone a less dramatic rise. These trends can only be sustained, however, if transport radically improves its energy efficiency and reduces its greenhouse gas emissions.

Linked with this is a drive to improve safety. Europe's roads have become safer in recent years: the number of road accidents involving personal injury fell by 12% between 1991 and 2007. More importantly, the number of road fatalities dropped by more than 44% over the same period. However, with still over 30 000 deaths in the EU in 2011, transport by road is still costly in terms of human lives. There is an objective in the 2001 European White Paper [120] to halve casualties with respect to 2001 levels in road transport. The increased use of ICT is seen as the answer to many problems to allow better scheduling of traffic flow to reduce congestion, enabling increased communication with infrastructure and between vehicles to reduce congestion and avoid accidents, and as a central element in introduction of more autonomy within systems to improve efficiency and increase safety.

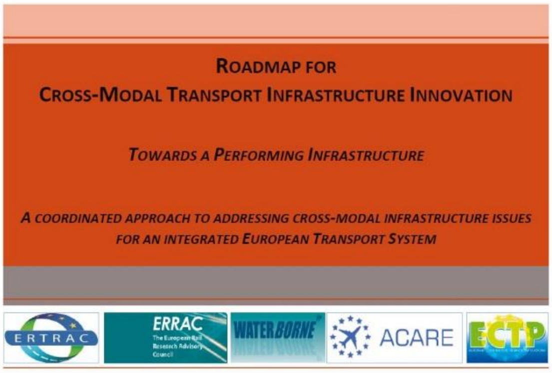
In the maritime sector, marine pollution and maritime accidents were considerably reduced and the EU has established one of the most advanced regulatory frameworks for safety and for pollution prevention (most recently with the third maritime safety package). In aviation, a comprehensive set of common, uniform and mandatory legislation has been adopted covering all the key elements affecting safety (aircraft, maintenance, airports, air traffic management systems, etc.). Safety agencies have been set up for aviation (EASA), maritime affairs (EMSA) and rail transport (ERA).

5.1.1 European Transport Technology Platforms

The European Commission supports a number of transport Technology Platforms that includes ERRAC [121], ERTRAC [122], ACARE [123] and WATERBORNE [124]. In 2001 the Commission issued a White Paper [120] setting a 10 year agenda for the European transport policy which was updated in the mid-term review of 2006 [125]. This highlighted that transport is a complex Systems of Systems that depend on multiple factors, including the pattern of human settlements, the organisation of production and the availability of infrastructure. Although the European transport system compares well in terms of efficiency and effectiveness with most advanced regions of the world, it is still not on a sustainable path. The open markets in Europe have led to more efficiency and lower costs which can be particularly seen in air transport, however, in other transportation areas there is a need to harmonise differences in taxation and subsidies.

5.1.2 TEN-T Guidelines

To coordinate the planning of infrastructure projects across Europe the Trans-European transport networks (TEN-T) policy has provided many benefits with an investment programme of EUR 400 billion. The TEN-T Guidelines [126] are the European Community’s instrument for policy definition and network planning. Adopted in 1996 and amended in 2004, the guidelines include two planning layers: a comprehensive network layer including outline plans for rail, road, inland waterway, combined transport, airport and port networks and a second layer of 30 priority projects. The TENs have already gone a long way in linking EU markets and people. Progress has been achieved in reducing air pollution and road accidents. Air quality in European cities has significantly improved through the application of stricter Euro emission standards addressing fine particles (PM10) which are particularly damaging for human health. The guidelines are also addressing expansion of transport infrastructure which result in habitat loss and landscape fragmentation. To co-fund identified TEN-T projects in the EU Member States a Connecting Europe Facility (CEF) Fund has been set up with a budget of €24.05 billion.



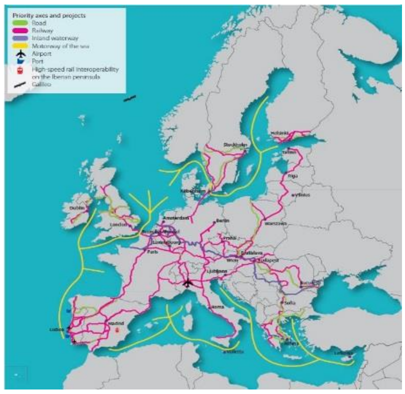
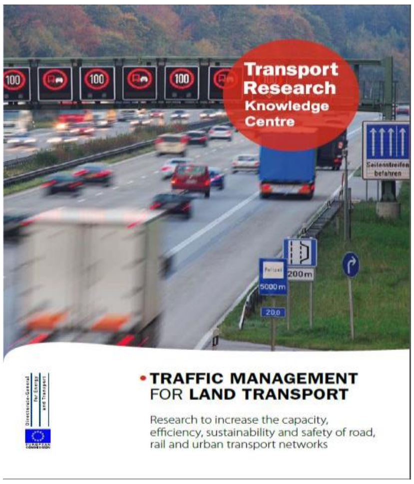


Figure 25. Cross-Modal Transport Infrastructure Innovation Roadmap [118] and Key Routes Identified [119]

Key transportation routes have been identified across Europe covering road, rail and marine transport [118] [[119] (See Figure 25). The European Parliament and Council introduced the Directive 2010/40/EU in July 2010 on the framework for the deployment of Intelligent Transport Systems (ITS) in the field of road transport and for interfaces with other modes of transport. The European Commission's aim is to accelerate and coordinate the deployment of ITS applications on the Trans European Road Network (TERN) across Member States in a consistent and harmonised way.

5.1.3 Traffic Flow and Integration with Infrastructure



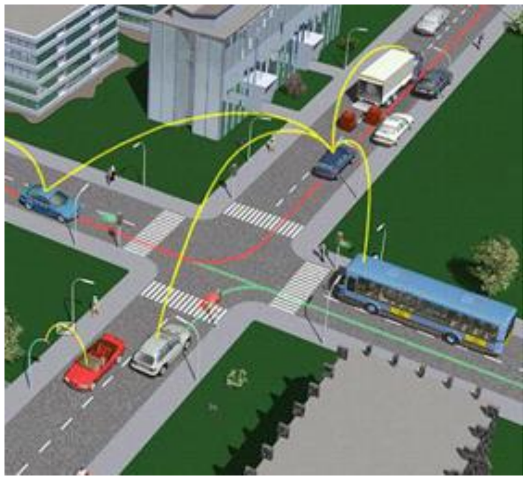


Figure 26. Policy Brochure on Traffic Management [126] and Intelligent Transport Systems [127]

The Transport Research Knowledge Centre (TRKC) consortium produced a Policy Brochure on "Traffic Management for Land Transport" [126] covering both road and rail on behalf of the European Commission's Directorate-General for Energy and Transport (See Figure 26). Although the use of railway signalling and traffic lights in cities have long been used for traffic management a key tenet of this brochure is the need for sophisticated integrated applications based on Intelligent Transport Systems. This is driven by realisation of the need to manage transport networks more effectively in order to maximise the use of existing infrastructure, provide a reliable service to the end user and increase safety, while reducing negative environmental effects. Urban and inter-urban traffic management research and applications are covered in this publication, including aspects such as network management, public transport priority, safety, punctuality and international traffic. Safety related to traffic management, e.g. speed management is also covered. The aim of traffic planning is to plan, monitor, control and influence traffic to maximise the effectiveness of the use of existing infrastructure, provide reliable and safe operation and address environmental goals. A further aim is to ensure fair allocation of infrastructure space (road space, rail slots, etc.) among competing users.

For road or rail transport the scope includes fleet management and timetabling, matching services and vehicles to meet demand and providing essential services while also fitting in with (or finding ways to improve) constraints caused by network capacity, driver shift patterns and technical aspects. For rail traffic the scope includes the bottom operational level of signalling systems and systems for train location; the intermediate level, consisting of the management of rail operations to enhance both the level of service to users and safety; and the higher strategic level, dealing with network access terms and capacity allocation. European policy has long promoted the use of rail in order to rebalance modal shift and encourage the use of this more environmentally friendly and safer transport mode. European rail policy has been developed in the last twenty years to open the competitive market for rail services, first in freight, then in passenger transport and to provide greater interoperability. This is expected to transfer more goods and passengers to rail, at a lower price and with better quality.

In the automotive domain the report highlights that research and deployment of ITS at the EU level is a key tool for traffic management and control to improve safety and user services and reduce the environmental impact of traffic, particularly at infrastructure bottlenecks. ITS applications for traffic management and control include rerouting, Variable Speed Limits (VSL) with automated enforcement, lane control, dynamic use of the hard shoulder on motorways or access control measures such as ramp metering, as well as specific measures for freight such as information on Heavy Goods Vehicle (HGV) parking and “stacking” of lorries in the case of disruption. Cooperative systems, whether vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I,) will play an increased role in traffic management and control in the future. To achieve this coordination across countries and regions, as well as with vehicle and equipment manufacturers, is required. Automatic Vehicle Identification and Location (AVI/AVL) and Automatic Number-plate Recognition (ANPR) is a prerequisite in order to ensure full use can be made of traffic management and enforcement strategies. Existing Automatic Incident Detection (AID) measures can be supplemented by linkage with “probe” (or “floating“) vehicle systems.

Additionally, in a report produced by Transport Research and Innovation Portal (TRIP) consortium for the European Commission’s Directorate-General for Mobility and Transport (DG MOVE) [128], it is highlighted that there is a need to manage conflicting factors, such as vehicles, loads and routes, to improve safety, and to reduce vehicle wear, transport times and fuel consumption. There are many challenges in relation to the increasing connectivity, amount of data (especially real-time), interoperability and integration of new and existing fragmented systems in multimodal and cross-border scenarios. New business models are needed and there is a need to promote behavioural change driven by policy. The challenges for policy-makers include data privacy and data sharing, regulation and planning of multimodal services, standardisation and cross border EU-wide trials. The Strategic Transport Technology Plan (STTP) is the long-term transport innovation policy to identify, in collaboration with all Research and Innovation actors, the most promising technologies. These objectives have been incorporated into the Horizon 2020 research and deployment focus, especially on the potential of ITS as a tool to link transport modes (promoting co-modality and inter-modality) with the services of different operators and infrastructure providers in a single transport mode. This cross modal perspective takes into account transport infrastructure, services and operations [129].

5.1.4 European Innovation Partnership on Smart Cities and Communities

The European Innovation Partnership on Smart Cities and Communities (EIP – SCC) was initiated in July 2012 (See section on Smart Cities). The Smart Cities concept was originally outlined in the digital agenda which contains 132 actions divided in 7 pillars to boost the EU economy and enable Europe's citizens and business to get the most out of digital technologies and infrastructures. Under the Priority Area "Sustainable Urban Mobility", 7 priority actions are being pursued:

- Improve clean power for transport: vehicles and infrastructure
- Foster seamless door-to-door multimodality in urban transport
- Further clean logistics
- Open up intelligence in urban transport systems
- Enable tools for seamless doors-to-door multi-modality
- Promote sustainable and integrated mobility planning
- Promote use of cleaner vehicles.

5.1.5 Smart Cities – Urban Mobility

Considering urban mobility POLIS has produced a policy paper [130] that highlights the needs to support the increased use of electric vehicles, introduce ITS and encourage behavioural change (sharing economy, focus on active travel) within smart cities. This advocates the use of governance with involvement of the Commissioner for Transport to raise the profile of sustainable urban mobility, include the Directorate Generals that govern EU legislation affecting Smart City objectives (DG Environment, DG Regio, DG Connect, DG Move, DG Energy) and engage with urban transport stakeholders.

5.2 Intelligent Transport Systems – European Activities

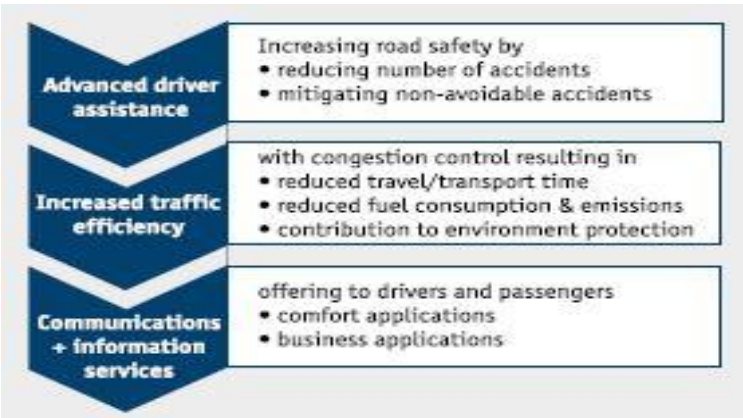


Figure 27. Intelligent Transport Systems [127]

Traffic management is a highly complex problem coming under increasing demands for additional capacity, greater safety and lower costs while meeting strict environmental regulations. At the same time the global car fleet is predicted to double from currently 800 million vehicles to over 1.6 billion vehicles by 2030. Without innovative thinking, integration of information and flow control systems severe congestion will be a major concern for mobility with long commutes and dramatic implications for road haulage of freight leading to logistical problems of late deliveries within highly complex scheduled systems. Already embedded intelligence, mobile phone, car-to-car and car-to-infrastructure communication are offering the opportunity for increased awareness, more efficient mobility and automated driver safety systems.

In the automotive sector Intelligent Transport Systems (ITS) [131] are being developed to provide innovative services relating to different modes of transport and traffic management (See Figure 27). These will enable various users to be better informed and make safer, more coordinated, and 'smarter' use of transport networks. The aims are to increase journey efficiency, reduce congestion, improve road safety and reduce air pollution. An EU Directive (2010/40/EU) was issued on the 7 July 2010 [132] defining the framework for deployment of intelligent transport systems in the field of road transport. Here ITS are defined as “systems in which information and communication technologies are applied in the field of road transport, including infrastructure, vehicles and users, and in traffic management and mobility management, as well as for interfaces with other modes of transport”. Air pollution and congestion are issues in areas of high population density and integrated approaches exploiting combinations of walking, bicycle, buses and trains are advocated.

Among the technologies being explored under ITS are car navigation, traffic signal control systems, vehicle message signs, automatic number plate recognition and speed cameras. Opportunities to link with parking guidance and with weather systems are also being considered. For congestion avoidance advanced modelling techniques are being explored against historical baselines to predict and redirect traffic. Already a lot of work is being performed on "floating car" or "probe" data collection for obtaining travel time and speed data for vehicles traveling along streets and motorways [131]. This can be done from triangulation from mobile phones (which periodically transmit their presence to the mobile phone network), vehicle re-identification using sets of detectors mounted along the road to track a unique vehicle serial number (provided by Bluetooth MAC addresses or RFID serial numbers, e.g. from toll tags) as it travels down a road to give travel times and speed, or from in-vehicle GPS (satellite navigation) systems that have two-way communication with a traffic data provider. A key advantage of floating car data technology is that it less expensive than sensors or cameras, it provides greater coverage, is faster to set up and maintain and works in all weather conditions including heavy rain.

Traffic control is not a new concept and inductive loops that detect magnetic field changes have been placed in road networks for many years to perform ramp metering in order to manage traffic congestion [131]. The simplest detectors simply count the number of vehicles during a unit of time that pass over the loop, while more sophisticated sensors estimate the speed, length, and weight of vehicles and the distance between them. Loops can be placed in a single lane or across multiple lanes, and they work with very slow or stopped vehicles as well as vehicles moving at high-speed. However, more recently advances in telecommunications and information technology, coupled with state-of-the-art microchip, (RFID Radio Frequency Identification), and inexpensive intelligent beacon sensing technologies, have enhanced the technical capabilities opening up new opportunities for more global control of traffic. Vehicle and infrastructure-based networked systems using infrastructure sensors installed or injected into the road or attached to buildings, posts, and signs, can be placed permanently or during road maintenance to provide better monitoring of vehicles operating in critical zones.

Cameras are a common sight on today’s roads and have been used for many years for traffic enforcement to detect and identify vehicles disobeying a speed limit (normally combined with radar detection), detect vehicles that cross red traffic lights, identify vehicles traveling in bus lanes, vehicles crossing railways, crossing double white lines or incorrectly utilising high occupancy vehicle lanes (reserved for car pooling). Number plate recognition systems can be used to automatically issue tickets to offenders. However, more recently cameras are also being used for traffic flow measurement and automatic incident detection.

Cameras are considered to be "non-intrusive" as there is no need to install components into the road surface but they do require some configuration, e.g. input of known measurements such as the distance between lane or the height of the camera above the roadway. The typical outputs from a video detection system are lane-by-lane vehicle speeds, counts, and lane occupancy readings. Some systems provide additional outputs including gap, headway, stopped-vehicle detection, and wrong-way vehicle alarms. These systems have been successfully combined with variable speed limits that change with road congestion and other factors such as weather conditions. One example is the M25 Motorway that circumnavigates London. On the most heavily travelled 14-mile (23 km) section (junction 10 to 16) of the M25 variable speed limits combined with automated enforcement have been in force since 1995 [131]. The results indicated savings in journey times, smoother-flowing traffic, and a fall in the number of accidents, so the implementation was made permanent in 1997.

5.2.1 ERTRAC Strategic Research Agenda for Road Transport

In Europe the ERTRAC Strategic Research Agenda [133] covers mobility, transport and infrastructure, safety and security, environment, energy and resources, design and production. It highlights a number of key research topics including traffic management, integration of vehicle and infrastructure systems, traffic management using ITS, data collection and processing, business models, optimisation of road space to ensure that vehicles (particularly HGVs) adopt routing systems that minimise adverse impacts, systems for segregating traffic with dedicated infrastructure and prioritised traffic management and methods to assist the booking of optimised slots for freight vehicles. The White Paper produced by ERTRAC [134] highlights a number of ongoing projects around Europe and also highlights the key role that exploitation of new ICT functionality will have on the future of ITS. In order to fulfill the aspirations of the Transport White Paper there is a need to coordinate the development of Systems of Systems for surface transport at an EU level with strong political commitment.

5.2.2 European Transport Network Alliance Plus (ETNA) Plus Project

The objective of ETNA Plus [135] is to foster innovation in trans-national cooperation in Transport. The aim is to encourage new actors and regions into EU research calls and projects. ETNA Plus targets transnational cooperation by raising awareness and giving support to national/regional research stakeholders to identify funding sources and partners, create coherence among research activities and transfer best practice via brokerage events, training and coaching.

5.2.3 New Mobility Services

The New Mobility Services project [136] is developing ways to better integrate and manage seamless multi-modal (door-to-door) mobility for urban transport. This is being achieved by providing open platforms and open data that allows public and private service providers to develop and test schemes that provide information, ticketing and planning of trips. The scheme targets large-scale roll out with involvement of 10 cities. A number of key challenges are being addressed, to overcome technological obstacles to deployment and target public investment, to reduce fragmentation and promote sharing of data to provide services, ensure quality of data, enable standardisation and interoperability, define business models and provide governance and reach a balance between private and public bodies.

5.2.4 DRIVE C2X

The DRIVE C2X project [137] has 34 partners, 13 support partners and an 18.6 million Euro budget. It aims to provide the foundations for cooperative systems in Europe with the aims of safer, more economical and more ecological driving. The project is carrying out field tests of systems leading on from the PRE-DRIVE C2X which implemented technologies in European test sites in Finland, France, Germany, Italy, Netherlands, Spain and Sweden. An aim is to raise public awareness, provide feedback for standard organizations and support for initiating public-private ventures. The work focuses on communication between vehicles (C2C) and the roadside and backend infrastructure system (C2I). Previous projects such as PReVENT [138], CVIS [139], SAFESPOT [140], COOPERS [141], and PRE-DRIVE C2X [142] have proven the feasibility of safety and traffic efficiency applications based on C2X communication. DRIVE C2X goes beyond the proof of concept and addresses large-scale field trials under real-world conditions at multiple national test sites across Europe. The systems being tested are built according to the common European architecture for cooperative driving systems defined by COMeSafety [143]. This guarantees compliance with the upcoming European ITS standards. This approach also ensures that the results of DRIVE C2X have long-term validity at a European level, giving system developers as well as decision maker’s confidence.

DRIVE C2X is also implementing and testing a concept for the integration of a data backend, enabling commercial services based on C2X communication data to be developed for private and commercial customers. Such services are expected to become a major revenue source for cooperative driving systems and are key for successful implementation of this technology on European roads.

5.2.5 Autonomous Cars - HAVEit – Highly Automated Vehicles for Intelligent Transport

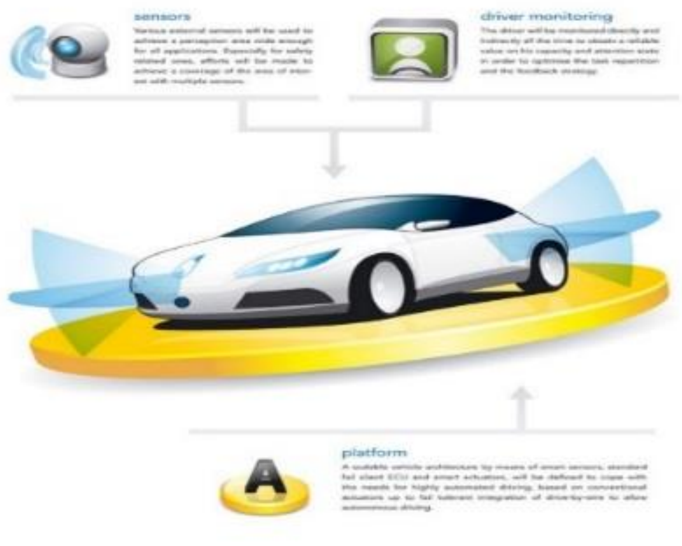


Figure 28. - HAVE-it Autonomous Car

There is great interest in the automotive industry in introducing autonomous driving features to improve safety. The EU funded HAVEit project [144] has developed concepts and technologies for highly automated driving (See Figure 28). The key drivers for automated driving are increasing traffic density, the growing flood of information available to drivers and the rising average age of the population. In Europe 1 in 3 people will be over the age of 60 in 10 years' time. Automation is needed to relieve drivers of some of the stress of driving guiding them through traffic more efficiently with a consequent environmental benefit. This will help reduce driver workload, prevent accidents, reduce environmental impact and make traffic safer. The HAVEit consortium (17 partners) consisted of vehicle manufacturers, Continental, Volvo Technology AB, Volkswagen AG, automotive suppliers and scientific institutes from Germany, Sweden, France, Austria, Switzerland, Greece and Hungary. In total, investments of 28 million Euro were made into HAVEit, 17 million Euro of which were EU grants and 11 million Euro were contributed by the 17 partners. Seven demonstration vehicles were produced.

Highly automated vehicles can take over three main driving functions: steering (lateral automation), path planning (longitudinal automation) and navigation. These make driving easier for people and create highly automated systems which can be used intuitively. As part of the HAVEit project, three automation modes which can be selected and activated by drivers were developed and implemented in all demonstration vehicles.

- Normal - Lane keep assist and emergency brake assist
- Longitudinal automation - no need to accelerate or brake
- Lateral automation - no need to steer

In the first mode, the driver steers the vehicle alone, assisted by already-available standard driver assistance systems, such as lane keep assist or an emergency brake assist. In partly or semi-automated mode, the vehicle drives with longitudinal automation, so drivers no longer have to accelerate or brake. At the level of high automation, lateral automation comes into play, meaning the driver no longer has to steer. Despite the level of automation selected, the driver is always fully responsible for manoeuvring the vehicle and can take control in place of the system at any time. The driver also has to monitor the vehicle's driving manoeuvers. In the partially and highly automated modes, the system observes the driver with the help of a camera located inside the vehicle. The moment the driver stops paying attention to the road, the assistant prompts them to take control of the wheel. The German Aerospace Centre (DLR) and the Wuerzburg Institute of Traffic Sciences (WIVW) developed the concepts of adaptive communication between the driver and the automated vehicle.

5.2.6 Drive Me

The Swedish Drive Me project [145] is a joint autonomous driving pilot project between the Volvo Car Group, the Swedish Transport Administration, the Swedish Transport Agency, Lindholmen Science Park and the City of Gothenburg with a vision for zero traffic fatalities. 100 self-driving Volvo cars will use 50 Km of selected public roads in everyday driving conditions around the Swedish city of Gothenburg to identify:

- How autonomous vehicles bring societal and economic benefits by improving traffic efficiency, the traffic environment and road safety
- Infrastructure requirements for autonomous driving
- Typical traffic situations suitable for autonomous vehicles
- Customers' confidence in autonomous vehicles
- How surrounding drivers interact smoothly with a self-driving car

The roads to be used are typical commuter arteries and include motorway conditions and frequent queues. The project started in 2014 with work on customer research and technology development, as well as the development of a user interface and cloud functionality. The first cars are expected to be on the roads in Gothenburg by 2017. An aim is that the Drive Me project will help define the role of self-driving vehicles in future city planning reducing infrastructure investments, lowering emissions and improving traffic safety. The driver will hand over responsibility to the vehicle, which can handle all driving functions at the driver's discretion, however, the driver is expected to be available for occasional control with a sufficiently comfortable transition time. For drivers autonomous driving is expected to provide more efficient time-management behind the wheel with the ability to interact safely via phone or tablets. The project will also investigate fully automated parking, without a driver in the car, such that the driver can walk away from the car at the parking entrance while the vehicle finds a vacant spot and parks by itself.

5.3 European Green Initiatives

In Europe there are a number of “green initiatives” to take up the use of electric vehicles and to reduce pollution across a range of transport modes.

5.3.1 Green Cars Initiative

Figure 29. Green Cars Initiative

The European Green Cars Initiative [146] (See Figure 29) provides funding for green technologies for cars, trucks and buses. Funding is via a combination of EC grants and loans from the European Investment Bank. Research is being funded on greener combustion engines for trucks, bio-methane, electric and hybrid vehicles and infrastructure.

5.3.2 Smart, Green and Integrated Transport European Programme

The Smart, Green and Integrated Transport programme [147] is targeted at boosting the competitiveness of the European transport industries. The aim is to create a European transport system that is “resource-efficient, climate-and-environmentally-friendly, safe and seamless for the benefit of all citizens, the economy and society”. This is being supported with a budget of €6339 million (2014-2020) from Transport Challenge funding for cleaner and quieter aircraft, vehicles and vessels. The intention is to minimise impact of transport on the climate and the environment.

A number of areas are being funded including research on smart equipment, infrastructures, services and technology. The aim is to provide better mobility and reduce congestion in urban areas, improve safety by reducing accident rates, provide better security, and reduce traffic congestion generally to provide greater mobility of people and freight. There is also an aim is to support improved policy making to promote innovation and meet the challenges raised by transport and the societal needs related to it. Calls for this programme have been made under Mobility for Growth, Green Vehicles and Small Business and Fast Track Innovation for Transport.

5.3.3 Electromobility

The EV4SCC (Electric Vehicle for Smart Cities and Communities) project [148] brings together cities and regions with companies to showcase innovative electro-mobility solutions. The aim is to promote “replication at scale” in intelligent management of public and private fleets of electric-vehicles, smart urban logistics with light e-vehicles, smart electrification of public transport, innovative integrated infrastructure solutions and smart electro-mobility solutions that serve multi-modal mobility services [149].

5.4 Trucking and Logistics - Drivers and Initiatives

Figure 30. Logistics Issues – Traffic Jams [photo dpa-Zentralbild], Congestion from Deliveries [photo Abenblatt.de]

Information provided by modern ICT systems is available at all levels of the supply chain offering unprecedented opportunities for optimization. Transport volumes keep growing globally (See Figure 30) leading to congestion on roads, however, the sizes of individual shipments are not increasing and indeed there is a move towards shipments of smaller loads [150]. The move towards global sourcing has changed the dynamics of logistics. For example, 10 years ago 90% of the parts for a car would come from factories within a 200Km radius, now the parts are sourced from a world-wide supplier base. Customer service expectations are high with demands for fast and efficient on-time delivery. In order to execute transport tasks efficiently transport service networks play a vital role. These networks are dedicated, e.g. to parcel, express or less-than-truckload-shipments and related logistic services. Analysis and optimization of their structure can provide great benefits in terms of efficiency and also fuel cost and emissions reductions. More efficient operation of nodes (depots, hubs, terminals) provides greater throughput and lower latency. To support this operators are increasingly turning to simulation models to achieve robust solutions that improve their efficiency, reduce handling costs and increase the performance of their terminal operations. A key challenge is to link between material flow simulation and arriving and departing traffic.

The task of delivery in urban areas increasingly is leading to congestion (See Figure 30) and ways of bundling deliveries at local hubs to reduce the numbers of vehicles making deliveries is also challenge. Urbanisation is a key challenge and air quality directives such as Euro 6 [151] are driving new truck and powerplant design. Likewise stricter standards are being introduced in the USA and China for fuel economy and emissions.

The trucking community is familiar with the use of ICT and already automatic tolling systems are used across Europe, however, there is a lack of harmonisation of systems and to operate across all of Europe a lorry driver needs a myriad of different devices on the dashboard. For main routes across Europe drivers typically need 7 different tolling devices. The use of telematics and connectivity is seen as the future to make major improvements in management of freight efficiency, emissions, safety and personal effectiveness. Take up of telematics is, however, still low in the industry as the average fleet size in Europe is 10 trucks. Medium to large companies account for around 25% of the trucking companies in Europe and small companies for the remaining 75%. There is also driver resistance to being tracked. Typical experience shows that just by introducing a tracking device on a vehicle there is a 5-10% saving in fuel – indicating that drivers do not always use their vehicles for work.

There are a number of potential benefits from introducing tracking. These include monitoring of driving behaviour which can be fed back to the driver (highly fuel efficient trucks do not make a difference if the driving is bad), provision of routing to the cheapest petrol stations, and reductions in insurance claims (from providing proof of speed, etc. in court cases). Additionally, monitoring of key truck parameters can be used to optimise efficiency, e.g. truck tyre pressures have a big impact on efficiency, and there is a great interest in moving from remote diagnostics to prognostics as batteries and tyres account for 50% of breakdowns. Already companies such as Scania give away a free telematics system with all of their trucks and currently 800,000 vehicles are fitted with it. Services are provided based on this and customers have the option of buying them.

The key benefit of telematics is in gathering and exploiting data in fleet management. Customers want to know every minute where a delivery is and there is a move from reactive to proactive operations through data mining of Big Data. A major issue that contributes to unnecessary fuel consumption and emissions is the shipping of goods in half empty trucks and the return of empty trucks (where there is an immediate 40% penalty in fuel consumption). Means of co-ordinating and optimising deliveries across fleets of vehicles can thus bring huge savings.

5.4.1 DHL GOGREEN Initiative

The DHL GOGREEN initiative [152] is introducing optimised transport routes, alternative drive vehicles and energy-efficient warehouses to reduce CO_2 emissions and other environmental impacts in the transportation and storage of goods. By 2020 the company aims to increase the carbon efficiency of its operations by 30% compared with 2007 levels. Already the company has achieved a 10% reduction. Sustainability is seen as a competitive factor driven by consumer demands and also by investors who consult sustainability rankings when looking for viable investment options. To address this the GOGREEN initiative is considering a complete view of emissions with the aim to "burn less and burn clean" across all vehicles, buildings and aircraft. Already there are 11,500 green vehicles on test utilising a mix of electric and alternative fuels. A systems approach is being adopted and solar panels are used to charge electric vehicles at warehouses and a new rail link to China is being used as an alternative to flying goods. This allows goods to be shipped from China priority within 7 days by air or 28 days by rail depending on customer requirements. The company provides Carbon Reports and a Green Optimization service to identify ways to minimise greenhouse gas emissions and improve overall environmental performance. Carbon accounting has been integrated into financial accounting systems so that the emissions are automatically calculated from fuel and electricity consumption data. To compensate for unavoidable emissions a "climate neutral" approach is offered using energy provided by solar panels and wind turbine energy.

5.4.2 Carbon Footprint

The growing freight transport sector is a major contributor to greenhouse gas emissions. Several initiatives exist for the calculation of the carbon footprint of freight transport chains. However, there are problems in terms of comparability, transparency and accuracy since these initiatives are based on different starting points, approaches or intentions in development. The EU co-funded project COFRET (Carbon Footprint of Freight Transport) [153] is developing a unified approach to calculate logistics related carbon footprint emissions along complex supply chains. Likewise there are efforts in the logistics industry to harmonise the measurement of emissions from trucks for specific driving cycles and introduction of badging of truck CO_2 efficiency by Green Freight Europe [154].

5.5 National Smart Transportation Initiatives

5.5.1 The Netherlands

The Netherlands is one of the leading European countries with respect to the implementation of smart traffic systems. The government is making substantial investments in new forms of smart mobility allocating more than €70 million for intelligent transport systems (ITS) by 2018 [155]. Nine projects are being set up to deploy new services and gain practical experience in traffic management, to better distribute traffic on the roads to avoid congestion, and to provide services that give travellers real-time driving and travel advice during their journey. This includes:

- Tools to improve supermarket logistics by reducing 200 lorry trips per day in the Groningen-Assen, Arnhem-Nijmegen, Midden-Nederland regions and the Amsterdam Metropolitan Area.
- Improved sharing of information to reduce the build-up of traffic due to incidents in Brabant and Noord-Holland where more than 20,000 trucks and 150,000 passenger cars break down on the main road network. (It is predicted with national coverage, this could reduce congestion by 2.5%).
- Use of event apps to reduce congestion around festivals, congresses and concerts (festival goers account for five million rush-hour car trips in and around cities per year)

Collaborations on traffic management are also being performed with the Amsterdam Pilot, the ITS Corridor, the Innovation Traffic Centre and the policy theme Autonomous Vehicle Travel, among other initiatives. These and other efforts are part of the Road Map for Better In-transit Information (Routekaart Beter Geïnformeerd op Weg).

5.5.1.1 City region of Eindhoven Initiative

IBM and NXP Semiconductors N.V. conducted a 12 month smarter traffic pilot in Eindhoven demonstrating how a connected car can automatically share anonymous information on braking, acceleration and location that can be used by the central traffic authority to improve traffic flow, reduce congestion and resolve road network issues [156]. Eindhoven is located at the hub of several international transportation routes, where relatively small incidents can have major consequences for the system as a whole. In an earlier six-month road pricing trial conducted by the city in conjunction with IBM and NXP, advanced road pricing technology was successfully used to incentivise drivers to change their driving behaviour, reduce road congestion and contribute to a greener environment. Seventy percent of drivers changed their behaviour to avoid rush-hour travel when presented with the right incentives, demonstrating that road pricing systems can have a positive effect on driving habits and help alleviate traffic. Participating pilot cars were equipped with a device containing a telematics chip which gathers data from the central communication system of the car (CAN-bus) that gives indicators of potholes or icy roads which is transmitted to the cloud-enabled traffic centre.

5.5.1.2 Enschede Vehicle Inductive Profile

In Enschede the travel times of vehicles is collected by Smart detection loops at traffic lights [157]. Travel time savings are stored in a database, processed and shown on four dynamic route information panels on Highway 35 allowing drivers to choose the most favourable route in rush hour. The system is co-funded by Regio Twente.

5.5.2 Spain

5.5.2.1 Electric Barcelona

Barcelona has a growing fleet of electric vehicles, taxis, cars and motorcycles that benefit from the facilities offered by the city, such as its 300 public recharging points (See Figure 31) that are free to use and found at a number of stations across the city [158]. The city is also introducing a new way of renting cars, based on car sharing, using electric vehicles. The city also has the cleanest bus fleet in Europe, including hybrid and compressed-natural-gas vehicles.

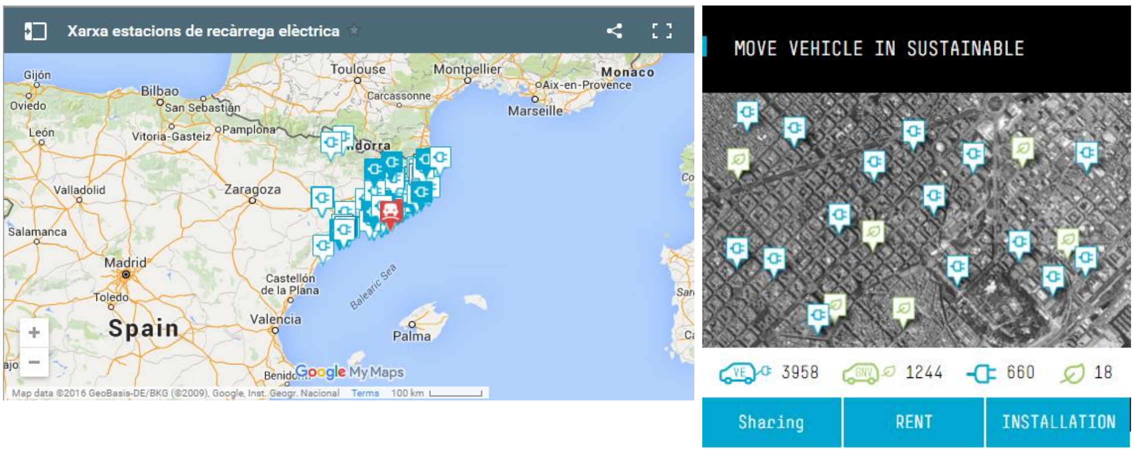


Figure 31. Barcelona Electric Charging Network [159]

A public-private partnership is being used to provide real time management of the city, an urban traffic sensor system, traffic management, mobile applications for public transport and fast ticketing. New business opportunities are being found in fleet management, pre-booking of parking slots and optimization of routes for emergency services [160].

5.5.2.2 Zaragoza traffic monitoring system (Spain)

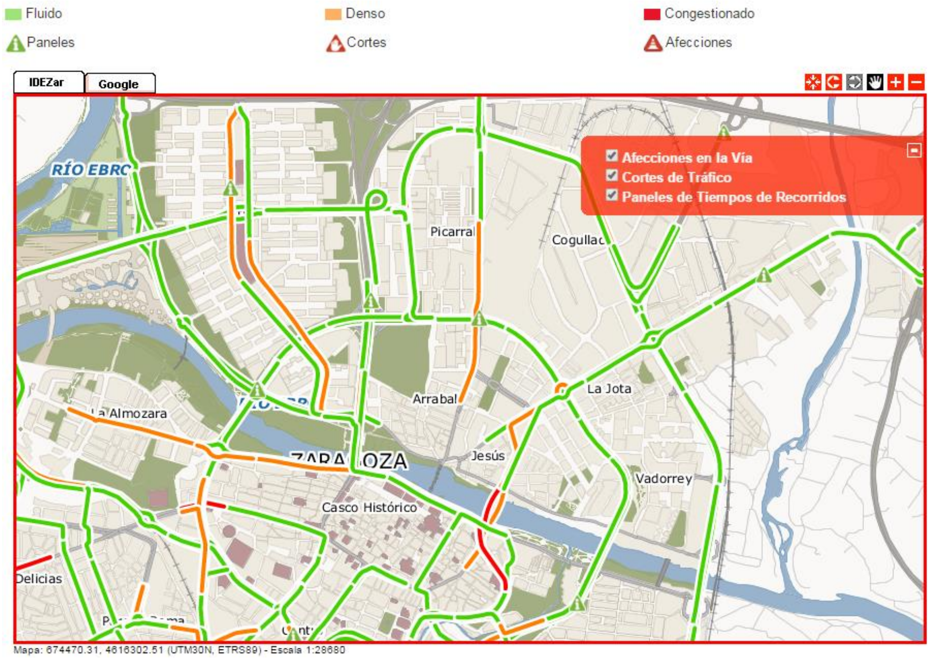


Figure 32. Zaragoza Municipal Traffic map

The Zaragoza City Council has installed 150 traffic sensors at traffic lights in 90% of the city's main thoroughfares which continuously send data to the Traffic Control Centre for traffic management. This is used to provide information to the public via mobile devices [161]. This gives information on the state of the main city roadways by means of a colour code, any incidents that have occurred, scheduled traffic diversions and serious traffic jams (See Figure 32).

Additionally, a metering infrastructure and vehicle passage sensors have been installed at numerous points in the city to allow traffic-light re-programming to cope with traffic conditions. Zaragoza has become Europe's least congested city with over 500,000 inhabitants according to TomTom [162].

5.5.3 Ireland

Dublin City Council (DCC) has made all of its data available to the IBM Smarter Cities Technology Centre in Dublin [163]. Big Data analytics is being used to gather insights from the data and this is being passed back to the council's roads and traffic department. As a result real-time geospatial data from 1,000 buses across Dublin and their timetables are integrated with a digital map of the city. Dublin's traffic controllers can use digital maps and dashboards to monitor the status of the entire bus network at a glance. It is also possible to see congestion forming in real time and instantly access CCTV feeds to identify the cause and apply appropriate traffic-calming measures. Data from rain and flood gauges is also being integrated to alert to potential hazards presented by extreme weather conditions. IBM is currently developing a predictive analytics solution combining data from the city's tram network with electronic docks for the city's free bicycle scheme to optimise the distribution of the city's free bicycles according to anticipated demand from tram passengers.

5.5.4 Greece

In Thessaloniki two different systems have been put into place [164]. The first, a new traffic control centre, manages incidents with real-time information, dynamically estimates traffic for the rest of the day, assesses and confirms estimated travel times, and dynamically manages traffic lights. The second system is a mobility planner that provides citizens with real-time traffic condition data, enabling them to choose between the shortest, most economical and most environmentally friendly route. This is funded 50% from Iceland, Liechtenstein and Norway through the 'European Economic Association' and 50% by National Funds. An aim is to promote the use of public transportation, walking and cycling. Finally, through special urban mobility training programmes, a new culture for urban mobility is being promoted.

5.5.5 UK Smart Transport Catapult

Figure 33. Intelligent Mobility

The Transport Systems Catapult [165] is one of seven elite technology and innovation centres established and overseen by the UK's innovation agency, Innovate UK. The centre promotes Intelligent Mobility (See Figure 33) using new and emerging technologies to transport people and goods more smartly and efficiently. The aim is to help UK businesses create products and service. The aim is to sell these at the world level and provide a test bed for the transportation industry. There is an emphasis on collaboration, bringing together diverse organisations across different modes of transport

5.5.5.1 MIRA

MIRA in the UK operates Europe’s most advanced ITS test track, innovITS-Advance [166] dedicated to research and development of intelligent transportation systems (ITS). This utilises modern communication technologies, private GSM and cellular networks, fully configurable wireless networks and state-of-the-art, vehicle-to-vehicle communications based on the draft 802.11p WAVE standard [167]. Work is investigating transport information, intelligent vehicles, and intelligent infrastructure, looking at a range of topics including OEM/aftermarket applications for congestion, hazards, tracking and fleet management, data management and modelling, real-time data sharing, integrated in-vehicle multimedia applications, aftermarket and integrated HMI solutions, secure communication networks, pedestrian safety, vehicle positioning and sensor systems, co-operative control systems and autonomous systems.

5.6 Rail Transport - Drivers and Policy

The European rail infrastructure is facing increasing congestion due to unprecedented numbers of passengers requiring innovative ways to increase capacity on existing infrastructure (faster scheduling of passengers through stations and shorter stopping times at stations) and demanding levels of punctuality never before seen with more people and improved journey times. Here management, control and sociological aspects need to be considered in unison.

The interoperability regulations and the 2011 Transport White Paper [168] require that the European railway system behaves as a single Systems of Systems. The commercial drivers in the industry are for 24/7 operation, high availability, low cost, safety, increased capacity, recovery from disturbance, low carbon emissions and customer satisfaction. Already trains are operating across the European continent and the Commission requires a level playing field without barriers to competition. The main competitors to the rail network are other modes of transport and in order for the railway to be the preferred transport mode, the industry must offer a guaranteed door-to-door or factory-to-point-of-sale service 24/7. To achieve this there is a drive towards Automatic Train Control and automated maintenance to increase capacity and reduce costs to the point where rail operations do not require subsidy from government.

Capacity is currently severely restricted due to controlling train movement through a system of blocks (sections of “reserved” track that no two trains can operate on). Moving blocks improve this but autonomous train-to-train communications and new infrastructure components could increase capacity by more than 100% with an asset value of billions. The 2011 Transport White Paper [168] requires the majority of medium to long distance journeys (freight and passenger) to be by rail. This is driven by congestion costs (1.5% of EU GDP) and the need for greatly reduced transport emissions. The EU is driving the railway industry towards a single system through interoperability requirements. The industry also aims for a more resilient infrastructure to route traffic in an optimal manner responding to an incident. Key improvements being sought:

- **Improved capacity** - Improved planning and operation with potentially more flexible timetables could deliver improvements in capacity, by optimising the timetables at peak periods to maximise traffic flow.
- **Reduced Emissions** - Improved timetable planning and operation, can lead to optimised driving to reduce stopping and starting to reduce emissions. A Systems of Systems approach may also provide the necessary planning that would allow hybrid rail vehicles to just run the combustion engine away from stations and urban areas, reducing noise and urban pollution.

5.6.1 ERRAC Strategic Research Agenda 2020 for Rail

ERRAC [121] was set up in 2001 with the goal of creating a single European body with both the competence and capability to help revitalise the European rail sector and make it more competitive, by fostering increased innovation and guiding research efforts at a European level. Within ERRAC, all major rail stakeholders are gathered, including 45 representatives from each of the major European rail research stakeholders: manufacturers, operators, infrastructure managers, the European Commission, EU Member States, academics and users' groups. ERRAC covers all forms of rail transport from conventional, high speed and freight applications, to urban and regional services. Since its launch in 2001, ERRAC has produced a number of important and influential documents, such as the Joint Strategy for European Rail Research – Vision 2020 [169], the SRRA – Strategic Rail Research Agenda [170] and its 2007 updated version, Suburban and Regional Railways Landscape in Europe [171], Light Rail and Metro Systems in Europe [172], Rail Research in Europe [173] and a comparison of the Member States public research programmes.

A set of roadmaps were developed in the EU funded (FP7) project ERRAC ROADMAP (2009-2012) and in 2012, an initial update of the ERRAC vision for the future of rail to support H2020 was released. This vision "Railroute 2050" [174], highlights the European effort required for research and innovation especially to meet the objectives of the European Commission 2011 Transport White paper "Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system". RailRoute 2050 offers a range of research opportunities for a competitive, resource-efficient and intelligent rail transport system that meets the future demands of European citizens, stipulates economic growth, creates European jobs, and strengthens the position of the European rail sector in global competition. The European vision for railway research and innovation outlined in RailRoute 2050 illustrates the research pillars that need to be supplemented by the corresponding investment pillar. Additionally, ERRAC launched FOSTER RAIL (2013-2016) [175] which will support development of a new full and complete vision, including the Rail Business Scenario and the Strategic Rail Research and Innovation Agenda.

5.6.2 European Rail Roadmap

Between 2009 and 2012, ERRAC carried out a 3 year rail research roadmapping project called ERRAC ROADMAP [176]. This highlighted the needs for intelligent mobility, competitiveness and enabling technologies and infrastructure as priority research areas related to traffic management. In the area of intelligent mobility, the main issues deal with the definition of new management techniques to enhance infrastructure use. These include timetable optimisation, new fleet management tools, and development of information systems, as well as harmonised information exchange between stakeholders in cross-border traffic. In the area of competitiveness and enabling technologies, priorities include the compatibility of on-board data collection systems and their integration with communication networks, as well as the analysis of passengers and traffic flows in order to reach a more efficient Europe-wide train path allocation. Finally, in the infrastructure area, priorities include the development of train control systems and new operational rules in order to optimise both capacity and service interchange.

5.6.3 European Rail Traffic Management System (ERTMS)

The European Railway Traffic Management System (ERTMS) [177] is a major industrial project developed by Alstom Transport, Ansaldo STS, AZD Praha, Bombardier Transportation, CAF, Mermec, Siemens Mobility and Thales in close cooperation with the European Union Railway stakeholders and the GSM-R industry. Currently there are more than 20 train control systems across the European Union. Each train used by a national railway company has to be equipped with at least one system but sometimes more are required to be able to run safely within one country. A problem is that each system is stand alone and non-interoperable. If traffic is cross border this leads to extensive integration and engineering effort with high associated costs. This restricts competition and also hampers competitiveness of the European rail sector versus road transport. As an example the Thalys trains running between Paris-Brussels-Cologne and Amsterdam have to be equipped with 7 different types of train control systems.

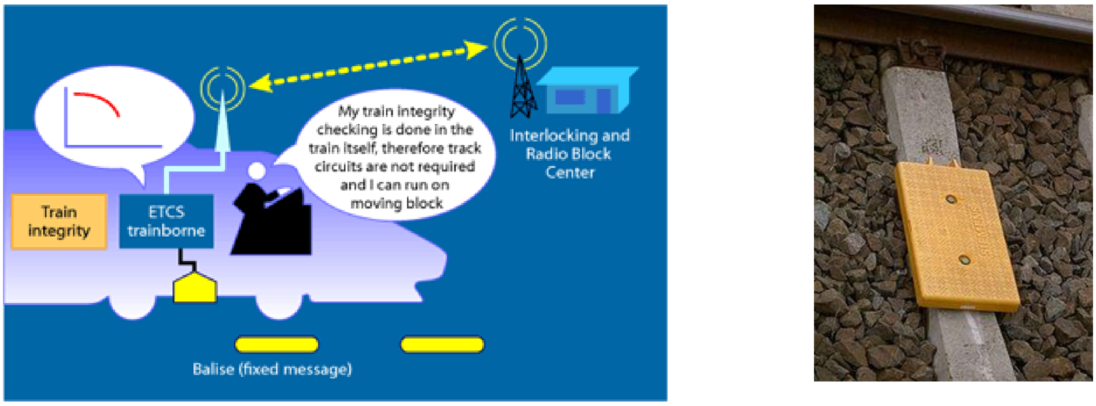


Figure 34. ERTMS Level 3 and Balise

To address this ERTMS aims to gradually replace the different national train control and command systems across Europe to create a seamless European railway system. Instead of lineside signals, a computer in the driver's cab controls the speed and movement of the train, whilst taking account of other trains on the railway (See Figure 34). Bringing the control system inside the train will allow more autonomous operation, so that drivers can always run at the optimum safe speed helping more trains run faster and recover from delays quicker. Each train will run at an appropriate safe speed, allowing more trains onto the tracks. This will increase passenger and freight capacity, reliability, reduce maintenance costs, improve punctuality and lead to safer trains and greater competitiveness for the supply market. By moving more people and freight onto trains and reducing delays there is also an expected reduction in pollution.

ERTMS has two basic components, the ETCS, the European Train Control System, which is an Automatic Train Protection system (ATP) to replace the existing national ATP-systems, and GSM-R, a radio system for providing voice and data communication between the track and the train. This uses standard GSM but on a reserved rail frequency. It should be noted that ERTMS is not a new concept and it has been successful outside Europe in countries such as China, India, Taiwan, South Korea and Saudi Arabia. The ERTMS/ETCS is split into a number of application “levels” which range from track to train communications (Level 1) to continuous communications between the train and the radio block centre (Level 2). Level 3, which is in a conceptual phase, will further increase ERTMS potential by introducing a “moving block” technology to increase capacity.

- ERTMS level 1 is used as an add-on to conventional lineside signals and train detectors. Communication between balises (See Figure 34) and the train ensures that it automatically brakes if exceeding maximum allowed speed.
- ERTMS level 2 does not use lineside signals (reducing maintenance costs by their removal). The movement authority is communicated directly from a Radio Block Centre (RBC) to the onboard unit using GSM-R. Balises are used to transmit “fix messages” such as location, gradient, speed limit, etc.
- ERTMS Level 3 is still in its conceptual phase but allows introduction of “moving block” technology. Removal of fixed blocks (sections of tracks where two trains cannot run at the same time) increases capacity greatly. The train itself becomes a “moving block” communicating accurate position data.

5.6.4 Foster Rail

FOSTER-RAIL is a European Level 1 Coordination and Support Action [175] driven by ERRAC aimed at supporting the land transport European Technology Platforms activities. The aim is to strengthen the research and innovation strategies of the transport industries in Europe. This will assist ERRAC and the other transport-related European technology platforms (ETP) in defining research needs for their strategies and programmes for H2020 in order to realise the objectives of the Europe-2020 strategy [169] and the White Paper 2011 vision for a competitive and resource-efficient future transport system [168]. An updated Strategic Rail Research and Innovation Agenda is being produced under Foster Rail. It is being performed in consultation with the European Commission and Member States and Associated States. FOSTER-RAIL will integrate the work of ERRACs Working Groups and progress this building upon the ERRAC ROADMAP project and RailRoute 2050 [174]. An aim is to support and enhance cooperation between stakeholders and decision-makers to provide an enhanced definition of strategic research and innovation needs and establishment of Business Scenarios. A key area is co-modality with other transport modes. The project will support the Strategic Rail Research and Innovation Agenda as well as a Rail Business Scenario for 2050. This Railway Business Scenario shall be the reference for future research agendas and technology roadmaps to be developed until 2050.

5.6.5 SHIFT^{2}RAIL

SHIFT^{2}RAIL [178] is a H2020 supported European rail joint technology initiative seeking focused research and innovation (R&I) and market-driven solutions by accelerating the integration of new and advanced technologies into innovative rail product solutions. The integration of systems is a core objective of the programme and it applies to all segments of the rail market: High Speed/Mainline, Regional, Urban/Metro & Suburban, and Freight. SHIFT^{2}RAIL aims to promote the competitiveness of the European Rail Industry and create a Single European Railway Area (SERA). SHIFT^{2}RAIL has set targets to double the capacity of the European rail system, increase its reliability and service quality by 50% and at the same time halve lifecycle costs.

The aim is to achieve this by introduction of better trains (more comfortable, quieter and more reliable), operating on an innovative rail network infrastructure in a reliable way from the first day of service introduction. This will be done at a lower life cycle cost, with more capacity to cope with growing passenger and freight mobility demand. SHIFT^{2}RAIL also aims to attract more users to rail. For passengers there will be more travel options, more comfort, and improved punctuality. For the freight forwarder/shippers rail freight will become more cost effective, punctual and traceable as a shipment option. There is an expectation of more job creation, less pollution and more efficient and optimised public investments.

5.7 Aerospace Transport - Drivers and Policy

Air passenger volume is predicted to double air traffic density over the next two decades in an already congested airspace. Movement of increasing numbers of passengers requires complex management and integration across the world of airport operations, baggage handling and air traffic control to maximise flow. Air traffic control systems by themselves integrate numerous functionalities which enable semi-automated operations in the en-route airspace. Tools and methods that partially automate some of what is manually performed by Air Traffic Controllers today is currently an active area of research. At the same time the need for unprecedented high levels of aircraft availability is driving the use of sophisticated information and communications technologies for predictive health monitoring, integrated with worldwide maintenance and logistics systems to ensure that aircraft are always fit to fly.

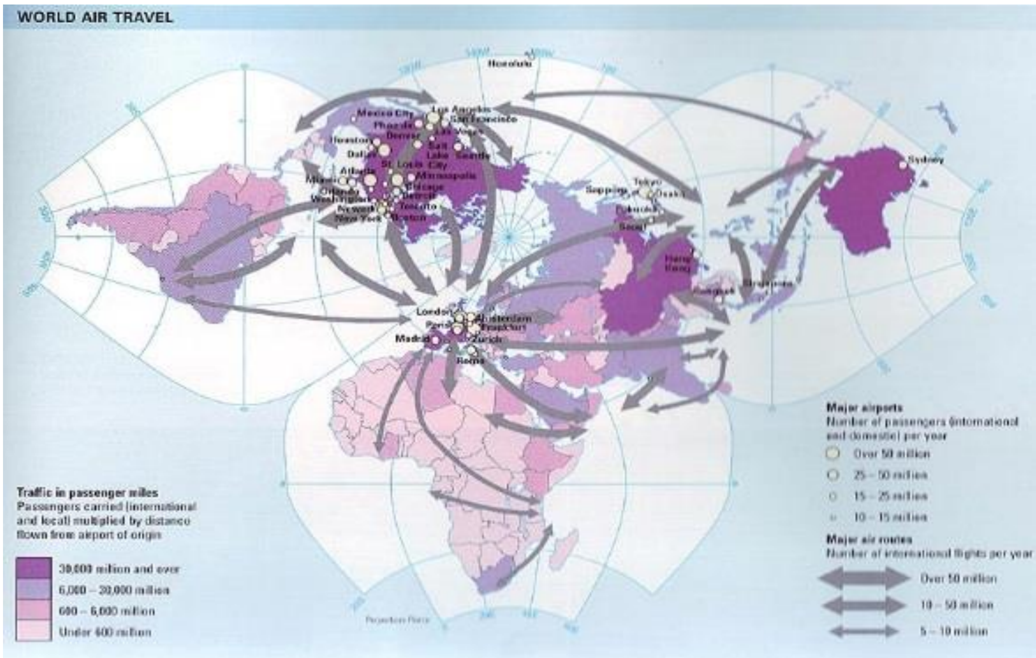


Figure 35. Major World Airports and Traffic Routes [179]

Figure 35 shows a map of major air traffic routes around the world [179]. On the key routes shown there are over 50 million passengers a year and on the majority of other routes there are 10-50 million passengers. Air traffic is increasing and the number of aircraft is expected to double by 2020 to meet demand. As a global aviation industry, the biggest and most important challenge is to continue to safely accommodate ever increasing air traffic in support of global economic growth and prosperity, whilst protecting the environment.

Air Traffic Management will be a major topic in the coming years, especially in Europe where separate systems will have to be integrated. The challenges here are not only technological, but also legislative/political and need to be tackled at a European (and even world-wide level). In the future Unmanned Aerial Vehicles will also be integrated with the normal ATM network presenting further technological, legislative and political challenges.

5.7.1 European Air Traffic Management - SESAR

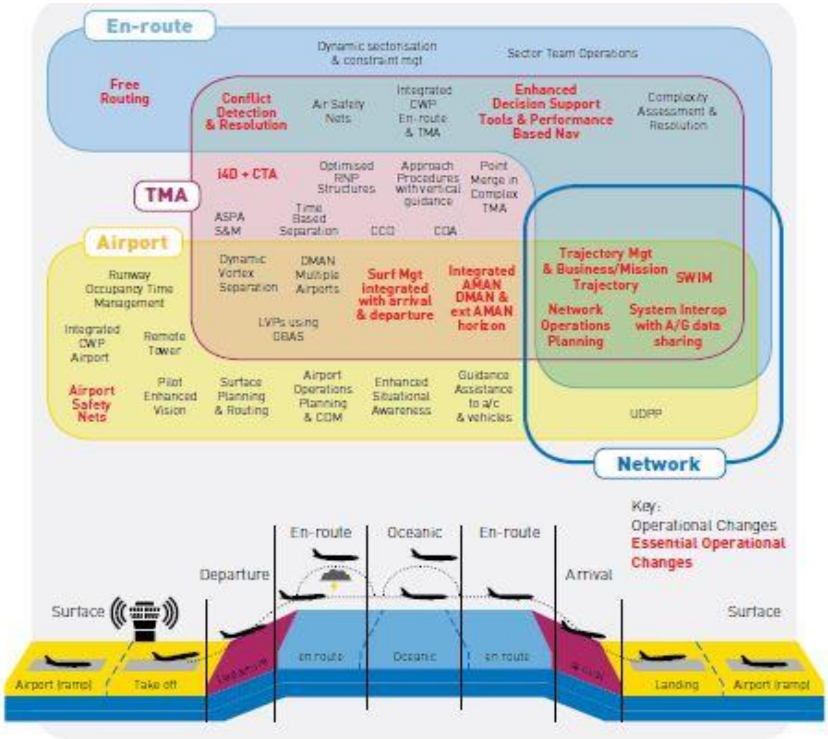


Figure 36. SESAR 4D Routing [180]

The Single European Sky programme [179][180] is reforming the architecture of European Air Traffic Control to meet future capacity and safety needs. Within Europe Eurocontrol predicts 20.4 million yearly flight movements by 2030 which is twice the current figure. In order to meet this need €2.1 billion is being invested in R&D to develop a new air traffic control system for Europe. This will exploit improved air traffic and aircraft positioning and communication technologies, such as GALILEO [181] to provide significant improvements in the efficiency and safety of air travel. The Single European Sky ATM Research programme (SESAR – formerly known as SESAME) is the name given to the collaborative project that will completely modernise the European Air Traffic Control infrastructure. SESAR (See Figure 36) aims at developing the new generation air traffic management system capable of ensuring the safety and fluidity of air transport worldwide over the next 30 years.

The first definition phase of SESAR ended in 2008, delivering an ATM master plan [182][183][184] defining the content, the development and deployment plans of the next generation of ATM systems. This activity was led by Eurocontrol, and co-funded by the European Commission under the Trans-European Network Transport programme. Work was executed by a consortium with representatives of all air transport stakeholders and included non-European members reflecting the global nature of ATM. The development phase (2007-2013) provided a new generation of technological systems and components. For this phase the Commission created the SESAR joint undertaking, based on the GALILEO model, supported by public and private funds from the European Community, Eurocontrol, industry and third countries. The current deployment phase (2013-2020) is seeking to build the new infrastructure necessary for the future within Europe and in partner countries. This is being carried out under the responsibility of industry without further public funding. Additionally, there has been some activities performed under the Clean Sky EU programme [185] looking at ATM and related issues to reduce emissions and fuel cost such as work by Thales on innovative Flight Management Systems.

5.7.2 Unmanned Aerial Vehicles (UAVs)

There are numerous UAV activities being undertaken within Europe with several major large programmes and many smaller programmes investigating and developing UAV technology for military and civilian use. It is not possible to cover all of these in a report but in this section a few key large programmes are highlighted.

Figure 37. Watchkeeper UAV [186]

The Thales Watchkeeper WK450 [186] (See Figure 37) is a remotely piloted air system (RPAS) for all weather, Intelligence, Surveillance, Target Acquisition and Reconnaissance (ISTAR) which has been developed for use by the British Army, in a €1bn contract awarded to UAV Tactical Systems (U-TacS) in 2005, a joint venture between Thales UK and Israeli Elbit Systems.

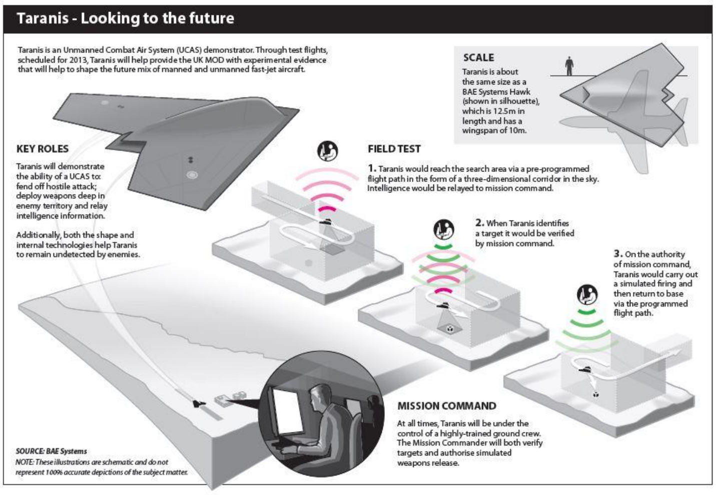


Figure 38. TARANIS UK Military Programme [187]

Named after the Celtic god of thunder, the £185 million Taranis concept aircraft [187] (See Figure 38) is jointly funded by the UK MOD and UK industry. The Taranis demonstrator aircraft was formally unveiled in July 2010. It is about the size of a BAE Systems Hawk aircraft and has been built by BAE Systems, Rolls-Royce, the Systems division of GE Aviation (formerly Smiths Aerospace) and QinetiQ working alongside UK MOD military staff, scientists and other smaller companies. The Taranis demonstrator is the result of one-and-a-half-million man hours of work by scientists, aerodynamicists and systems engineers from 250 UK companies. The aircraft was designed to create an unmanned air system which is capable of undertaking sustained surveillance, marking targets, gathering intelligence, deterring adversaries and carrying out strikes in hostile territory. The aircraft has low observability, high levels of systems integration, supporting control infrastructure and full autonomy elements. The aim of TARANIS was to help the UK MOD and Royal Air Force make decisions on the future mix of manned and unmanned fast jet aircraft and how they will operate together in a safe and effective manner.

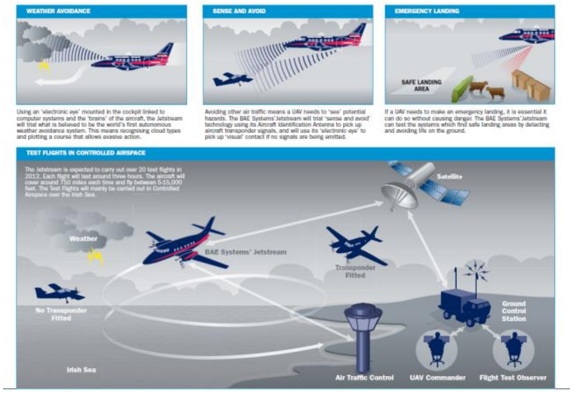


Figure 39. ASTRAEA UK Civil UAV Programme [188]

The ASTRAEA (Autonomous Systems Technology Related Airborne Evaluation & Assessment) programme [188] (See Figure 39) is a UK industry-led consortium focusing on the technologies, systems, facilities, procedures and regulations that will allow autonomous vehicles to operate safely and routinely in civil airspace over the UK. The aim of the ASTRAEA programme is to enable the routine use of UAS (Unmanned Aircraft Systems) in all classes of airspace without the need for restrictive or specialised conditions of operation. This will be achieved through the coordinated development and demonstration of key technologies and operating procedures required to open up the airspace to UAS. The programme is a £62 million effort led by seven companies: Autonomous Decision Making SOftware (AOS), BAE Systems, Cassidian, Cobham, Qinetiq, Rolls-Royce and Thales. The programme consists of two separate projects, one addressing *Separation Assurance & Control* that covers the particular technologies required to control the flying vehicle in the airspace considering the ground control station, the spectrum, security and integrity of the communication system and the vehicle's sense and avoid sensor system, and the second addressing *Autonomy & Decision Making*, providing the intelligence within the vehicle through a variable autonomy system that shares decision making for the mission and contingency management with the human operator.

5.8 Marine Transport – Drivers and Initiatives

By far the most efficient mode of transport for the movement of goods, the shipping sector is expected to grow by 150-250% over the next 30 years. Here integrated world-wide ship management systems are being linked with ship fouling efficiency metrics and navigation systems to optimise performance to reduce fuel consumption and emissions. The introduction of emissions monitoring [189] has led to new operational approaches such as “slow steaming” to reduce emissions for products that are not time critical. Logistically there are complex interactions in the movements of containers around the world to ensure that shipping and handling costs are minimised with tight linkage into the appropriate rail or road haulage network to move the goods onwards as quickly and efficiently as possible.

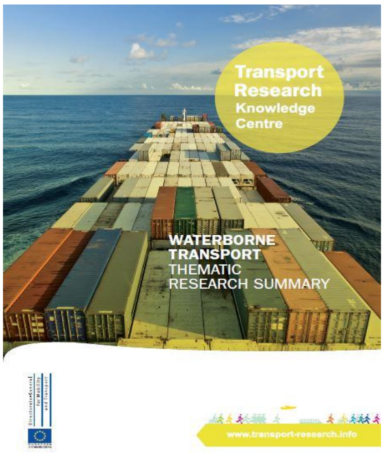
European shipbuilders are world market leaders by turnover [124]. In particular Europe produces nearly all the high value cruise ships in the world, around 50% of all equipment suppliers' products are exported outside Europe and almost 100% of the dredging technology and know-how is European. From a fleet management perspective around 40% of the world merchant fleet is controlled by European companies and approximately 25% are flying the European EEA flag. Of the top 5 world ports, 3 are European and the European Oil & Gas Service Industry is also a world technology leader, exporting 70% of products. The European maritime industry is spearheading environmentally friendly technologies, e.g. European equipment suppliers have provided on-board total waste management systems ahead of future environmental regulations.

A key driver in the maritime industry is improving safety of waterborne operations. This is because recent maritime disasters and accidents in inland navigation have shown that accidents come with high costs in terms of loss of life, environmental damage and with high economic impact. Additionally, high profile accidents have tarnished the overall image and public perception of the waterborne sector. With the increase in cargo traffic in busy North Sea lanes there is a need to maintain safe operations of cargo vessels. At the same time passenger cruise ships have also got bigger and are now operating in non-traditional, remote and difficult regions such as the artic with new and increasing risks. Passengers have high expectations for comfort and a wide range of on-board amenities. The current research drivers are to develop and demonstrate innovative solutions for ship design and waterborne operations to avoid and mitigate passenger risks and ensure high levels of safety.

The industry believes that new technologies for maritime traffic management will be key for safer and more secure operations. In the marine sector there is great interest in optimised shipping operations and voyage optimisation, condition based maintenance, reducing costs and reducing emissions. Local legislation has resulted in emissions monitoring being introduced in ports and local governments have introduced their own requirements. The drivers are for reduced maintenance, enhanced asset life, reduction in crewing levels through increased automation and fleet optimisation via shore based decisions. Key enablers in the industry are the introduction of VSAT systems [190] that allow much greater data rates for data transfer.

There is also a drive for a more integrated transport chain. To reduce congestion in ports and port fairways, port traffic guidance systems need to be at the same time cost efficient and easily deployable. Synergies with existing systems should be ensured, with the aim of integrating the use of port traffic guidance tools by all relevant authorities and ensuring the full interoperability between Information and Communication Technologies (ICT) systems, which monitor vessels, freight and port services.

5.8.1 Waterborne



- Sustainable Waterborne Transport
  - Assuring security of supply
  - Increasing the energy efficiency of ships and vessels
  - Minimising the environmental impact of ships and vessels
  - Building safer ships and vessels
  - Increasing competitiveness
  - Recruiting and retaining a skilled workforce
  - Developing advanced waterborne infrastructure including e-maritime solutions
- Support for the harvesting of offshore resources
  - Renewable energies: wind, wave and tidal energy
  - Fossil fuels and raw materials
  - Fisheries and aquaculture
- Minimising impact on the oceans
  - Developing a better understanding of the oceans
  - Improving products and services for marine monitoring
  - Increasing direct collaboration with the marine sciences

Figure 40. WATERBORNE European Technology Platform Research Summary [192]

In Europe the marine industry has come together with the aim of providing sustainable waterborne transport for the future. To support this the WATERBORNE European Technology Platform [191] has been created with the aims highlighted in Figure 40 [192]. The WATERBORNE initiative was driven by the Maritime Industries Forum (MIF) and its R&D Committee in 2005. Waterborne identifies R&D requirements for European competitiveness in the industry and the innovation and research needs to meet new regulations for safety and environmental goals. The programme is very wide with stakeholder coverage from deep and short sea shipping, inland waterways, ship yards, equipment manufacturers, the marine leisure industry, research and university institutions and classification societies. In addition to a stakeholder Support Group, there is a Mirror Group of government appointed delegates. The WATERBORNE TP published a Vision 2020 paper in 2012 [193], a Strategic Research Agenda in 2011 [194] and an Implementation Plan in 2011 [195]. These documents are being used by the European Commission to direct calls under the R&D workprogrammes and also national R&D programmes. They are also being used by industry to guide research and development.

5.8.2 Marine Vision 2020 and Strategic Research Agenda

Although waterborne transport is the most sustainable, fuel-efficient and environmentally friendly transport mode, special consideration is required of the consequences of accidents, particularly in sensitive coastal areas. As a mode of transport ships have the capacity to transport very large quantities of cargo or large numbers of passengers, consequently there must be a strong focus on safety and also environmental protection. The development and expansion of port capacity is required but care must be taken to preserve natural habitats in surrounding areas. Additionally, security is increasingly an issue and operations must be protected against threats. The EU shipping industry working with public authorities has progressively enhanced safety at sea and introduced greater measures to protect the environment and is actively promoting this on an international basis. The Vision 2020 paper [193] produced by the industry highlights the needs for effective designs, systems, procedures and techniques to increase the level and reliability of the ship system's performance, with the goal of a "zero accident" record. To do this there is a need for research into:

- Effective means to avoid accidents
- Robust ships and reliable equipment
- Improved survival in extreme conditions (ice, freak, waves, etc.)
- Competent crew, ship management and shore operations

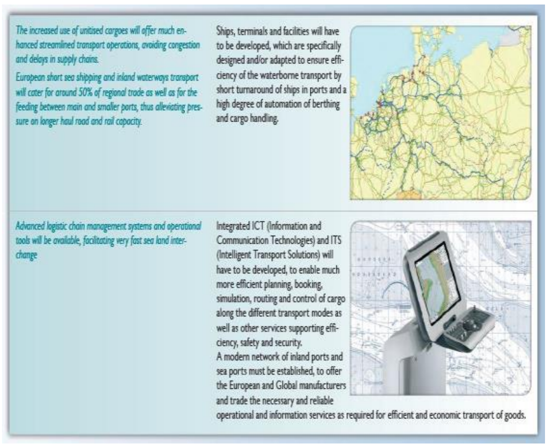


Figure 41. Innovation Challenges: Traffic Management, Integrated Supply Chains, Port Efficiency [193]

Supporting this there is a drive to improve man-machine interfaces and decision support systems to minimise impact of human error. Ships built in Europe will be equipped with on board systems for performance monitoring (See Figure 41) with the aim of reducing life cycle maintenance and also providing safer operation. This will include monitoring and failure prevention strategies and systems for corrosion and wear monitoring. This requires the development of predictive maintenance and inspection capabilities to support the whole life cycle. For safer operations cheap, fool-proof and safe communication and identification equipment needs to be developed to support smaller coastal craft (e.g. fishing and recreational craft, craft with amateur crew) that can be integrated within traffic management systems. This is required if a political decision driven by safety is made to include all small coastal craft in traffic management systems. Additionally, safe and efficient data models and algorithms will be required to cope with the expected huge numbers of traffic participants. If no political decision is made safety is still a concern and so alternative safe and user friendly strategies to traffic management also need to be developed.

Emissions are another key issue and a 'zero emission' approach, to SOx, NOx, CO_2, PM, VOCs presents a technological and economic challenge particularly as approaches to reducing one pollutant tend to increase the emissions of other pollutants with different options being needed for a variety of ships. The Waterborne Strategic Research Agenda [195] focuses on three themes:

- Safe, sustainable and efficient waterborne operations
- Maintaining a competitive European maritime industry
- Managing and facilitating growth and changing trade patterns

These are subdivided into topic area such as short-sea shipping, inland waterways, ship design, operation, and maintenance, maritime safety and ports and port operations.

Under Horizon 2020 a major call was made [196] supporting innovation actions to develop safer and more efficient waterborne operations through new technologies and smarter traffic management. The key areas related to ICT and Systems of Systems are:

- New and improved systems for the surveillance, monitoring and integrated management of waterborne transport and other activities (commercial and non-commercial).
- New and cost effective European Global Navigation Satellite System (European GNSS)-based procedures for port approach, pilotage and guidance, ICT-enabled shipping lanes and maritime services that will reduce the risk of accidents and incidents in port approaches and dense traffic lanes, and minimise both delays and turn-around times.
- For traffic management, solutions that support the extension, integration and optimisation of waterborne transport information and communication systems with the aim of contributing to build a comprehensive "e-maritime" environment (including e-Navigation components that are compatible with existing or emerging international standards). The objective here is to build a “European Maritime Transport Space without Barriers” allowing waterborne transport (including inland navigation) to be used to the full potential within an integrated intermodal logistic chain.

Of particular note is that the call is asking for solutions that will also provide the foundation for the deployment of autonomous and actively guided ships as well as the possibility to verify all related safety certificates before a vessel enters the port. This is to support the future long term goal to reduce crew numbers still further and move towards autonomous and actively guided ships. In parallel with the research activities there is a need to also provide inputs into EU and international regulatory regimes. An aim is to promote standardisation and international research co-operation particularly in the areas of safety devices and e-Navigation solutions.

5.8.3 e-Maritime

The EU e-Maritime initiative [197] aims to foster the use of advanced information technologies for working and doing business in the maritime transport sector. Maritime transport administrative procedures are complex, time-consuming and, even today, are quite often done on paper so there is a need to embrace modern ICT technologies and ways of doing business. Major European ports have deployed advanced information systems. These deliver considerable quality and efficiency gains, however, currently there is no interoperability between port information systems. This limits the potential for new services and economies of scale. Small ports may not have any electronic data transmission capabilities at all. The usual practice is that shipping companies at each port manually enter the same data repeatedly, resulting in duplication and errors.

For the next generation of sailors (the "Internet" generation) access to cyber-space is a must. e-Maritime aims to stimulate coherent, transparent, efficient and simplified solutions in support of cooperation, interoperability and consistency between Member States and transport operators.

Going beyond e-Maritime there is also an activity to provide a mechanism for a Common Information Sharing Environment (CISE). This is currently being developed jointly by the European Commission and EU/EEA member states [198]. It will integrate existing surveillance systems and networks and give all concerned authorities access to the information they need for their missions at sea. CISE will make different systems interoperable so that data and other information can be exchanged easily.

5.8.4 Unmanned Ships

Figure 42. Rolls-Royce Unmanned Ships Concept [199]

Modern ships are operated with much lower numbers of crew than in the past. This has been achieved by introducing much greater levels of automation and also through more advanced on-ship monitoring systems. Rolls-Royce's Blue Ocean development team has set up a virtual-reality prototype that simulates 360-degree views from a vessel's bridge [199]. The idea is that eventually captains on dry land will be able to use similar control centres to command hundreds of crewless ships. Drone ships would be safer, cheaper and less polluting for the shipping industry (See Figure 42). The European Union is funding a €3.5 million study called the "Maritime Unmanned Navigation through Intelligence in Network" project which will produce a prototype for simulated sea trials to assess costs and benefits. The Rolls-Royce design for an autonomous ship (See Figure 42) has no bridge with just containers from front to back. By replacing the bridge and the systems that support the crew, e.g. electricity, air conditioning, water and sewage, the ships can be 5 percent lighter before loading cargo and would burn 12% to 15% less fuel. Additionally, from a financial perspective figures show that a crew costs $3,299 a day and account for about 44 percent of total operating expenses for a large container ship.

There are considerable hurdles to adoption of unmanned ships coming from regulators who are concerned about safety and unions who are concerned about job losses. In fact current regulations dictate minimum crew levels by international conventions. The country where a ship is registered is responsible for regulating vessels within its own waters and for enforcing the international rules. The international IMO regulations apply to seagoing vessels trading internationally and exceeding 500 gross tons, except warships and fishing boats. If drone ships do not comply with the IMO rules, they would be considered unseaworthy and ineligible for insurance. There is, however, interest in deployment of unmanned ships in the Baltic Sea. The expectation is that computers will gradually increase their role in navigation and operations reducing crew levels further. Container ships and dry-bulk carriers are the most likely first candidates for total autonomy as tankers carrying hazardous materials such as oil and liquefied natural gas will probably remain manned longer because of the perception that having a crew on board is safer.

To successfully replace crews unmanned ships will need constant and comprehensive computer monitoring to anticipate failures in advance and "redundant" systems to maintain availability. Computer systems can also be used to analyse ship information and optimise performance. Cameras and sensors can already detect obstacles in the water far better than the human eye. Of particular note is that human error causes most maritime accidents which are often related to fatigue. Unmanned ships would also reduce risks such as piracy, since there would be no hostages to capture, however, ships would become vulnerable to a different kind of piracy from computer hackers.

5.9 US Transportation Drivers and Policy Initiatives

The drivers in the US for intelligent transportation are similar to those in Europe, however, another key driver is homeland security. There is a desire to provide surveillance of roadways and also a means for mass evacuation of people in urban areas as a result of natural disaster or threat. Each state has an Intelligent Transportation Systems chapter that holds a yearly conference to promote and showcase ITS technologies and ideas. Representatives from each Department of Transportation (state, cities, towns, and counties) within the state attend this conference. The Department of Transportation (DOT) has defined a set of goals for the US transportation system:

- Reduce or eliminate deaths and serious injuries among all users of the transportation system – drivers, passengers, cyclists, and pedestrians
- Increase the reliability and efficiency of the transportation system – for the movement of both people and goods
- Drive innovation in the development of safe, affordable mobility options for all Americans
- Increase the service life and optimise the maintenance of transportation structures in a state of good repair
- Reduce the environmental and energy impacts in the development, operation, and maintenance and use of the transportation system
- Increase the resilience of the transportation system to withstand severe weather and climate change impacts.

5.9.1 Intelligent Transport Systems Joint Program Office (US)

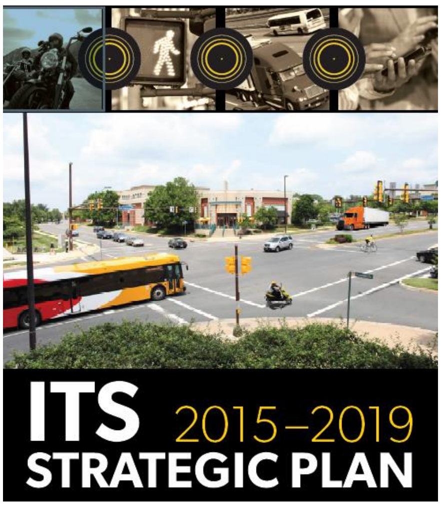


Figure 43. Intelligent Transportation Systems

The ITS Joint Program Office (ITS JPO) [200], within the Office of the Assistance Secretary for Research and Technology (OST-R) has responsibility for executing the “Subtitle C- Intelligent Transportation System Research of Public Law 109-59 Safe Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users, enacted August 10, 2005”. Specifically the ITS-JPO will:

“Conduct an ongoing intelligent transportation system program to research, develop, and operationally test intelligent transportation systems and to provide technical assistance in the nationwide application of those systems as a component of the surface transportation systems of the United States”

The office works with the Federal Highway Administration, Federal Motor Carrier Safety Administration, Federal Railroad Administration, Federal Transit Administration, Maritime Administration, and the National Highway Traffic Safety Administration to plan, programme, and execute the ITS Research Program (See Figure 43). The focus of the programme is on vehicle-to-vehicle and vehicle-to-infrastructure connectivity through the application of advanced wireless technologies. The ITS Research Program specifically develops and tests the underlying technology and applications.

5.9.2 State Smart Transportation Initiative (SSTI)

The State Smart Transportation Initiative [201] promotes transportation practices that “advance environmental sustainability and equitable economic development, while maintaining high standards of governmental efficiency and transparency”. The SSTI, housed at the University of Wisconsin, operates in three ways:

- As a community of practice, where participating agencies can learn together and share experiences as they implement innovative smart transportation policies.
- As a source of direct technical assistance to the agencies on transformative and replicable smart transportation reform efforts.
- As a resource to the wider transportation community, including local, state, and federal agencies, in their efforts to reorient practice to changing social and financial demands.

The SSTI and Smart Growth America (SGA) has produced a handbook [202] that provides 34 specific recommendations to help state transportation officials position their agencies for success.

5.9.3 Smart Growth America

Smart Growth America [203] is a national organization dedicated to researching, advocating for and leading coalitions to bring smart growth practices to communities nationwide. The aim is to use smart growth strategies to create transportation systems for businesses and communities. This includes coverage of transit options like buses, trolleys, subways, light rail, street cars and ferries. It is driven by the need to accommodate more travellers in the same space and create better options for travel between home, jobs and stores. There is also an aim to make neighbourhoods safer and more appealing. Notably there is a drive to avoid new construction, cut expenses and improve the environment. Communities can provide travel choices by making it easy for residents and visitors to drive, walk, bike, or take public transport.

5.9.4 Smart City Challenge

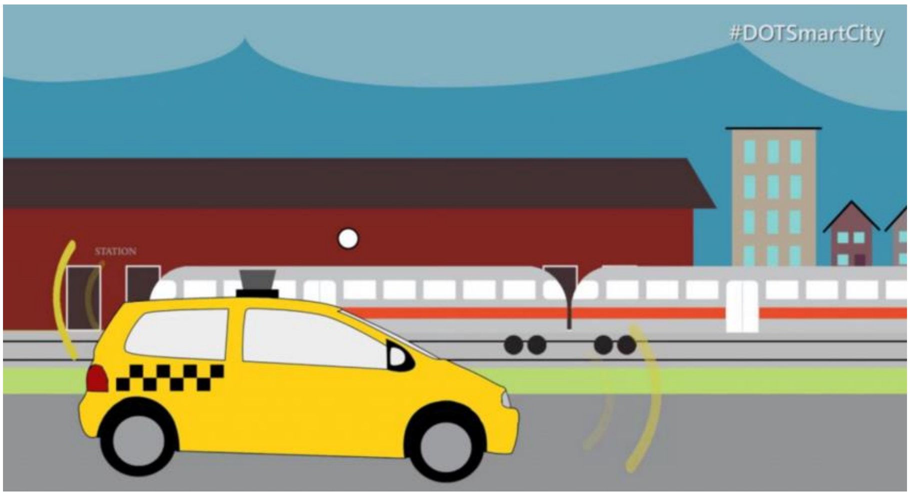


Figure 44. Smart City Challenge

The United States Department of Transportation (DOT) has launched a Smart City Challenge [204] (See Figure 44). This is targeted at mid-sized American cities (200,000 and 850,000 residents). The DOT will award the winning city with $50 million of funding to implement proposed ideas and create a model for other cities to follow. The winner will be announced in June 2016. The funding will come from a public-private partnership between the DOT ($40m) and Vulcan ($10m) which is an investment vehicle founded by Microsoft co-founder Paul Allen [205]. The aim is to promote Intelligent Transportation Systems, connected vehicles and automated vehicles.

5.9.5 Beyond Traffic

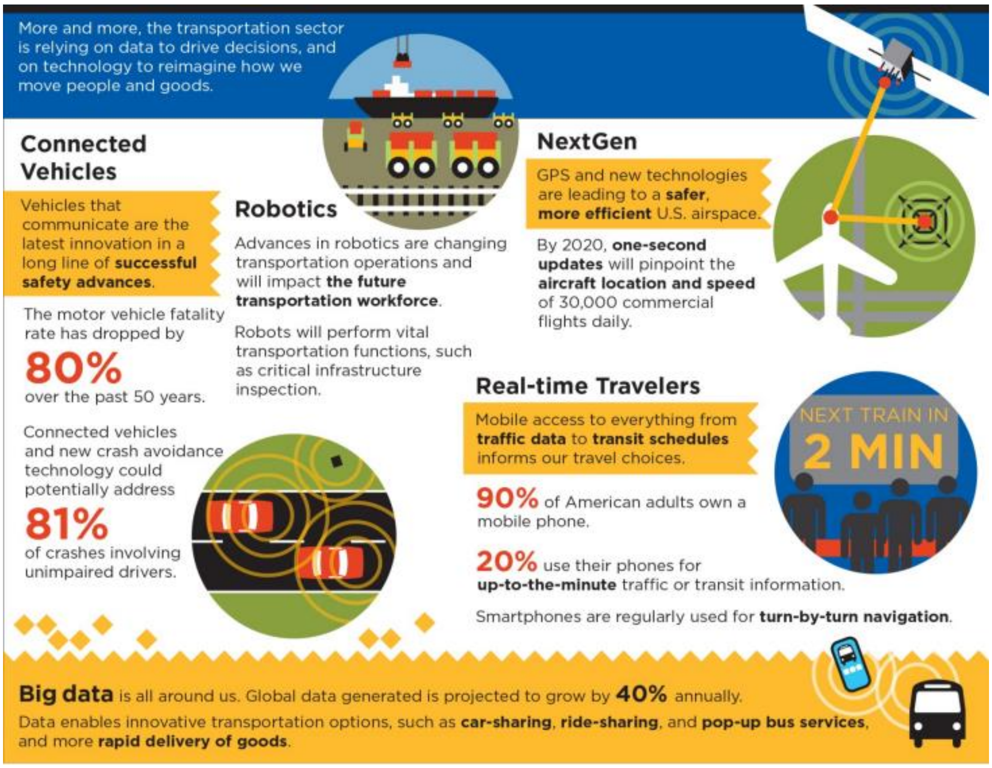


Figure 45. Technology Applications for Transportation

Beyond Traffic [206] is a new forward-looking analysis from the U.S. Department of Transportation outlining the expected trends in the transportation system over the next three decades (See Figure 45). The aim is to promote a national conversation about the future of the U.S. transportation system and objectively frame critical policy choices that need to be made. Beyond Traffic is structured into three parts. The first part discusses the major trends in the transportation system. The second part discusses the implications of these trends for each mode of transportation: highways, transit, pedestrian and bicycle, aviation, intercity and freight rail, maritime and pipeline. The final part presents a future scenario and highlights policy options based on the implications of the trends.

5.9.6 Mobile Millennium

Mobile Millennium [207] was developed by the California Center for Innovative Transportation (CCIT), the Nokia Research Center (NRC), and the University of California (UC) at Berkeley. The partnership began in 2006 when the National Science Foundation co-funded a joint US/European Union workshop in Helsinki. The aim of the work is to use mobile phone and navigation technologies to monitor real-time traffic flow. In 2008 the Mobile Century project performed proof of concept work to test traffic data collection from GPS-equipped cell phones in one hundred vehicles driven on a 10-mile stretch of a highway located in the San Francisco Bay Area. The phones, which effectively served as vehicle probes, stored vehicle speed and position information every three seconds. These measurements were sent wirelessly to a server for real-time processing. The Mobile Century experiment enabled the design and development of algorithms and data collection systems to assemble traffic data from GPS-equipped mobile phones.

The aim of the Mobile Millennium project was to demonstrate the potential of GPS in cell phones to alter the way traffic data is collected, by using the existing cell phone infrastructure to collect data and transmit it directly back to drivers as a 24/7 consumer service. This was demonstrated in the Bay Area and New York City in November 2008 and remained operational until June 2010 with 2000 registered users. Mobile Millennium highlighted some future challenges that need to be addressed by transportation agencies and businesses before similar systems become more commonplace. These challenges include new procurement approaches that are focused on purchasing information rather than equipment, defining the respective roles (and business models) of the public and private sectors in provided traffic information to consumers, and trade-offs between individualised information delivered to a smart phone and distracted driving.

5.9.7 SMART

SMART (Sustainable Mobility & Accessibility Research & Transformation) [208], is a project of UMTRI, the University of Michigan Transportation Research Institute and TCAUP, the Taubman College of Architecture and Urban Planning, in Ann Arbor. The US SMART programme is working with Ford to undertake research and demonstration projects related to the sustainable future of transportation in an urbanizing world. This is driven by the need for sustainable transportation to cope with accelerating urbanization, population growth, globalization, and demographic shifts. Key issues are the environment, energy security, social equity, productivity, urban economies and livability. Recognizing the complexity of the challenge and the sophistication of the innovation required, SMART takes a systems approach to urban mobility. It is a university-wide initiative, working on new theoretical perspectives and practical, innovative, systems solutions.

5.9.8 ITS Deployment – Sensys Networks

Sensys Networks is an American SME developing wireless sensor technologies for ITS. They are working on control of intersections and within the PATH project [209] on “connected corridors” which is concerned with coordinated control of freeways and adjacent urban streets. The timeline for implantation of these systems is 2-3 years with drivers for 24/7 operation, low cost, safety and mobility. A key aim is to provide the ability to get a layered view of operations from network level to individual intersections and compare real time and historical performance of road networks. The expectations are that this will reduce traffic congestion and the associated costs and emissions.

5.9.9 Google and Apple

In general telematics in the automotive sector, even for fleet and insurance operations, has a low uptake because of the cost of retrofitting it to vehicles which is far higher than for factory installed equipment. Customers are not currently asking for connectivity, however, presently there is interest is in providing in-car WiFi so that passengers can connect to services such as Apple CarPlay [210] and those being provided by the Android Open Automotive Alliance [211]. Going one step further Apple is producing wearable computing that connects with cars. The future could well be a “Google Dashboard” and Google are very interested in the automotive industry as collecting information from cars gives free mapping information.

5.9.10 INRIX

Founded in 2005 in Kirkland, Washington and with offices in the UK and Germany INRIX [212] combines data from 1 million miles of roads in North America and 1 million kilometres in 28 European countries to provide services in the car, online and on mobile devices for personal navigation, mapping, telematics and other location-based services. The company has 200 customers and industry partners worldwide including the Ford Motor Company, MapQuest, Microsoft, NAVIGON AG, TeleNav, I-95 Corridor Coalition, Tele Atlas, deCarta, TCS, Telmap, ANWB and ADAC.

Using the services drivers get information on the fastest routes and travel journey times that save time, money and reduce fuel consumption. The traffic data services include accurate real-time and predictive information, real-time incident and weather safety alerts, personalised traffic reports and route advice as well as historical traffic information. For fleet operators the company provides traffic congestion information which can be used to reduce fuel costs and optimise schedule planning. The company also provides information to media broadcasters on traffic congestion and estate agents on actual drive time to and from home and work based on traffic conditions.

5.9.11 Electric Vehicles

Transportation is responsible for 70% of petrol consumption in the US and is the second costliest expense for most American households [213]. President Obama has made historic investments in advanced vehicle and fuel technologies, public transit, and high speed rail under the Recovery Act. One goal was to have one million electric cars on America’s roads by 2015 which was incentivised by tax credits. New fuel economy standards are also being introduced for cars and trucks to raise average fuel economy to 35.5 miles per gallon by 2016. This is predicted to save 1.8 billion barrels of oil over the lifetime of the vehicles subject to the standard. The aim is to lower transportation costs, reduce dependence on oil, revitalise the U.S. manufacturing sector and provide more transportation choices to the American people. Steps have been proposed to improve the efficiency of all modes of transportation, air, road, rail and marine and to develop alternative biofuels

5.9.12 Google Car

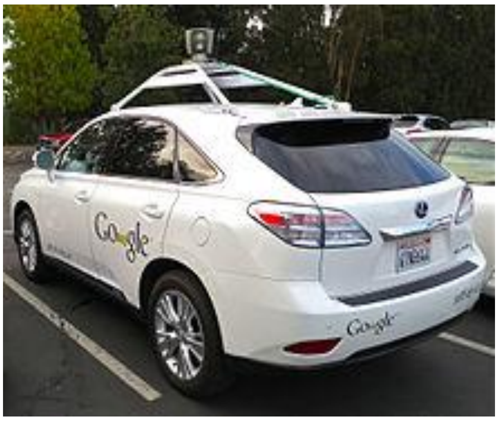


Figure 46. Google Equipped Lexus Autonomous Car and Google Prototype Driverless Car

Google has been working on a Self-Driving Car project [214] for several years to develop autonomous car technologies. This has resulted in the Google Chauffeur software. The company has equipped 6 Toyota PRIUS, 3 Lexus RX450h (as shown in Figure 46) and an Audi TT with $150,000 of equipment to allow autonomous operation. Google has also been active lobbying American states to allow operation of autonomous cars and has been successful in Nevada, Florida and California. By April 2014, the 10 cars had logged nearly 700,000 autonomous miles (1.1 million km). Google have also announced their own driverless car that has no steering wheel or pedals (See Figure 46). Google is not planning on commercial development of the system but they are interested in selling the system and the data behind it to automobile manufacturers.

5.10 Trucking and Logistics - United Parcel Service

UPS perform 17 million shipments per day and are moving from being a trucking company to being a “technology company with trucks” with extensive use of package routing technology and telematics. All road, rail, air and shipping hubs are connected by a private IT system to provide a single network for all categories of service. They have obtained huge savings from network efficiency through development of the ORION (On Road Integrated Optimisation and Navigation) system [215] for predictive and prescriptive operations. This provides an in-cab computer that performs highly complex optimisation of deliveries to minimise miles driven and minutes vehicles spend idling while at the same time maximizing the number of pickups and deliveries made per litre of fuel used. It also gathers information on driving behaviours (which can be used to identify and improve driving performance to reduce fuel consumption) and collects information on mechanical variables from the engine and drive train that can be used for on-condition maintenance saving money and reducing waste (parts, oil, etc.).

Vehicles are used as rolling laboratories and data has been collected for 185 million miles since 2000. Small adjustments to operations can be made with large payoffs over 100,000 drivers around the world. For example, the most efficient vehicles can be matched to routes, the number of stops and starts performed can be minimised and safety can be improved by minimising backing up required in residential areas (which are full of other vehicles, fixed objects, people and pets).

In 2010 telematics-equipped vehicles eliminated more than 15.4 million minutes of idling time saving 103,000 gallons of fuel (and avoiding of 1,045 metric tonnes of CO_2). Additionally, the number of stops per mile were reduced delivering more packages with fewer engine restarts that consume fuel. The use of telematics saved 1.7 million miles of driving in 2010, equating to more than 183,000 gallons of fuel or 1,857 metric tonnes of CO_2. The company is also increasingly using electrification for local deliveries and biomethane as an alternative fuel for larger trucks to further reduce emissions.

5.11 US RAIL

The US lags behind Europe and China in terms of it rail network [216]. The introduction of High-speed rail was one of President Obama's signature transport projects and $11 billion has been spent since 2009 on development of faster passenger trains. However, implementation has been slowed down by Republican and community opposition. The money provided went into upgrading the existing Amtrak service, however, this will only allow trains to go 110 miles per hour and not provide high speed rail services (See Figure 47). Also none of the money went into services in the Northeast Corridor which is the most likely place for a high-speed rail line. Congress has been subsequently asked for a further $10 billion to support high-speed initiatives.

Figure 47. Amtrak's Acela Near Baltimore. The 150 mph Acela Averages Only 80 mph on the New York to Washington Corridor. (Credit Luke Sharrett: The New York Times)

Notably there is opposition from republican governors in Florida, Ohio and Wisconsin, who all canceled high-speed rail projects and returned federal funds after deeming the projects too expensive and unnecessary. The most likely and controversial high speed rail project to be implemented is the 520-mile route between Los Angeles and San Francisco which has begun track construction and put out bids to build the trains. The project will be partly supported by the states cap-and-trade programme which requires business to pay for excess pollution.

In Florida, a 125 mph train will be introduced by a private company, All Aboard Florida, but this will operate at much slower speeds between Miami and West Palm Beach, with a stop in Fort Lauderdale. An extension to Orlando is also planned. Although private, the builders have applied for a $1.5 billion loan from the Federal Railroad Administration which must be paid back with interest over 25 years. Several counties along the route, however, are opposing the project.

In Texas, the private Texas Central Railway company, has proposed a high-speed rail line by 2021 with trains that could reach speeds of up to 205 mph using Japanese bullet trains. This would cut the trip between Houston and Dallas to 90 minutes.

5.12 US Aerospace

5.12.1 NEXTGEN – Air Traffic Control USA

The equivalent of SESAR in the US is NEXTGEN [217]. The National Airspace System (NAS) is the collection of all the components (airspace, facilities, equipment, services, workforce, procedures, etc.) that enable the US air transportation system. Within the United States one organization, the FAA, operates 350,000 airplanes and 18,000 landing facilities. NEXTGEN, is short for the “Next Generation Air Transportation System,” and is a programme to comprehensively transform the NAS into a system to meet future needs that will be safer, more reliable, more efficient and which will reduce the impact of aviation on the environment. The system is due for implementation across the United States in stages between 2012 and 2025.

There are some fundamental differences in the approaches being taken by the SESAR and NEXTGEN programmes. SESAR’s emphasis is on i4D (using aircraft RTA capabilities – the aircraft themselves calculate the Required Time for Arrival) while the FAA’s emphasis is on ADS-B (Automatic Dependent Surveillance - Broadcast) for Interval Management (where aircraft positions are sent by the aircraft and are monitored and controlled centrally). This potentially may lead to a global harmonisation problem for aircraft operators and manufacturers, i.e. multiple solutions for the same operational problem in the same timeframe.

5.12.2 US Unmanned Air Vehicles

Figure 48. Lockheed Martin Unmanned Aerial Vehicles [218]

Lockheed Martin are key proponents of military autonomous vehicles for aerospace, land and underwater marine [218]. Two examples of their autonomous air vehicles are shown in Figure 48. These are the Aerial Reconfigurable Embedded System (ARES) design concept which was developed as part of the Transformer (TX) programme in 2009 [219]. Transformer aimed to develop and demonstrate a prototype system that could provide flexible, terrain-independent transportation for logistics, personnel transport and tactical support missions for small ground units. The other is the K-Max unmanned cargo helicopter which is designed to keep forward operating bases supplied, reducing the number of truck convoys, and the troops that protect them, on the dangerous roads of Afghanistan. Here difficult terrain and threats, such as ambushes and Improvised Explosive Devices (IEDs) can make ground-based transportation to and from the frontline a dangerous challenge. While manned helicopters can easily bypass those problems, they often present logistical challenges of their own and can subject flight crews to different types of threats. Additionally Lockheed Martin also make smaller Unmanned Aerial Systems such as Desert Hawk III that enables soldiers to see what is over the next hill and “persistent surveillance” platforms like PTDS, High Altitude Airships, Hybrid Air Vehicle and ISIS [219] to keep “eyes in the sky” over large areas for weeks, months and even years at a time.

5.12.3 Amazon Prime

Figure 49. Amazon Prime Air [220]

In the civilian domain Amazon are developing their Prime Air concept [220] with drones that deliver packages to customers (See Figure 49). The main goal of the new delivery system is to get packages into customers' hands within 30 minutes or less using unmanned aerial vehicles. Actual deployment of such systems is still many years away as a major barrier is gaining approval from the FAA for operations. This would be even more complicated for operations in Europe due to the more fragmented nature of airspace control. Amazon do not see this as an issue but rather an opportunity to work on the underlying technology and improve payload and endurance which are presently very limited.

5.12.4 US Maritime

In 2009 IHS Global Insight completed a study for the Maritime Administration of the US Department of Transportation, titled “An Evaluation of Maritime Policy in Meeting the commercial and Security Needs of the United States” [221]. This highlighted that US maritime policy supported the nation's domestic maritime trades but did not support international shipping trade and policy reforms were needed. The Maritime Administration has reacted by creating offices at major US gateway ports (See Figure 50), starting with 10 of the largest ports on the West, East and Gulf Coasts, the Great Lakes and the inland river system. These offices interact with stakeholders, including headquarters staff, state and local authorities, port operators, shippers and carriers to identify Federal and state funding and cooperate on planning, environmental and community projects.

Figure 50. Gateway Office Locations in the US

The aim is to identify bottlenecks and ways of improving freight movement. For instance, the Gateway Office in Southern California has worked with public and private sector participants to tackle congestion and better understand the connection between improved cargo flow, economic vitality, community improvement and environmental sustainability. The gateway in Anchorage is leading a public-private partnership with the Port of Anchorage to redevelop its port complex to enhance the transportation of goods within the state. This includes construction of new berths and piers, introduction of new container cranes, on-site rail and railroad trailers, and development of a modern container yard to improve efficiency and reduce truck traffic. This approach to port revitalization is expected to be replicated at other U.S. ports. Public Private Partnerships are seen as the key way of funding and developing ports within America [222].

Here there are major infrastructure funding initiatives such as the Transportation Infrastructure Finance and Innovation Act (TIFIA) programme that provides direct loans, loan guarantees and standby lines of credit as a result of the Fixing America's Surface Transportation Act (FAST Act) with $1.435 billion in capital over five years. This can be used for surface transportation infrastructure including highways, passenger and freight rail, port access, public transit, intermodal freight facilities and international bridges and tunnels. To date, the TIFIA programme has provided $22.7 billion in credit assistance to support more than $82.5 billion in transportation infrastructure investments to help build 56 major transportation projects around the country [223].

At the maritime industry level the American Maritime Partnership (AMP) represents the domestic industry. It has 450-plus members including vessel owners and operators, shipboard and shoreside workers, shipbuilders and repair yards, equipment manufacturers and vendors, dredging and marine construction contractors, numerous maritime associations and national security organizations. A strong domestic maritime industry is seen as being critical for America's economic, national, and homeland security. This is supported by maintaining the Jones Act [224] as the foundation of America's domestic maritime policy. This requires that any vessel transporting goods or passengers between two points in the United States or engaging in activities in US waters must be US owned, US built, and US crewed.

5.13 Rest of the World

MarketsandMarkets [225] forecasts the smart transportation market to grow from $46.72 Billion in 2015 to $138.76 Billion by 2020, growing at a Compound Annual Growth Rate (CAGR) of 24.3% from 2015 to 2020. North America and Europe are expected to be the largest markets in terms of revenue contribution. The Asia Pacific (APAC) region, however, is expected to show the highest potential growth due technological advancements and the presence of leading smart transportation vendors. Additionally, the Middle East and Africa (MEA) and Latin America (LA) markets are also showing a growth in smart transportation. The report also describes the competitive landscape of the smart transportation market. This includes a comparative analysis of the technological and marketing strategies that the key players are adopting. Notably at present the key strategies followed are providing customised solutions or buying smaller domain expert companies. The leading players in this market are Cisco Systems, Inc., General Electric Company, IBM Corporation, Accenture PLC, Indra, Cubic Corporation, Kapsch, and LG CNS.

5.13.1 Worldwide ITS

A Network of National ITS Associations was officially launched on 7 October 2004 in London [226]. This Network is a grouping of national ITS interests formed in order to ensure that ITS knowledge and information is transmitted to all actors at the local and national level. The Network currently consists of 27 member organisations. The Network Secretariat is at ERTICO-ITS Europe and is a multi-sector, public/private partnership pursuing the development and deployment of Intelligent Transport Systems and Services. It connects public authorities, industry players, infrastructure operators, users, national ITS associations and other organisations together and works to bring “Intelligence into Mobility”. The ERTICO work programme [227] focuses on initiatives to improve transport safety, security and network efficiency whilst taking into account measures to reduce environmental impact. The vision is of a future transport system working towards zero accidents, zero delays with fully informed people, where services are affordable and seamless, the environment is protected, privacy is respected and security is provided. At a worldwide level the ITS World Congress [228] is an annual event to promote and showcase ITS technologies organised by ERTICO – ITS Europe, ITS America and ITS Japan. This event attracts over 8,000 people.

5.13.2 Asia-Pacific Economic Cooperation (APEC)

The Smart Transport pillar of the Energy-Smart Communities Initiative [229] will examine clean and efficient ways of moving both goods and people throughout the Asia-Pacific region. Work will focus on options to reduce travel times, costs, energy use, and carbon emissions for urban and freight transport. A number of multinational projects have been proposed across the region:

- Energy-Efficient Urban Transport Network (Australia, Canada, Indonesia, Singapore, Chinese Taipei, US)
- Energy-Efficient Freight Transport Network (Australia, Indonesia, Chinese Taipe, US)
- Electromobility Survey And Road Map (Australia, Canada, Hong Kong, China, Indonesia, Japan, Malaysia, Singapore, Chinese Taipei, US)

- Electric Vehicle Demonstrations (Canada, Indonesia, Japan, Malaysia, The Philippines, Chinese Taipei, US)

5.13.3 Japan

In Japan there is great interest in car-to-car, car-to-infrastructure, and infrastructure-to-car communication. This is being used for warning drivers of upcoming hazards. Already Japan has installed sensors on highways which are used to notify motorists that a car is stalled ahead. Transmission of car data to infrastructure opens up the opportunities to centrally fuse and process data to detect events such as rain (wiper activity), congestion (frequent braking activities) and ice detection (from ABS activations). Transmission from infrastructure-to-car can be used to provide driver recommendations to avoid traffic or warn of hazards increasing road safety.

5.13.4 India

The US Department of Transportation is offering sustainable transport solutions for the Indian cities of Allahabad, Ajmer and Visakhapatnam [230]. This is a bilateral cooperation in the field of smart cities including efficient public transportation systems, intelligent transport systems, traffic information and control, multimodal integration and capacity building and training in the field of urban transportation. Task forces have been set up for Allahabad, Ajmer and Visakhapatnam in association with United States Trade Development Agency. The US Government is focusing on promoting regional transport solutions and the planned initiative will build 100 smart cities in India.

5.14 ICT Regulations

5.14.1 eCall

EU regulation is driving the introduction of car emergency vehicle notification systems (eCall) [231]. Additionally, insurance companies are interested in introducing driver behaviour tracking functionalities. Using the system in an emergency the vehicle occupants can manually eCall or the vehicle can automatically call via activation of in-vehicle sensors after an accident. The eCall device establishes an emergency call carrying both voice and data directly to the nearest emergency point. The voice call enables the vehicle occupant to communicate with the eCall operator. At the same time, data is sent containing information about the incident, including time, precise location, the direction the vehicle was traveling, and vehicle identification. The pan-European eCall system aims to be operative for all new type-approved vehicles as a standard option. Depending on the manufacturer of the eCall system, it could be mobile phone based (Bluetooth connection to an in-vehicle interface), an integrated eCall device, or a functionality of a broader system like navigation, telematics device, or tolling device. Going one stage further the EC funded project SafeTRIP [232] is developing an open ITS system that will improve road safety and provide resilient communication through the use of S-band satellite communication. This would allow greater coverage of the Emergency Call Service within the EU. Work on the eCall standard has been ongoing for a number of years and currently it is targeted for implementation in 2017. This has been slowed by lack of support for it from some member states and currently other technologies are overtaking it combining the same functionality with congestion and traffic management information.

5.14.2 Decarbonisation of Transport

The legal requirements introduced by the European Commission toward decarbonisation of transport have defined an EU global target of 40% reduction of greenhouse gas emissions with a need for 27% of the global energy mix coming from renewable energy sources by 2030. To meet this there is a need to deploy alternative fuel vehicles infrastructures and introduce electrification of cars and public transportation systems. This requires full integration of electric vehicles in urban mobility policies and in the electricity grid, both as energy consumers and potential storage facilities. The European Commission will present a strategic transport R&I agenda by 2016 and under action 11 the EC will propose a comprehensive road package. This will include a number of measures including promoting procurement of clean vehicles, a more efficient pricing of infrastructure, the roll-out of ITS, as well as creating the market conditions to support deployment of alternative fuels.

5.15 Standards

5.15.1 CAR 2 CAR

There is a key need for communication standards. The CAR 2 CAR Communication Consortium (C2C-CC) [127] [233] is a non-profit, industry driven organisation initiated by European vehicle manufacturers and supported by equipment suppliers, research organisations and other partners.

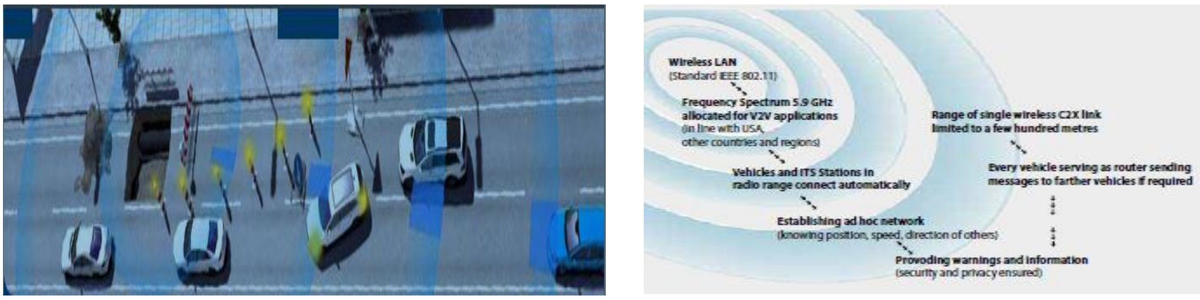


Figure 51. Car2car Communication Consortium [233]

The C2C-CC (See Figure 51) [233] is dedicated to the objective of further increasing road traffic safety and efficiency by means of cooperative Intelligent Transport Systems (C-ITS) with Vehicle-to-Vehicle Communication (V2V) supported by Vehicle-to-Infrastructure Communication (V2I). It supports the creation of European standards for communicating vehicles spanning all brands. As a key contributor the C2C-CC works in close cooperation with European and international standardisation organisations. In cooperation with infrastructure stakeholders the C2C-CC promotes the joint deployment of cooperative ITS. The European Commission forecasts a heavy increase of vehicle kilometres. To improve mobility on the roads it has adopted the ITS Action Plan [234] indicating 24 concrete measures in 6 priority areas.

An iMobility Forum has been created and a number of R&D programmes including large scale field operational tests are being performed. These are expected to contribute to the reduction of road fatalities, improve efficiency and reduce the environmental impact of road traffic in all areas including smart cities.

Figure 52. Partners in CAR 2 CAR

The C-ITS will provide new active safety measures to enhance the existing passive safety systems preventing or mitigating traffic accidents. As energy consumption and emissions need to be reduced and road traffic needs to exploit road capacities, new driving measures are required to support sustainability, e.g. green driving. The aim is to develop an open European standard for C-ITS with an associated validation process focusing on V2V Systems. This is to be supported with realistic deployment strategies and business models to speed up the market penetration and a roadmap for deployment of C-ITS (for V2V and V2I). A key aim is to also provide a validation route for these technologies. Specifications will be provided to the standardisation organisations, in particular ETSI TC ITS, in order to achieve common European standards for ITS. There is also an aim to push for the harmonisation of C2C Communication Standards world-wide such as 802.11.P and promote the allocation of a royalty free European wide frequency band for V2V Applications and joint deployment of C-ITS by all stakeholders.

The mandate M/453 of the European Commission [235] paved the way for the development of the minimum set of standards by the European standardisation organisations ETSI, TC, ITS, and CEN in 2012, ensuring the interoperability of C-ITS**.** Deployment will rely on these standards covering the whole communication chain, including security and privacy issues and starting with day-one applications based on selected common message sets. The project will demonstrate the C2C-System as proof of technical and commercial feasibility.

The standard is close to finalisation and a MOU exists between the major OEMs (See Figure 52). Critically day-one applications will reveal if the standards are sufficient and whether further work is required. The initial applications are concentrated on providing all the information in the car rather than having centralised information being sent to the car so cars will also utilise sensing radar, laser scanners and image detection. Importantly cars must be able to understand each other and not be dependent on the OEM.

5.15.2 POSSE

Project POSSE ("Promotion of Open Specifications and Standards in Europe") is funded under the INTERREG IVC programme. The POSSE project is led by Reading Borough Council in the UK with national co-funding from the Department for Transport towards this project. The project will help share experiences from UK and German open specifications frameworks with roads authorities in a number of Member States (Czech Republic, Italy, Lithuania, Norway and Spain).

5.15.3 ISO ITS Standards

ISO/TC 204, Intelligent transport systems, focuses on standardization of information, communication and control systems in the field of urban and rural surface transportation and covers the areas of traveller information, traffic management, public transport, commercial transport, emergency services and commercial services in the intelligent transport systems (ITS) field.

Other Working Groups (WG) and sub-committees (SC) of ISO are developing technical standards which include:

- WG on Infrastructures of sensor network e.g. ISO/IEC DIS 29182-1 Information technology – Sensor networks: Sensor network reference architecture
- WG on Governance of IT which incorporates the mechanisms, methods, and models which ensure the conformance of IT to underlying and required policies, regulations, laws, and ethical guidelines.
- SC on Telecommunications and information exchange between systems

6 5G

6.1 European Drivers and Policy Initiatives

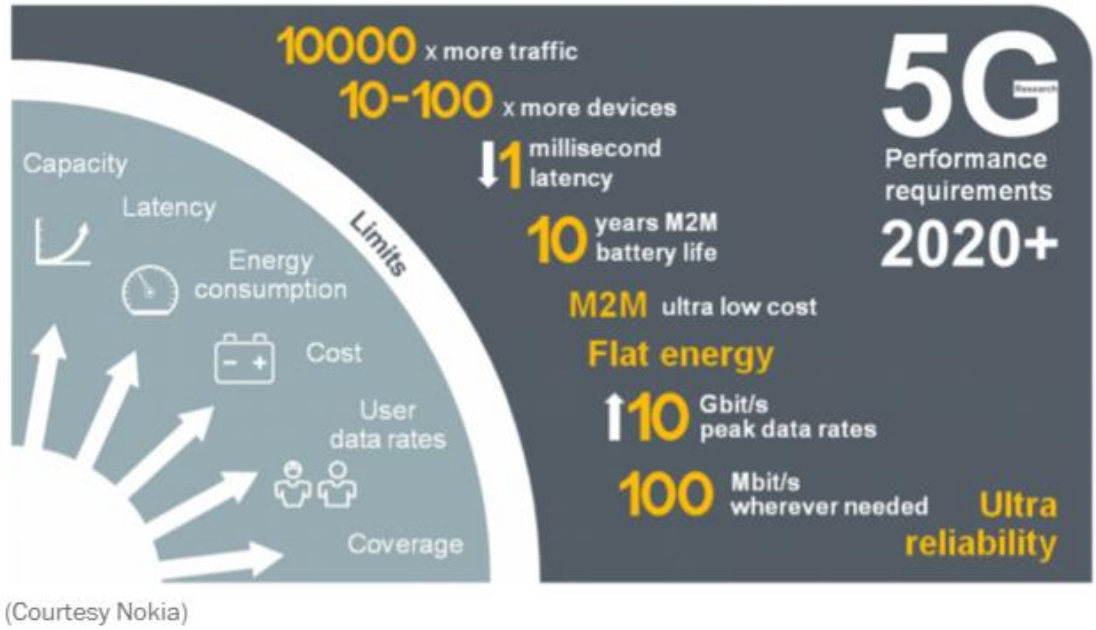


Figure 53. 5G Performance Requirements 2020+

To support the development of smart cities there is a need for substantially improved communication networks. The fifth generation (5G) cellular network (See Figure 53) is a revolutionary change to the traditional cellular network which was utilised for the sole purpose of moving content (voice, video, data, etc.) between two or multiple parties. 5G extends the cellular network from content delivery to a “Control Network” that opens up new doors to new applications such as:

- manoeuver and monitoring of virtual or real time objects (e.g., robots in a factory or hazardous environment),
- controlling smart vehicles on a street (automated traffic control and driving),
- walking using an exoskeletons, remote surgery, (E-health care services), etc.
- new levels of human-IoT interaction including immersive augmented reality and immersive gaming.

However, these new applications also add new stricter requirements on cellular networks such as a need for an end-to-end latency of 1ms for control purposes, extremely high reliability of 99.9999% for controlling cars, very dense connectivity and increased battery life to enable IoT or device-to-device communication. Future networks will also be highly heterogeneous.

Since the demand for the data rate increases by one order of magnitude every 5 years (according to Moore’s law) the 5G network needs to be a massive broad band to cope with increasing throughput requirements. This is estimated to be in excess of 10 Gbps for cellular and above 100 Gbps for short range communications beyond 2020. To meet future data rate requirements more spectrum will be needed. The current cellular network utilises the frequency band below 6GHz but large chunks of contiguous spectrum are only available at higher frequencies from 10 GHz to 100 GHz. There is thus a need to harmonise these spectrum bands globally and provide spectrum access policies, e.g. licensed spectrum access, license assisted access and dynamic spectrum access. A key requirement is a "global standard" based on a consensus between major players around the world most notably in the EU and US. The EU Commissioner specifically highlighted this at the Mobile World Congress 2015 in Barcelona. Multiple white papers have been published on the future of 5G in the EU and US which set out a similar set of requirements for latency, security, reliability and data rates.

The increased ability for interconnectivity and transfer of content will lead to new services and here there are also privacy issues. As new services will also involve car manufacturers, the automation industry, etc. there is a need to engage with key manufacturers to understand requirements, usage scenarios, definitions and business cases.

As part of the Digital Agenda the EC has put forward a number of actions to support the development of 5G [236]. In 2012, the European Commission, under the lead of Neelie Kroes, committed €50 million for research to deliver 5G mobile technology by 2020.

6.2 European 5G Initiatives

6.2.1 The 5G Infrastructure Public Private Partnership

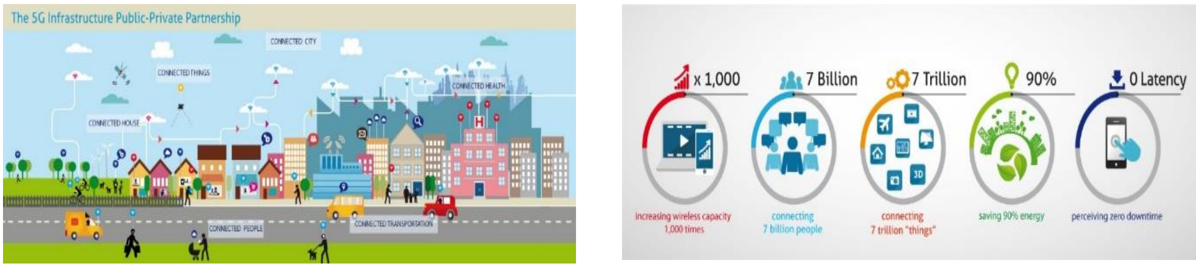


Figure 54. 5G Infrastructure Public Private Partnership

In a report from the GSMA Intelligence [237] it is highlighted that network operators need to move fast with 5G. The EC is supporting this, in particular, by helping to establish PPP structures for 5G development (See Figure 54). To support 5G in Europe the 5G Infrastructure PPP has been set up. The EU is investing €700 million over the next seven years into the 5G PPP through the Horizon 2020 programme. EU industry will match this investment by up to 5 times, to more than €3 billion.

The aims of the 5G Infrastructure PPP are to:

- Provide 1000 times higher wireless area capacity and more varied service capabilities compared to 2010
- Save up to 90% of energy per service provided. The main focus will be in mobile communication networks where the dominating energy consumption comes from the radio access network
- Reduce the average service creation time cycle from 90 hours to 90 minutes
- Create a secure, reliable and dependable Internet with a “zero perceived” downtime for services provision
- Facilitate very dense deployments of wireless communication links to connect over 7 trillion wireless devices serving over 7 billion people
- Ensure for everyone and everywhere the access to a wider panel of services and applications at lower cost

Driving research within Europe is the 5G Roadmap that has been put together by the European Commission and the 5G PPP. This is shown in Figure 55.

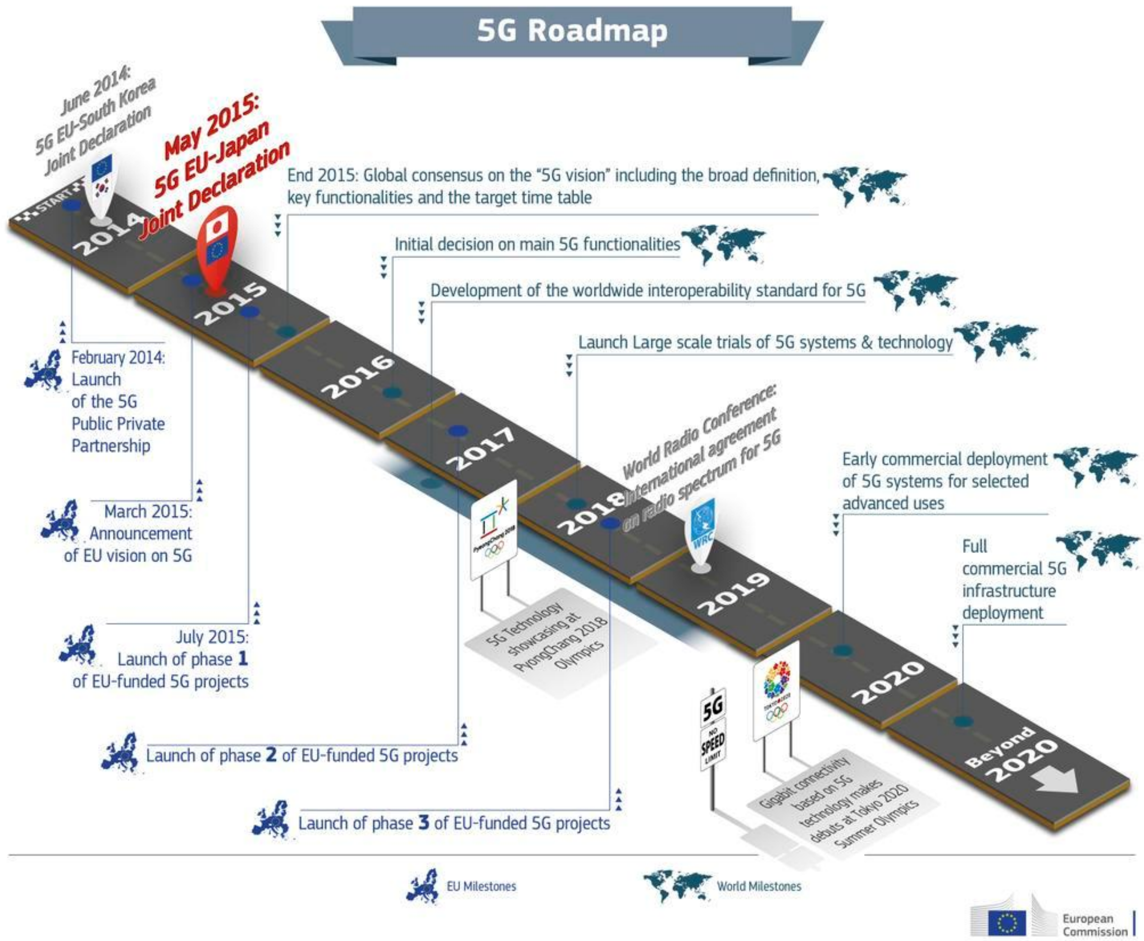


Figure 55. European 5G Roadmap

This roadmap defines stages e.g. definition of 5G functionalities, development of a worldwide interoperability standard and large scale trails of 5G systems leading to 5G standardization and commercial deployment of the technology in 2020. This is being very actively supported through 3 calls funding supporting projects. In the first call for phase 1 a total of 19 5G PP projects were launched and currently Phase 2 proposals are being prepared for 2017.

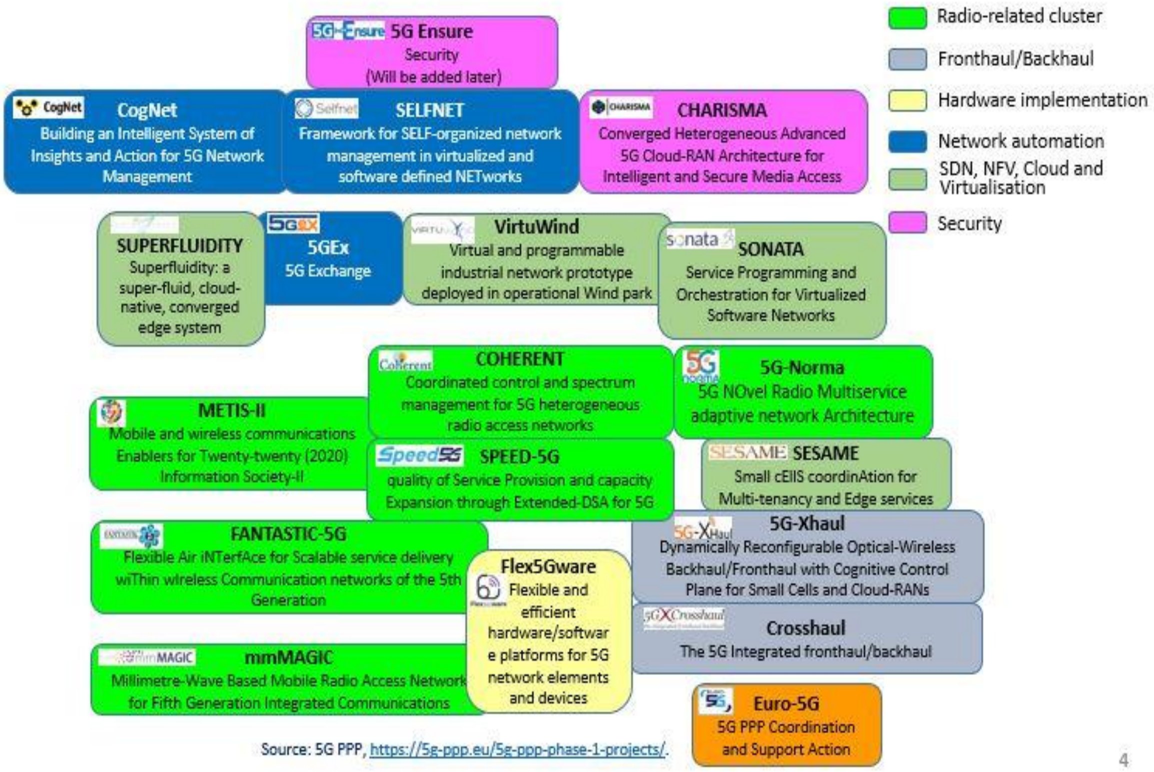


Figure 56. Projects Funded under Phase 1

The projects funded under Phase 1 are targeted at definition of functionalities for 5G. An overview of these is given in Figure 56. It can be seen that the project portfolio addresses a number of key aspects considering the radio technology, fronthaul and backhaul, hardware implementation, network automation, SDN and Cloud interfaces and security. By bringing all of these technologies together and developing the fundamental functionalities for 5G a coordinated approach is being adopted to development across Europe. This is also supported by a specific Coordination and Support Action Euro-5G.

Another key output from the 5G PPP is the development of a White Paper (5G Infrastructure Association: Vision White Paper, February 2016) [238] . In this paper the requirements for key sectors have been gathered.

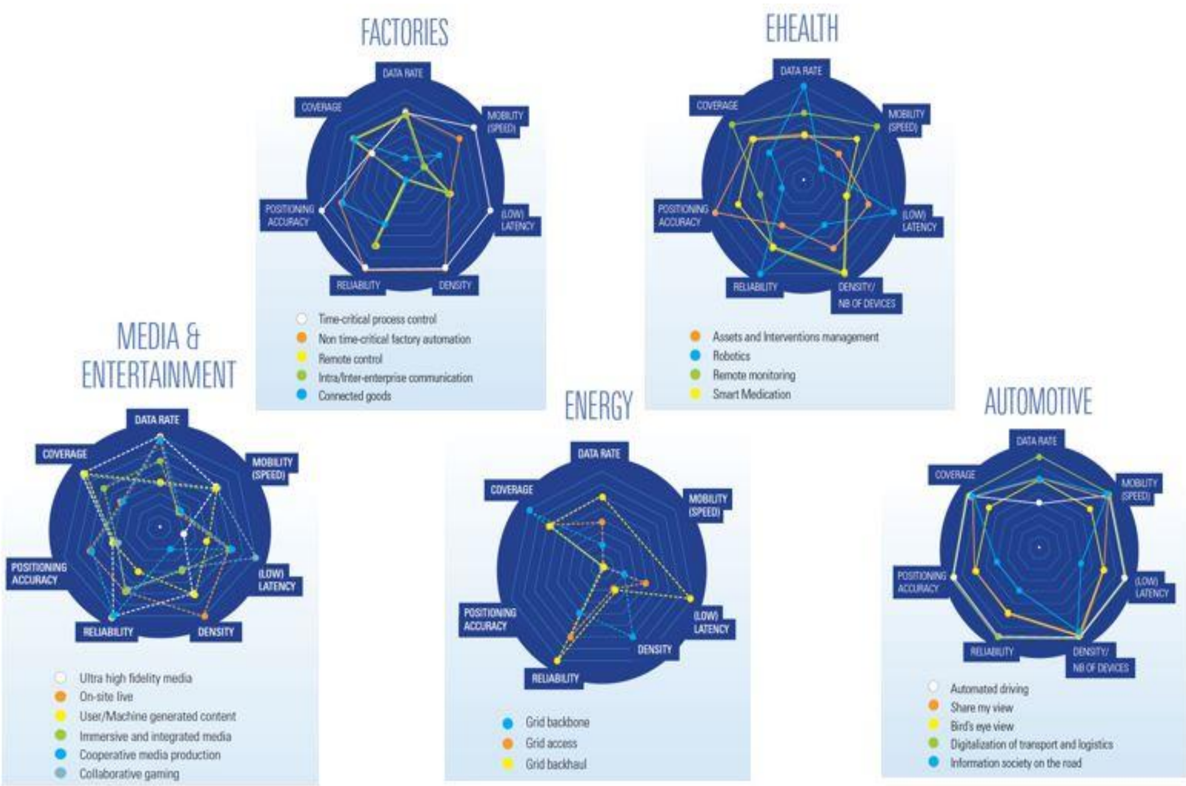


Figure 57. Requirements on 5G from Different Key Sectors

The sectors considered have relevance to the Smart Cities, Smart Energy and Smart Transportation sectors giving the needs for specific identified applications in the domains. In addition to the 5G PPP projects funded under FP7 there are also a number of other projects that have been funded under FP7 and other initiatives. These are briefly described below.

6.2.2 METIS 2020 and METIS-II

The METIS project [239] is driven by several telecommunications companies. A key aim is to reach world-wide consensus on the future global mobile and wireless communications system. The overall technical goal is to provide a system concept that will provide 1000 times higher mobile system spectral efficiency compared to current LTE. The METIS-II [240] project launched as part of the PPP builds on the METIS project and is developing the overall 5G radio access network design and will provide the technical enablers for efficient integration and use of the various 5G technologies and components that have been developed. METIS-II will also provide a 5G collaboration framework within the 5G-PPP for a common evaluation of 5G radio access network concepts and help prepare actions towards regulatory and standardisation bodies.

6.2.3 5GrEEn

The 5GrEEn project [241] (which is linked to METIS and started in 2013) is focused on the design of Green 5G Mobile networks. Here the goal is to develop guidelines for the definition of new generation network with particular care of energy efficiency, sustainability and affordability aspects.

6.2.4 i-JOIN

In November 2012, the FP7 Interworking and JOINt Design of an Open Access and Backhaul Network Architecture for Small Cells based on Cloud Networks (i-JOIN) project [242] was launched. This project coordinated by the IMDEA Networks Institute (Madrid, Spain) is exploring the RAN-as-a-Service (RANaaS). The RAN functionality is flexibly centralised through an open IT platform based on a cloud infrastructure. The aim is to allow joint design and optimisation of access and backhaul, operation and management algorithms, and architectural elements, integrating small-cells, heterogeneous backhaul, and centralised processing. iJOIN is studying the requirements, constraints, and implications for existing mobile networks, specifically 3GPP and LTE-A.

6.2.5 CROWD

The EU funded CROWD (Connectivity management for eneRgy Optimised Wireless Dense networks) project [243] was launched in 2013. Led by the IMDEA Networks Institute it is investigating the design of sustainable networking and software solutions for the deployment of very dense (1000x higher than current density – users per square metre), heterogeneous wireless networks. Sustainability is defined in terms of cost effectiveness and energy efficiency. Heterogeneity arises from a number of factors including coverage radius, the technologies used (4G/LTE vs. Wi-Fi), and deployment strategies (planned vs. unplanned distribution of radio base stations and hot spots).

6.2.6 NetWorld2020 ETP

NetWorld2020 [244] is a European Technology Platform for communications networks and services. Here the communications networks being considered are between users of both mobile and fixed equipment. The NetWorld2020 European Technology Platform brings together industry leaders, innovative SMEs, and leading academic institutions from the communications sector. The NetWorld2020 Expert Group drafted a White Paper on the topic “5G Experimental Facilities in Europe” and a public consultation was held which finished in January 2016. Based on the received feedback the document will be updated and finalised, followed by an endorsement process within the ETP.

6.3 National Initiatives

6.3.1 UK

6.3.1.1 5G Innovation Centre

In September 2015 the University of Surrey opened a 5G Innovation Centre (5GIC) with the aim of securing the UK's role in leading the development of the next generation communications technology. The centre has over 170 researchers and has attracted over £70 million of investment, including £12 million from the Higher Education Funding Council for England (HEFCE). The 5GIC is the world's largest academic research centre dedicated to next generation mobile and wireless connectivity. In addition to academic expertise, major industry partners, e.g. EE, Huawei, O2, Vodafone, HEFCE, Enterprise M3, TEOCO Corporation, BBC, BT, Cobham, Anite, Ascom, Digital Catapult, Fujitsu, Rohde & Schwarz, Samsung, Roke, McLaren Applied Technologies, Ofcom, Imagination Technologies, ITRI, MYCOM OSI, Three and Ordnance Survey, are also partners in the centre. Technology that enables one terabit per second (Tbps) communications speeds has been developed (more than 1,000 times faster than the highest 4G speed) and 15 patents have been filed. The aim is to work in cooperation rather than competition to create a single standard for 5G. Commercialisation of technologies is expected from 2020. A 4G network testbed is also available to researchers which will be upgraded to 5G and Internet of Things. The testbed will provide 10Gbs/per cell by 2018, ten times faster than the highest speed available over 4G. Already Huawei, BBC R&D and the 5GIC have demonstrated ultra-high-definition (4k) video streaming to a mobile device. The centre has also demonstrated a 5G-Sparse Coding Multiple Access radio waveform that can support at least three times the number of IoT devices than would be possible with 4G [245][246].

6.3.1.2 Kings College London

Work at Kings College [247] being led by Mischa Dohler on 5G is concentrated on the Tactile Internet [248]. The term "Tactile Internet" was first coined by Gerhard Fettweis from the Technical University of Dresden to describe using touch for real-time remote control of robotic systems. Work at Kings College is trying to "close the data cycle" so that it is possible to touch and feel remote systems. What is needed for this is very fast data rates thousands of times faster than today's average 4G. Work is being supported by Ericsson and Samsung to investigate the potential of the technology [249].

6.3.2 Germany

6.3.2.1 fast zwanzig20

Zwanzig20 [250] is a partnership for innovation initiated by the Technische Universität Dresden and funded by the BMBF. The programme "Zwanzig20 - Partnership for Innovation" aims at a supra-regional, inter-, trans- and multi-disciplinary collaborations of partners and promotes openness and transparency. In particular, a project called "Fast Wireless" is looking at reliable and real-time-capable 5G-Mobile. Here the target is low latency of 1 ms in distributed control applications with a very high number of users, i.e. sensors and actuators, per radio cell.

6.3.2.2 5G Lab Germany

Within the 5G Lab Germany [251], 20 professors from TU Dresden are collaborating in an interdisciplinary team with more than 500 scientists to advance research on the key technologies for the 5th generation of mobile communications (5G) and its applications. Applications being considered include Tactile Internet applications, e.g. automated driving and robotic-aided tele-surgery. To achieve this goal, the whole value chain is being addressed: from the semiconductor chips across wireless data transmission, networking and mobile edge clouds to Tactile Internet applications.

6.3.2.3 Berlin 5G

5G Berlin [252] is a Fraunhofer led research initiative formed in 2014 by Fraunhofer HHI and FOKUS based on existing technology knowhow and testbed activities in wireless broadband. The aim is to integrate core technologies and expertise from academia and industry across disciplinary borders and test latest technologies, system components and applications in a real world setup. New partners can join this initiative to develop and test innovative 5G infrastructures, products, and applications.

6.3.2.4 5G:Haus

5G-Haus [253] is a virtual laboratory set up by Deutsche Telekom to develop the architecture and steer standardisation work in cooperation with Continental, Fraunhofer ESK and Nokia Networks. They have demonstrated communication between vehicles via the LTE cell network. The aim is for vehicles to share hazard information on the motorway which requires very short transmission times.

6.3.3 Finland

The 5G Test Network Finland (5GTNF) [254] coordinates and combines the research and technology development activities from the 5G infrastructures built under the Tekes 5thGear programme [255]. 5GTNF creates a single coherent entity from the numerous smaller test networks around Finland and represents them as an integrated innovation platform to the research community, industry and other interested parties on a national and international level.

6.3.4 Multinational

6.3.3.1 Ericsson 5G

Within Europe Ericsson is very active in 5G and is supporting a number of projects. The company is working with King's College London and the Technische Universität Dresden to collaborate on 5G research, addressing technical implications and the societal challenges of implementing 5G technology. A focus for the work is machine-type communication and Smart Sustainable Cities applications. Ericsson is also funding work at the Royal Institute of Technology, Chalmers University of Technology and Lund University in Sweden. At a world-wide level Ericsson has also teamed with LG Uplus and MTS for 5G development, as well as working to unify 5G infrastructure through the 5GEx project.

6.4 US 5G Drivers and Policy Initiatives

Internet access provides substantial economic benefits across the U.S. economy. It was noted that in 2000, about 200 million American homes had traditional phone lines, however, now there are only 73 million showing how voice over Internet and wireless connections have altered lives. Wireless networks carry more than 100,000 times the traffic that they were supporting in 2008. Under the Recovery Act $5 billion was made in broadband investments, including support for more than 114,000 miles of broadband infrastructure, especially in under-served areas. Tax incentives were also introduced in order to encourage wireless companies to invest tens of billions in infrastructure and services.

With the rapid growth in the use of wireless communications a key commitment was made to make 500 MHz of Federal and nonfederal spectrum available by 2020 for mobile and fixed wireless broadband use. Commercial companies and Federal agencies have collaborated to achieve this by finding new ways to free up spectrum, e.g. by sharing. Half of the 500MHz has been made available and was auctioned off in the FCC's Advanced Wireless Services 3 (AWS-3) auction raising $40 billion. There are also plans to auction off a further 126 MHz of spectrum. White spaces between television channels are being made available and expansion of access to the 5 GHz band (currently used for Wi-Fi services) is being considered.

6.4.1 4G

The United States leads the world in 4G deployment with 4G/LTE coverage for more than 98 percent of U.S. citizens and speeds 10 times faster than 8 years ago. There has been nearly $150 billion in 4G LTE investment by wireless operators since 2010. This includes an $85 million investment in advanced wireless testing platforms by a public-private effort between NSF and more than 20 technology companies and associations. These platforms, and the fundamental research supported on them, allows academics, entrepreneurs, and the wireless industry to test and develop advanced wireless technology ideas. America's success in 4G is seen as being down to a clear policy strategy that favoured making spectrum available early and establishing flexible-use rules to enable innovators and entrepreneurs to define the future of wireless technologies and applications. By avoiding a rigid, top-down, standards-setting process and technology roadmap, there has been an open competition and development of technologies that have been taken up by the marketplace. Another policy strategy is to provide sustained Federal investments in fundamental academic research that leads to technology breakthroughs. NSF, the Defense Advanced Research Projects Agency (DARPA), and other Federally-funded academic research into new channel access, antenna, modulation, and other technologies has made important contributions to the 3G and 4G revolutions, the broad deployment of Wi-Fi, millimeter wave (mmWave) technologies, and new dynamic spectrum-sharing arrangements.

6.4.2 MOBILE NOW Act

On the 3rd of March, 2016, the Senate Committee on Commerce, Science, and Transportation approved the MOBILE NOW Act [256] to boost the development of next-generation 5G wireless broadband by ensuring more spectrum is made available for commercial use and reducing the red tape for building wireless networks. This came in advance of a new major initiative the Advanced Wireless Research Initiative.

6.4.3 Advanced Wireless Research Initiative

President Obama announced a $400 million Advanced Wireless Research Initiative in July 2016 to boost research for next-generation mobile networks [257]. This will establish four city-sized testing grounds for 5G wireless services from October 2017. Funding for the initiative will come from the National Science Foundation with $35 million also being invested by American Telecoms companies including AT&T, Intel, Sprint, T-Mobile, Qualcomm and Verizon. The drivers for this investment are to support the growth in net connected smart phones. Here there is an expected growth from the current 370 million smart phones (a doubling in phones over the last 10 years) and other devices to 200 billion devices globally in the next 5 years (25 connected devices per person). The Internet of Things alone is expected to add 50 billion connected devices globally by 2020. Data traffic in North America is expected to grow at a 42 percent compounded annual growth rate between 2015 and 2020.

Supporting this the Federal Communications Commission voted to make a large block of spectrum available for permission-less 5G development and use. The vote supports the US strategy of spectrum first with flexible-use and makes the US the first country in the world to make large amounts of high-frequency millimeter wave spectrum available for both licensed and unlicensed use. The spectrum designated is in the higher bands: 28 Gigahertz, 37 GHz and 39 GHz which makes it possible to deliver more data than current networks giving high-capacity, high-speed, low-latency wireless networks. It is expected that 5G services will provide speeds of up to 100 times faster than existing 4G wireless networks. Regulation was also updated to make it easier for telecommunications providers to transition from older telephone networks to newer wireless and Internet-based voice networks. Although it was noted that some of the regulations remained onerous, e.g. to provide connections to existing fax machines until 2025 and prove cybersecurity standards (which are the remit of other government agencies).

A number of potential applications for the new technology are cited in the areas of smart cities and the Internet of Things. These include:

- Mobile phones and tablets that can download full length HD movies in less than 5 seconds, 100 times faster than 4G (6 minutes) and 25,000 times faster than 3G (26 hours).
- First responders and emergency room doctors who get live, real-time video and sensor data from police vehicles, ambulances, and drones, along with patient vitals and medical records—all before the patient arrives at the hospital door.
- Semi- or fully-autonomous vehicles that can communicate with the outside world and with each other to improve travel efficiency and safety.
- Factories equipped with always-connected smart manufacturing equipment that self-diagnose and repair themselves before they break.
- Gigabit-speed wireless broadband available in businesses, public transportation stations, stadiums, campuses, schools, malls, parks, and other public spaces.

- Virtual reality training environments and simulators that allow entry-level workers to develop and demonstrate skills in high-demand fields like solar energy installation—anytime, from anywhere.

6.4.4 NSF

In support of the Advanced Wireless Research Initiative NSF is committing $50 million over the next 5 years to design and build four city-scale advanced wireless testing platforms. $5 million has been reserved to create a project office to manage the design, development, deployment, and operations of the testing platforms. Each platform will deploy a network of software-defined radio antennas city-wide to mimic the existing cellular network. This will allow academic researchers, entrepreneurs, and wireless companies to test, prove, and refine their technologies and software algorithms in a real-world setting. The aim is to allow "at-scale experiments" and perform proof of concept work. The open competition will establish 4 US cities as centres for wireless research and development.

A further $350 million will be invested over the next 7 years in fundamental research on advanced wireless technology projects that will exploit these platforms. Additionally NSF will fund two prize challenges to enhance wireless broadband connectivity. The first challenge will focus on providing rapid, large-scale wireless connectivity to restore critical communication services in the aftermath of a disaster. The second will seek innovative solutions to provide low-cost, seamless connectivity in urban areas using fibre optics in overhead light poles.

Jointly with Intel Labs NSF will provide $6 million for research on information-centric wireless edge networks, with the goal of developing the ability to process very large quantities of information with response times of less than one millisecond. Jointly with the Academy of Finland $4.7 million will be made available to support joint US-Finland research projects on novel frameworks, architectures, protocols, methodologies, and tools for the design and analysis of robust and highly dependable wireless communication systems and networks, to support the Internet of Things. To bring communities together funding will be provided to support biannual meetings of international researchers in the Millimeter Wave Research Coordination Network and a large-scale networking platforms "Communities of Practice" workshop will be supported to gather international expertise on best practice. A workshop on ultra-low latency networks will also be supported.

6.4.5 DARPA

DARPA has announced the Spectrum Collaboration Challenge (SC2) which will use the advanced wireless test platforms to develop a new wireless paradigm for collaborative, local, real-time decision-making. Here the idea is that radio networks will autonomously collaborate and reason about how to share the RF spectrum.

6.4.6 National Institute of Standards and Technology (NIST)

NIST is creating a multi-disciplinary working group to establish a Future Generation Communications Roadmap. The aim will be to identify key gaps and also the R&D opportunities related to future-generation communications systems and standards. NIST via the 5G mmWave Channel Model Alliance is also performing a coordinated channel measurement, verification, and comparison campaign for indoor environments. The results will be presented at the First International Workshop on 5G Millimeter Wave Channel Models scheduled for December 4, 2016 at IEEE's GLOBECOM conference.

6.4.7 National Telecommunications and Information Administration (NTIA)

The Institute for Telecommunications Sciences (ITS) plans to sponsor undergraduate and graduate student wireless spectrum research on the test bed at the University of Colorado-Boulder. The aim will be to explore campus-scale wireless networking, spectrum sharing, and mobile applications in collaboration with ITS, CU-Boulder, and the City of Boulder. With the Center for Advanced Communications, ITS will demonstrate the Measured Spectrum Occupancy Database (MSOD) project at the International Symposium on Advanced Radio Technologies in Colorado. MSOD has been recording spectrum utilization 24 hours a day, seven days a week, over several years, using data from multiple sensor installations. Additional networked sensor sites are also being installed. ITS will also expand its Urban and Indoor Radio Frequency (RF) Propagation Measurement campaign by using the four city test sites to provide additional data to improve the accuracy of RF propagation models in urban terrain. Electro Magnetic Compatibility (EMC) analyses and tests will also be performed at the test beds.

6.4.8 Industrial Support for the Advanced Wireless Research Initiative

More than twenty private-sector companies and associations in the U.S. wireless industry have pledged more than $35 million in cash and in-kind support to the design, development, deployment, and ongoing operations of the testing platforms. This includes design support, technical networking expertise, networking hardware, including next-generation radio antennas, software-defined networking switches and routers, cloud computing, servers, and experimental handsets and devices, software, and wireless network testing and measurement equipment. Specifically:

- **AT&T** will provide on-site mobile connectivity in the cities selected.
- **Carlson Wireless Technologies** will contribute equipment, technology, and expertise in TV white spaces and dynamic spectrum sharing
- **CommScope**, will contribute connectivity solutions such as antennas, RF cabling, cabinets, small cells, and fiber optics.
- **HTC** will provide technical expertise, mobile devices, IoT sensors and virtual reality systems.

- **Intel** will contribute its portable 5G mobile trial platform and server equipment to assist in research on mmWave, multi-antenna array, steerable beamforming, novel radio interface techniques, and anchor-booster architecture.
- **InterDigital** will contribute financial support and provide tools for spectrum and bandwidth management, heterogeneous networks and backhaul.
- **Juniper Networks** will contribute software, systems, and expertise to advance orchestration and authentication of massively-scalable, massively-distributed IoT networks, as well as new approaches to secure these networks.
- **Keysight Technologies** will provide a range of current and next-generation cellular and WLAN hardware and software products and also consulting and testing assistance.
- **National Instruments** will provide equipment from its software defined radio platform to support next-generation wireless communications research in areas like mmWave and Massive MIMO.
- **Nokia** (with **Nokia Bell Labs)** will provide financial contributions, research collaborations, governance, and product platform support, and will focus on software-defined radios, the Internet of Things, remote sensing, mmWave, security, new use cases and applications, and dynamic spectrum sharing.
- **Oracle** will provide core network controls, analytics, and network orchestration to researchers and help them understand the impact of subscriber behaviors, enhance orchestration, and bolster security.
- **Qualcomm** will contribute financial support as well as engineering equipment and guidance to explore new and innovative communication systems.
- **Samsung** will contribute research design and engineering expertise, with a particular emphasis on technologies for future wireless networks in the 28GHz and other millimeter wave bands, as well as continued enablement for the Internet of Things.
- **Shared Spectrum** will provide technical expertise in dynamic spectrum sharing to support the design and architecture of research platforms.
- **Sprint** will support research and development and technical expertise on network design, use cases, and architecture requirements for core and radio access networks and the devices that will access them.
- **T-Mobile USA, Inc.** will provide technical expertise, including staff engineering assistance or advice in the design and deployment of the testing platforms.
- **Verizon** will contribute technical expertise, such as staff engineering assistance in the design and deployment of the testing platforms, and in fixed and mobile systems, indoor and outdoor environments, and residential and commercial buildings.
- **Viavi Solutions** will provide test, measurement, assurance, and optimization solutions for lab and field trials for network and services to enable next-generation technologies for the always-connected society and Internet of Things.
- The **Alliance for Telecommunications Industry Solutions (ATIS)** will provide technical assistance and staff time on the design and deployment of the testing platforms. ATIS will also identify potential opportunities for research to be conducted on the platforms.

- **CTIA** will contribute engineering and technical assistance to help align industry R&D and university research with next-generation wireless networks, devices, and applications.
- The **Telecommunications Industry Association (TIA)**, will provide technical and engineering expertise in wireless network deployment, Internet of Things, interoperability, and software-defined networking. TIA will also assist with expanding industry awareness of the testing platforms.

6.4.9 5G Americas and 5G Forum USA

5G Americas (which changed its name from 4G Americas on February 12, 2016) is an industry trade organisation headquartered in Bellevue, Washington, composed of leading telecommunications service providers and manufacturers [258]. The organization's mission is to advocate for and foster the advancement and full capabilities of LTE wireless technology and its evolution to 5G. 5G Americas' key aim is to develop a connected wireless community while leading 5G development for all of the Americas. The organisation is addressing the following areas:

- Standards recommendations, technical requirements and advocacy for LTE, LTE-Advanced and LTE-Advanced Pro technologies and beyond
- Supporting the 3GPP technology path as it evolves to 5G technology
- Serving as a resource for information on LTE wireless technology and 5G throughout the Americas

5G Americas organises a 5G Forum in the US every year which brings the community together. 5G Americas is also supporting the Global 5G Event in Beijing, China on May 31-June 1, 2016 collaborating with The Fifth Generation Mobile Communications Promotion Forum (5GMF) (Japan), 5G Forum (Republic of Korea), IMT-2020 (5G) Promotion Group (China) and The 5G Infrastructure Association - Public Private Partnership (5G PPP) (Europe).

In October 2015 5G Americas published a white paper on 5G Technology Evolution Recommendations, expanding upon 4G Americas' view of 5G recommendations in 2014.

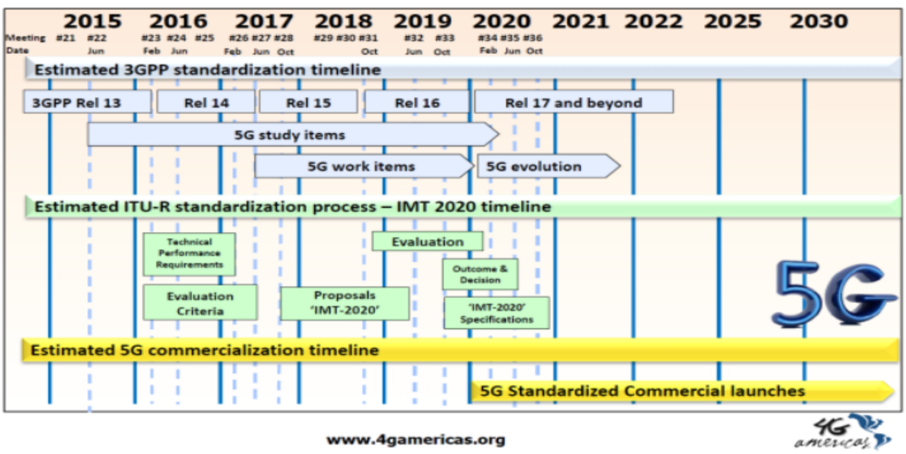


Figure 58. 5G Americas Roadmap

A key recommendation from this was that the US should invest in a national programme in 5G in order to compete with the rest of the world which led to the Advanced Wireless Research Initiative. Figure 58 shows the proposed technology roadmap to 2030.

6.4.10 Other US Research Initiatives in 5G

6.4.3.1 University of Texas at Austin

The Wireless Networking and Communications Group (WNCG) is an interdisciplinary center for research and education at the University of Texas at Austin with an emphasis on industrial relevance. Founded in 2002, the group includes 20 faculty from the departments of Electrical and Computer Engineering, Aerospace Engineering, Mathematics, and Computer Sciences. The group is addressing mmWave research to demonstrate the viability of the technology for 5G systems [259]. The research is investigating data rates, blockages, system coverage, sensitivity to interference and antenna arrays. Approaches to hybrid beamforming transmission strategies are being developed to adaptively configure arrays at the transmitter and receiver. Ways for propagation channel estimation are also being developed, which can aid the beamforming process [260].

Five technologies that could lead to both architectural and component disruptive design changes have been identified: device-centric architectures, millimeter wave, massive MIMO, smarter devices, and native support for machine-to-machine communications. UT Austin and Stanford work together in this area and they have recently been awarded a NSF grant of $978,000.

6.4.3.2 Stanford University & Berkeley

Stanford University in collaboration with Berkeley are working on Software-Defined Networking (SDN) and Network Functions Virtualization (NFV) which are two key pillars for 5G [259]. The concept for SDN arose out of the graduate work of Martin Casado at Stanford in collaboration with Nick McKeown, also at Stanford, and Scott Shenker at Berkeley. The Open Networking Research Center (ONRC), was set up to develop and support open source SDN tools and platforms for OpenFlow and SDN. This is sponsored by CableLabs, Cisco, Ericsson, Google, Hewlett-Packard, Huawei, Intel, Juniper, NEC, NTT Docomo, Texas Instruments and VMware.

The ONRC is made up of the networking research groups at Stanford University and UC Berkeley and an independent, non-profit Open Networking Laboratory (ON.Lab). The ONRC collaborates with the Open Networking Foundation (ONF) on the Open Networking Summit, bringing a cross-section of industry engineers, business leaders and researchers together for tutorials and keynotes.

6.4.3.3 NYU Wireless

New York University Wireless [261] was established in 2012 as a multi-disciplinary research centre, focusing on 5G wireless research in the medical and computer science fields. The centre is funded by the National Science Foundation and has an Industrial Affiliates board of 10 major wireless companies. NYU WIRELESS has conducted and published channel measurements that show that millimeter wave frequencies are viable for multi-Gigabit per second data rates for future 5G networks.

6.4.3.4 Rutgers University

The Cyber-Physical System (CPS) Lab at Rutgers University, funded by NSF, are working on dynamic provisioning and allocation under the Cloud Radio Access Network (C-RAN). They have shown that the dynamic demand-aware provisioning in the cloud will decrease the energy consumption while increasing the resource utilization. They have also implemented a testbed to demonstrate the feasibility of C-RAN and developed new cloud-based interference cancellation techniques.

6.4.3.5 T-Mobile and Ericsson US

Figure 59. Ericsson Prototype 5G Networking Technology at Mobile World Congress. (Source Stephen Shankland/CNET)

T-Mobile US, Inc. and Ericsson are working together and will jointly evaluate the performance and applicability of potential 5G key technologies. An aim is to develop a pre-standards 28GHz 5G test system for lab and field trials in the United States [262] with trials beginning in the second half of 2016. The work builds upon Ericsson's long-standing partnership with T-Mobile for LTE-Advanced and network transformation. The two companies will develop and test selected 5G use-cases and services to enable T-Mobile to evaluate emerging 5G technologies, drive 5G technology development and standardization, and to explore new business opportunities. Notably Ericsson are predicting 150 million 5G subscriptions by the end of 2021. T-Mobile currently delivers Americas fastest 4G LTE network and the market is waiting for 5G consumer smartphones and standards. Ericsson has announced more than 20 agreements to test 5G with operators across the globe and Ericsson 5G Radio prototypes, designed for operator field trials, are already achieving more than 25 Gbps mobile throughput.

6.4.3.6 Verizon

Verizon is planning field trials of 5G networks in 2017 at their Basking Ridge headquarters with a planned entry into service date of 2020 with 1GBps. Already, U.S. providers are promising to deploy 5G networks as early as

2020 [263]. The 5G network will provide speeds 200 times faster than the 5 Mbps generally available today on Verizon's LTE network. Verizon is also working with Alcatel-Lucent, Ericsson, Cisco, Nokia, Qualcomm and Samsung to test 5G in their innovation centers. The company is also talking to NTT Docomo. Verizon has set up a 5G Technology Forum that includes venture capitalists and the company has created "sandbox" testing areas for 5G technology, in its innovation centers in Waltham, Massachusetts, and San Francisco. Verizon considers other countries to be further ahead when it comes to policies on spectrum allocation for 5G [264].

6.5 Rest of the World

6.5.1 Next Generation Mobile Networks

The NGMN Alliance [265] has defined a 5G Roadmap that shows an ambitious time-line with a launch of first commercial 5G systems in 2020. The roadmap also defines how the industry players can achieve the required standardisation, testing and trials to make mature technology solutions available. The white paper highlights that standardization of technology is essential for the global success of future 5G solutions and the related ecosystem. This is to allow multi-vendor interoperability and economies of scale whilst reducing complexity and cost of interfaces. A challenge is that there are a range of interfaces, network elements and legacy systems, so it is necessary for numerous standardization bodies to come together for 5G standardization. Here the main risks are diversity of interests, conflicting standards and redundant options. This could result in development delays. Some coordination is required to avoid parallel work on similar areas (with potentially conflicting standards) so that solutions from different organisations are harmonised into an absolute minimum set.

6.5.2 South Korea

The South Korean mobile operator KT has said that it will launch a live 5G service for the 2018 Winter Olympic Games in the city of Pyeongchang. There have been a number of initiatives in South Korea. In 2008, the South Korean IT R&D programme of "5G mobile communication systems based on beam-division multiple access and relays with group cooperation" was formed. South Korea's Ministry of Science, ICT and Future Planning have also announced that it is committing $1.5 billion to its "5G Creative Mobile Strategy" [266]. The Korea 5G Forum was set up a collaborative development environment for 5G wireless communication and was founded by the Ministry of Science, ICT (information and communications technology) and Future Planning (MSIP). The forum brings together telecommunications companies, semiconductor companies, academia, and research institutes [267] and cooperates internationally on standardization, research and development. The 5G Forum collaborates with the EU (5GPPP and NGMN Alliance), Global Forum Wireless World Research Forum (WWRF), China (IMT-2020 PG, FuTURE Forum), Japan (5GMF) and India (GISFI) [268].

In 2014 an agreement was signed between the European Commission and South Korea to work towards a global definition of 5G, to cooperate in 5G research and also on harmonization of radio spectrum to ensure global interoperability. A coordinated call for research project proposals was held in 2016. Additionally a memorandum of understanding has been signed between the EU's 5G Infrastructure Association 5G-PPP (whose members include Alcatel-Lucent, Atos, Deutsche Telekom, Ericsson, Nokia, Orange, Telecom Italia, Telenor and Telefonica) and South Korea's 5G Forum.

The ETRI (KR) 5G Giga Communication Research Laboratory [269] is researching and developing 5G mobile communications technologies to lead the Korean governments "creative economy". The research areas being addressed include core technologies of mmWave wideband mobile communication and C-P-D-N (Contents-Platform-Device-Network)-interconnected technologies. Specifically, 5G high-capacity/low-latency wireless communications, 5G massive/low-power wireless communications, L1/L2/L3 core technologies for mmWave devices and base stations, and antenna/RF/channel research for mobile communications are being considered. Research is also being performed in digital holography broadcast-communication.

6.5.3 China

In China, three ministries: the Ministry of Industry and Information Technology (MIIT), the National Development and Reform Commission (NDRC) and the Ministry of Science and Technology (MOST) have set up an IMT-2020 (5G) Promotion Group to coordinate all 5G activities in Chinese industry and academia [266]. In November 2013, Chinese telecom equipment vendor Huawei stated that it will invest $600 million in research for 5G technologies in the next five years. This does not include additional investment to productise 5G technologies for global telecom operators. Huawei plans to test 5G technology in Malta and is looking to set up a research and development facility there depending on the 5G test results [270]. Huawei sees Europe as a big potential market and also sees the need for a global 5G standard. Huawei is actively engaged in many European projects. Huawei is a key contributor to the 5G PPP and is working on five of the 19 co-funded projects. Huawei is a Board Member of the 5G Infrastructure Association, and is contributing to several projects in the H2020 Framework Programme, e.g. METIS-II, FANTASTIC-5G, mmMAGIC, 5G-Xhaul and 5Gex projects. It is also providing £5 million of funding for 5G research and a testbed at the 5G Innovation Centre at the University of Surrey, UK. Additionally, it has launched the 5G Vertical Industry Accelerator (5G VIA) in Munich targeting a large-scale 5G testbed to simulate real-world scenarios. In 2015 Huawei won an award for the "Biggest Contribution to 5G Development" [271].

In 2015 Huawei and Ericsson began testing 5G-related technologies in rural areas in the Northern Netherlands at Loppersum in North Groningen. As most tests are currently being performed in urban areas testing in the countryside allows new applications to be considered, e.g. autonomous cars, tractors, home appliances, monitoring elderly people, delivering music, etc. [272].

6.5.4 GSMA

The GSMA represents the interests of mobile operators worldwide, uniting nearly 800 operators with more than 250 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and Internet companies, as well as organisations in adjacent industry sectors. The GSMA will play a significant role in shaping the strategic, commercial and regulatory development of the 5G ecosystem [273]. This will include areas such as the definition of roaming and interconnect in 5G, and the identification and alignment of suitable spectrum bands. The GSMA released a major new report at the GSMA Mobile 360-Europe event held in Brussels in December 2014, outlining its perspectives on the development of 5G. The report, 'Understanding 5G: Perspectives on Future Technological Advancements in Mobile' [237], provides an overview of network technology innovation today and how this is setting the agenda for the 5G future. It outlines the technical requirements of future 5G networks and explores potential use cases as well as the implications for operators and other mobile ecosystem players.

Many of the 5G technical requirements, e.g. network functions virtualisation (NFV), software-defined networks (SDN), heterogeneous networks (HetNets) and Low Power, Low Throughput networks are already being brought to market by vendors and deployed by operators as part of 4G. There is still an opportunity for growth in 4G, which still only accounts for 5 per cent of the world's mobile connections. At a country level, however, 4G has taken off far more in South Korea (69% of connections), Japan (46% of connections) and US (40% of connections) [237]. In the developing world 4G penetration is 2%. The expectation is that mobile operators will invest US$1.7 trillion globally in network infrastructure over the period 2014-2020, the majority of which will be spent on 4G networks.

An application that requires at least one of the two key 5G technical requirements (greater than 1 Gbps downlink and sub-1ms latency) can be considered a "true" 5G use case. Here there are opportunities for virtual reality/augmented reality/immersive or tactile internet, such as gaming, wearable tech. or health services, autonomous driving/connected cars and wireless cloud-based office/multi-person videoconferencing.

6.5.5 Alvarion Israel

In April 2012, Alvarion set a world record, reaching a tenfold increase in mobile broadband infrastructure capacity. Their $6 million Beyond Next Generation (BuNGee) project achieved a 'pre-5G' throughput of 1Gbit/s per square kilometre at Alvarion's premises in Tel Aviv. (A speed of 1Gbit/s allows you to download a standard high-definition movie in about a minute.) The demonstration confirmed that high-speed mobile Internet solutions were commercially viable.

6.5.6 Japan

Japan's government has been less active than the EU, China and South Korea in setting up national 5G R&D initiatives, however, Japanese companies are active and NTT Docomo plans to showcase the technology at the 2020 Tokyo Summer Olympics [274]. The company is performing "experimental trials" of 5G technologies with six vendors: Alcatel-Lucent, Ericsson, Fujitsu, NEC, Nokia and Samsung. Companies in Japan are also leading the Japan's Association of Radio Industries and Businesses (ARIB) "2020 and Beyond AdHoc" group.

Docomo produced a White Paper [275] in a July 2014 outlining the need for data rates 100 times higher, latency of 1 millisecond and a 1,000-times increase in systems capacity while at the same time reducing energy consumption. Docomo expect a 100-fold increase in the number of simultaneously connected users compared to 4G LTE. To achieve this they are working on millimeter-wave band technology, which allows multi-beam multiplexing and massive multi-input-output (MIMO) technologies [276]. These operate in the 6 to 66 gigahertz frequencies. Rather than broadcasting signals from a base station in all directions, individual signals can be transferred between individual terminals and a base station as required. In a crowded hot spot a cluster of smaller antennas can be used to eliminate interference from nearby terminals which reduce data rates and make better use of signal power. Already Docomo and Ericsson have shown that multi-beam MIMO technology can transmit data with a cumulative 20-Gbps throughput with four mini base-stations, each equipped with 64 antenna elements. Two Ericsson 5G prototype terminals, located 9 meters and 3 meters respectively from the base station were each able to simultaneously download over 10 Gbps over a 15-GHz wireless band. The companies also showed that it was possible to transmit data at 10 Gbps over a distance of 70 meters from the base station and then at 9 Gbps over a distance of 120 meters.

6.6 Standards

The uptake of 5G depends upon standardisation to be in place, however, the roll out of 5G is expected to be gradual allowing equipment upgrades to occur before some of the key 5G standards are formalised in 2018 and 2019 [277]. Pre-standard "5G-ready" equipment using software defined network (SDN) technology will allow network operators and enterprise customers to move to upgrade to full 5G once standardisation is in place for spectrum allocation and licences are issued.

6.6.1 ITU

The International Telecommunications Union wants to bring together people along with things, data, applications, transport systems, and cities in a smart networked communications environment. Mobile data traffic across the globe grew 69 percent between 2013 and 2014, reaching 2.5 exabytes (over a billion billion bytes) per month, according to Cisco. Analysts expect data consumption to climb to 24.3 exabytes per month by 2019 which cannot be met by existing 4G LTE. The International Telecommunication Union (ITU)'s IMT-2020 Focus Group [278] has defined a 5G network blueprint for technology improvements taking into account 60 research proposals to address gaps in the 5G wire-line network infrastructure, such as software and high-level network architecture [279].

6.6.2 Joint Standardisation

In order to get joint agreement on technical fundamentals and 5G spectrum bands globally by 2018 NTT Docomo (Japan), KT and SK Telecom (South Korea), and Verizon(US) are forming the 5G Open Trial Specification Alliance with the aim of driving the technology forward and creating standards for network equipment makers to follow.

6.6.3 3GPP

3GPP standardization is progressing and speeding up towards 5G. At a RAN 5G Workshop in September 2015, 5G standardisation activities were initiated with first study items on new waveforms, channel modelling >6GHz, and Radio Access Network architecture.

6.6.4 IEEE 5G initiative

The IEEE 5G initiative, which targets the creation of a global meta-standard for 5G, is considering the requirements of the BRIC and low-income countries in Africa, Asia and South America to enable 5G applications in large rural, low populated areas.

6.7 Need for Regulation

The U.N.'s International Telecommunications Union is working on spectrum harmonization for 5G so that the same frequencies are used worldwide [280]. Global harmonization is important for networking equipment and device manufacturers who operate in an international market to keep prices low for potentially billions of mobile devices. Harmonization was not possible for 4G LTE and this has resulted in slower deployment in some parts of the world, particularly Europe. The US has a dominant lead in 4G LTE but the rest of the world is now concentrated on 5G. A problem with 5G is that multiple international bodies are involved in different aspects of 5G standardisation. Long-term spectrum planning is essential and an issue in the US is that valuable 5G spectrum is already allocated to departments and agencies of the federal government, however, much of it is unused or underutilised.

In many countries development depends on government funding but in the US innovation is driven by entrepreneurs. Putting in place the right policies is crucial to unlocking private investment for 5G technologies. It is estimated that network upgrades will cost $2 trillion [281]. There is a push for "permissionless innovation in 5G design" in the US but this has been challenged by "net neutrality" regulations adopted by the FCC which are the subject of a legal challenge [282] [283]. This may inhibit 5G experimentation.

7 Big Data

7.1 European Big Data Drivers and Policy Initiatives

The area of data is seen as vital to support one of Europe's 10 political priorities: the vision of a Digital Single Market (DSM)[284]. This strategy aims to open up digital opportunities for people and business and enhance Europe's position as a world leader in the digital economy. The DSM covers the free movement of persons, services and capital to allow individuals and businesses to seamlessly access and exercise online activities under conditions of fair competition. Key to this is to provide a high level of consumer and personal data protection, irrespective of nationality or place of residence.

Big Data has become increasingly important. It is used to describe large amounts of data being produced very quickly by a high number of diverse sources, either by people or sensors (e.g. gathering climate information, GPS signals, etc.) that needs to be processed at the same speed. The applications of Big Data cover many sectors, from healthcare, manufacturing, transport and energy. Key to the future knowledge economy is the ability to extract value at the different stages of the data value chain. The US currently leads in this area with big companies being at the centre of the data revolution. Within Europe the research and innovation funding for Big Data is currently considered insufficient with many smaller uncoordinated activities being performed. It is also more difficult to translate technology advances within Europe into business opportunities due to a complicated European legal environment and a lack of access to large datasets. This is particularly a problem for SMEs. The EC has thus identified a number of issues that need addressing:

- Lack of a "scale" in the EU research/ innovation
- Fragmented legal landscape
- Skills shortage
- Delays in the take-up of data technologies
- Lack of a supporting data "ecosystem"

In July 2014 the European Commission defined a new strategy on Big Data with the aim of supporting and accelerating the transition towards a data-driven economy in Europe [285]. The strategy has a number of goals to accelerate innovation create productivity growth, and increase European competitiveness in data within the global market to make Europe a key player. Over 150 research and innovation projects have been funded with over €50m EU funding in 2014 targeted at Big Data and over €89m in 2014-15 targeted at Big & open data. A key initiative is the setting up of the Big Data Public Private Partnership.

7.1.1 Open Data

Critical to maximising the benefits from Big Data is the ability to access and share data. Here there is a need for Open Data to allow authorised people from the same, or different, organisations to share information (EU Open Data policy framework) [286]. Through the use of shared data it is possible to develop a holistic view of a theme that is being addressed or gain new insights about areas. However, the sharing of data between agencies is difficult unless trust exist. Typically each department sharing data will have a separate set of Information Governance requirements which makes it difficult to combine data. Here there is a need for a common information governance regime with agreed common measures in place for data exchange. This requires three key elements:

- Semantics – the meaning of information
- Syntax – the format of information
- Data Quality – the confidence to re-use information

7.2 European Initiatives

7.2.1 Big Data Value Public-Private Partnership

Figure 60. Big Data Value Public-Private Partnership

The Big Data Value Public-Private Partnership [287] (See Figure 60) was set up in January 2015. This aims to strengthen the data value chain within Europe and allow it to play a role in the global market. In order to achieve this the European Commission has teamed up in a Public-Private Partnership with both large and SME industry, researchers and academia. The PPP cooperates in data-related research and innovation to support a data driven economy in Europe. A key aim is to build a community working on data in Europe. This is supported by the Big Data Value Association which includes data providers, data users, data analysts and research organisations. The association is a non-profit, industry-led organisation whose founding members include ATC, IT Innovation, IBM, SINTEF, University of Bologna (CINI), Polytechnical University of Madrid, NOKIA Solutions and Networks, THALES, University of Duisburg Essen, Siemens, SAP, Engineering, TIE Kinetix, ANSWARE, Software AG, Orange, Atos, INDRA, ITI, VTT, Fraunhofer, DERI, and the Technical University of Berlin.

7.2.2 Open Data for Smart Cities

There are two key projects in Europe considering the exploitation of open data for smart cities [288]. Coordinated by ESADE Business School, the Open Cities and Commons for Europe projects aim to enable open innovation in the public sector in seven European cities: Amsterdam, Barcelona, Berlin, Bologna, Helsinki, Paris and Rome.

7.2.2.1 Open Cities Project

The Open Cities project [289] enables participating cities throughout Europe to publish their data as Open Data. By doing so the project is encouraging the creation of web and mobile civic applications to enable better services, lower costs and improve transparency. Two Open Data Platforms have been created (one for static data and one for dynamic data). Around this data two Worldwide App Challenges have been organised. The

Open Cities Open Data Platform [290] developed by Fraunhofer FOKUS can be easily customised to match an organization's specific requirements. The platform supports the entire Open Data lifecycle process, which includes identifying, publishing, discovering, enriching, and consuming data. Additionally, the Pan European Federation of Open Data Platforms has created the Open Cities data catalogue which is a federated repository of existing city open data catalogues. This has been used for the pan-European Open Data App Challenge. The platform provides a consolidated view on all of the open data available from the city portals of Amsterdam, Berlin, Barcelona, Helsinki and Paris.

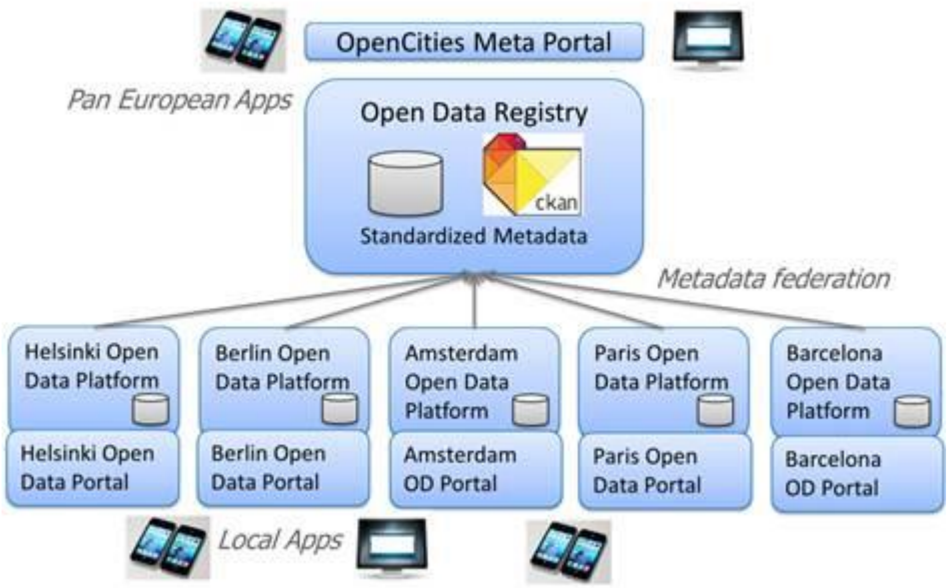


Figure 61. Open Cities Open Data Platform

The key benefit of the Open Cities data catalogue is that it provides easy access to searching, finding and browsing of data. This is useful for both apps developers and users. The Open Cities Open Data Platform, as shown in Figure 61, has been used as the basis for the Open Data Portal in Germany [291].

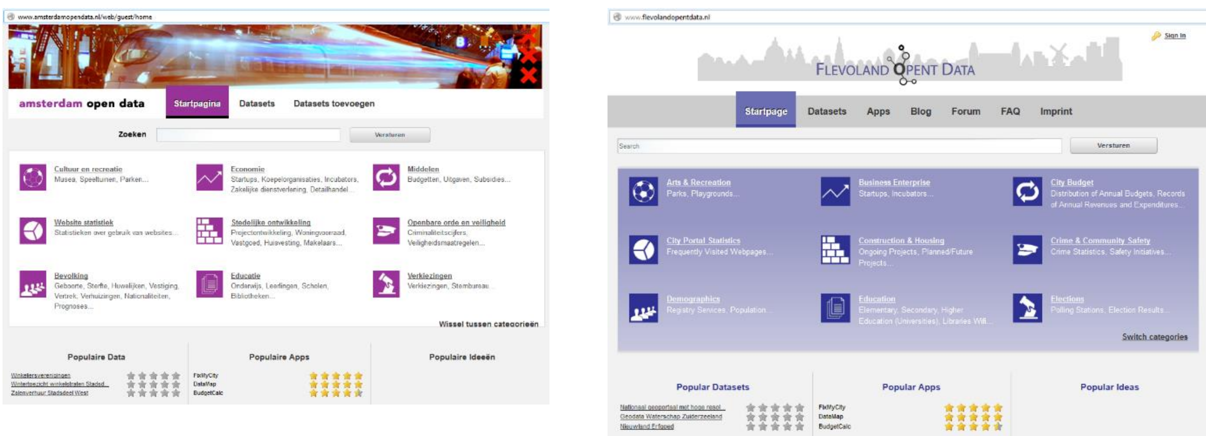


Figure 62. Open Data Amsterdam and Open Data Flevoland

The model has been already successfully "replicated" and implemented on-site in the city of Amsterdam [292] and in the province of Flevoland (See Figure 62).

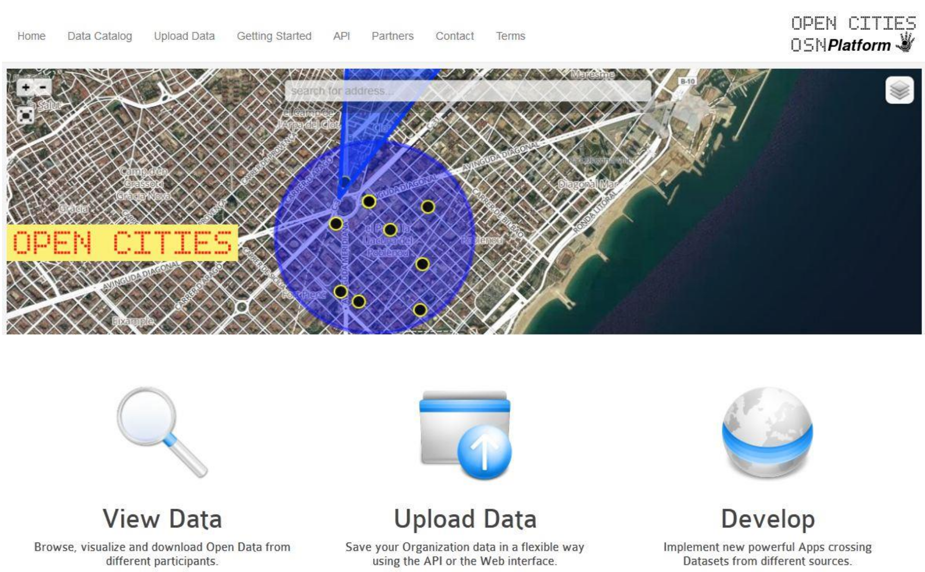


Figure 63. Open Sensor Network Platform (OSNP)

The Open Sensor Network Platform (OSNP) developed by Pompeu Fabra University (see Figure 63) offers a central point to publish and store data generated by sensors and dynamic services. This can be used for services related, for instance, to urban transport, tourism and demographics. The Consortium also runs an Open Cities App Challenge, inviting developers from all over the world to submit innovative ideas. Developers can turn their ideas into working apps through the "Hack at Home platform", which allows participants to present their ideas, get advice from mentors and form teams with designers, developers and coders to build apps for the participating cities. The Open Cities App Challenge has attracted more than 100 applications submitted from all over the world including apps for audio-based augmented-reality application for blind people, for providing real-time information on transportation and parking systems, for providing information on public transportation, bike rentals and traffic status. The aim of the challenge is to help Paris, Amsterdam, Barcelona, Berlin, Helsinki, Rome and Bologna to benefit from the talents of app creators and find solutions for managing tourism in cities.

7.2.2.2 Commons for Europe

The Commons for Europe Project [293] has three key aims to:

- Identify the needs of public administrations and citizens,
- Promote the creation of new applications that can provide innovative services, and
- Share the results openly among multiple European cities.

Solutions based on mobile and web applications that improve services, reduce costs and provide greater transparency are being developed. The team led by ESADE Business School [294] works with seven European cities to identify the types of new services that could benefit citizens and meet their needs through innovation. The Consortium has a team of technical experts in web and mobile applications that can offer services. The project intends to create an organisation called Code for Europe [295] similar to Code for America [296] to engage with promising young developers to provide mentoring and training. The aim is that any applications developed should be designed to work in any of the seven cities involved in the programme but, in addition, they will also be made available to any other European city.

The second key aim of the project is to promote knowledge and deployment of user driven Bottom-UP Broadband (BUB) networks in metropolitan and rural environments. This is being led by Guif.net [297] and the Universitat Pompeu Fabra who are driving creation of a pan-European network. The work is exploring new technologies such as Super-WiFi, bandwidth-sharing in fibre networks and sensor integration.

To avoid duplication of effort the Commons for Europe is developing a Code for Europe [295], a marketplace for civic innovation that features apps and digital services aimed at improving the lives of communities. Developers can add their apps to this resource with information on how they have been deployed and where to measure impact. World Bank sponsorship has been obtained to promote the Europe Commons repository in Latin and Central America.

7.2.2.3 Citadel on the Move

Citadel on the Move is a European Commission funded (CIP PSP) project [298] which aims to make it easier for citizens and application developers from across Europe to use Open Data to create innovative mobile applications. Citadel on the Move is addressing this by:

1. Defining strategies that make it easier for local government to release data in useable, interoperable formats;
2. Creating and providing templates that make it easier for developers and citizens to create mobile applications that can be potentially used and shared across Europe;
3. Pooling tools and resources into an Open Data Commons that facilitates access to data in different formats by shared templates and applications.

7.3 National Initiatives

7.3.1 Open Helsinki – Hack at Home programme

The City of Helsinki is looking for new ways to support developers who want to use open data in order to create digital services for their citizens. The Open Helsinki – Hack at Home programme [299] was launched on the 14 of June 2013 and encourages developers to create useful applications. The underlying themes of the programme are transparency of city decision-making and enabling better feedback to civil servants from citizens. Notably the Helsinki Region Infoshare service received the European Prize for Innovation in Public Administration. Hack at Home is an international concept, linking developers and organisations around the world. The collaboration of mentors, who give feedback and support in the different phases of the development process, and developers is at the very heart of the programme. Hack at Home also offers a forum to put together teams around ideas

7.3.2 UK

Data and digital are considered key to growth in the UK economy. In the UK a Public Data Group (PDG) was formed in 2011 bringing together 4 organisations (Companies House, Land Registry, Met Office and Ordnance Survey) that all concentrated on collection, management and distribution of vital data sets [300][301]. The aim of the group was to identify shared improvements, best practice and efficiencies across areas including surveying and IT. Work focused on data policy and a number of projects were run, e.g. the GeoVation challenge on Housing to use data to solve housing issues. The PDG Board met for the final time in April 2015 and the PDG was replaced by a Business Innovation and Skills Digital Culture, Services Platforms and Data Board bringing together senior officials and a wider family of organisations.

7.3.2.1 Open Data Institute

The Open Data Institute (ODI) [302] was officially launched at Shoreditch in December 2012 with the support of £2 million per year for five years from the Technology Strategy Board and $750,000 from the Omidyar Network. The Institute, founded by Sir Tim Berners-Lee and Prof Nigel Shadbolt, is an independent, non-profit, limited company. Its remit is to catalyse an open data culture that has economic, environmental and social benefits.

7.3.2.2 Data Governance

One of the main barriers to effective information sharing is the fact that different organisations collect information for different purposes and attach a different meaning or interpretation to it. The key to success is to rationalise Information Governance regimes across public services and address the semantics, or context, of information. To support this in the UK Sheffield City Council is developing an 'Information Governance Toolkit' for Local Government. The Local eGovernment Standards Body (LeGSB) is also creating a Public Service Concept Model (PSCM). This recognises that departments each have their own concept model that describes their own activities and outcomes but there are common concepts that are typically linked. A Smart Cities Concept Model, derived from the PSCM, would allow city organisations to share data by providing a simple neutral language to describe each data source, and the links between them. This approach could lead to a common agreed expression of the issues facing a city, such as: Congestion, Families, Skills, Ageing Population, Dementia, etc., and identify the different organisations and interventions necessary. Importantly, the information should not be stored in one place. Each organisation publishes their shared data over the web, and this is brought together (as linked data) to meet a particular need.

7.4 US Big Data Drivers and Policy Initiatives

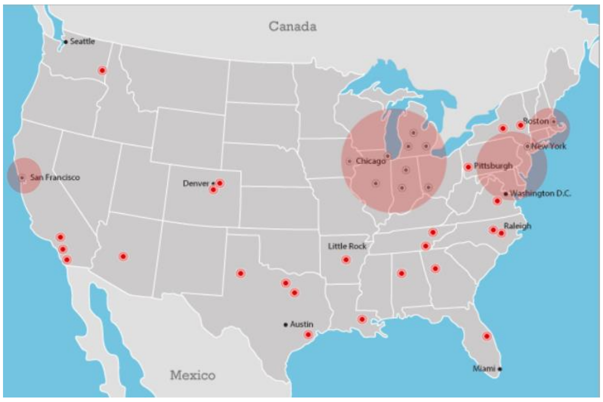


Figure 64. Map of US Big Data Research Activities

In the US Big Data has been a key theme since 2012 through the Big Data Research and Development Initiative. This was set up to support the ability to extract knowledge and insights from large and complex collections of digital data. Following commitment from central government, a number of agencies such as NSF, NIH, DARPA and the Department of Defence (DoD) launched their own Big Data Initiatives. Figure 64 shows a map of Big Data research across US universities.

7.4.1 Policy Issues in the US

As more initiatives were established it was clear that there were a number of policy issues that need to be tackled by coordinated policy measures. More specifically, the report “Report To The President Big Data And Privacy: A Technological Perspective”, Executive Office of the President President’s Council of Advisors on Science and Technology May 2014 [303], presents a number of policy issues on regulation and promotion of research capacity of RTOs, while the National Science Foundations (NSF), through its NITRD Big Data Senior Steering Group, has highlighted the need for formulating policies on enhancing Education and Workforce Development and establishing new ways of partnerships.

The Policy in the US focuses more on the actual uses of Big Data and less on its collection and analysis. The US wants to lead both in the international arena and at home. The policies and regulation do not promote technological solutions, rather they state intended outcomes. There is also an accent on policies to strengthen and stimulate US research in practical privacy-related technologies. Additionally, there is focus on aspects of social science that promote the successful application of technologies.

7.4.2 NSF Big Data Research Initiative

NSF through its Big Data Research Initiative [304] is developing new methods to derive knowledge from data; constructing new infrastructure to manage, curate and serve data to communities; and supporting new approaches for associated education and training. This includes actions on core techniques and technologies for advancing Big Data in science and engineering and Big Data solicitation, graduate education and training.

Core to NSF’s strategy is the Cyberinfrastructure Framework for 21st Century Science and Engineering, or "CIF21” [305]. The aim of CIF21 is to foster the development and implementation of a national cyberinfrastructure for researchers in science and engineering to achieve a democratization of data. NSF announced new awards ($10 million each) under its CIF21 and Expeditions in Computing programmes, as well as awards to expand statistical approaches to address Big Data, support cyberinfrastructure, the geosciences and training. The aim is to support fundamental science and underlying infrastructure to enable the Big Data revolution. As an example a $10 million award under the Expeditions in Computing programme was given to the University of California, Berkeley to integrate algorithms, machines and people to turn data into knowledge and insight. The objective is to develop new scalable machine-learning algorithms and data management tools that can handle large-scale and heterogeneous datasets, novel data center-friendly programming models, and an improved computational infrastructure. Another example is the NSF geosciences programme called EarthCube, supported by the CIF21 framework, which is developing a community-guided cyberinfrastructure to integrate Big Data across geosciences. A key requirement is to integrate data from disparate locations and sources with eclectic structures and formats that has been stored as well as captured in real time.

NSF is also supporting proposals under a Big Data Solicitation in collaboration with the National Institutes of Health (NIH). This programme aims to extract and use knowledge from collections of large data sets in order to accelerate progress in science and engineering research. Specifically, it will fund research to develop and evaluate new algorithms, statistical methods, technologies, and tools for improved data collection and management, data analytics and e-science collaboration environments.

In the near term, NSF is providing opportunities and platforms for science research projects to develop the appropriate mechanisms, policies and governance structures to make data available within different research communities. In the longer term, the aim is to build a larger-scale national framework, for the sharing of data among disciplines and institutions. Considering the longer term NSF is supporting a focused research group that brings together statisticians and biologists to develop network models and automatic, scalable algorithms and tools to determine protein structures and biological pathways. At an educational level NSF is funding a research training group in Big Data that will provide training for undergraduates, graduates and postdoctoral fellows to use statistical, graphical and visualization techniques for complex data. In addition, cross-disciplinary efforts are being encouraged such as data citation to increase opportunities for the use and analysis of data sets; participation in an Ideas Lab to explore ways to use Big Data to enhance teaching and learning effectiveness; and the use of NSF's Integrative Graduate Education and Research Traineeship (IGERT) mechanism to educate and train researchers in data enabled science and engineering.

7.4.3 NIH BD2K Centers

BD2K funds research and training activities that support the use of Big Data to advance biomedical research and discovery [306]. This includes efforts in enhancing training, resource indexing, methods and tools development, and other data science-related areas. The BD2K Centers of Excellence programme has established 11 Centers of Excellence for Big Data Computing and one Center that is a collaborative project with the NIH Common Fund LINCS programme, called the LINCS-BD2K Perturbation Data Coordination and Integration Center. The Centers are located all across the United States. They are large-scale projects aiming to develop new approaches, methods, software tools, and related resources. The Centers also provide training to advance Big Data science in the context of biomedical research. The 12 BD2K Centers function as a consortium and collaborate with one another, with the purpose of furthering the field of biomedical data science research. To harness the full potential of Big Data scientists must be able to readily find, cite, and access existing data and other digital objects, such as software. There is no existing infrastructure or incentive available for this which presents barriers to data use and sharing. This leads to duplication of effort and makes it difficult to see where there is sparse research coverage. Funding is thus being provided to address resource discovery, citation, and access. Training and Targeted Software Development awards are also being made to fund software and tools development to tackle data management, transformation, and analysis challenges in the biomedical research community.

7.4.4 DARPA

The Defense Advances Research Projects Agency (DARPA) is promoting the creation of open-source software tools that can help with the processing and analysis of Big Data, providing US$3m in funding to the software company Continuum Analytics. The company is using the open source Python programming language which is currently used by the oil and gas, physics and signal processing industries. The plan is to add support for large, multi-dimensional arrays and matrices and create visualization techniques for the interactive exploration of large multi-dimensional data sets. DARPA's XDATA programme [307] is focused heavily on the warfighting environment, where a "virtual" net of sensors and communications systems offer levels of battlefield awareness. In times of battle, data usage can see extreme spikes and current Department of Defence systems cannot handle or analyse the information that results from this. There is a need to efficiently fuse, analyse and disseminate the massive volumes of data this network produces. The volume and characteristics of the data, and the range of applications for data analysis, require a fundamentally new approach to data science, analysis and incorporation into mission planning, on timelines consistent with operational tempo.

Notably DARPA will allow results produced by XDATA funding to be used by other US government agencies. The tools are seen as solid foundation for continued innovation in the rapid analysis and visual exploration of massive, interconnected data from heterogeneous sources. XDATA plans to release open-source software toolkits for the applied mathematics, data visualization and computer science communities to address challenges being raised by Big Data.

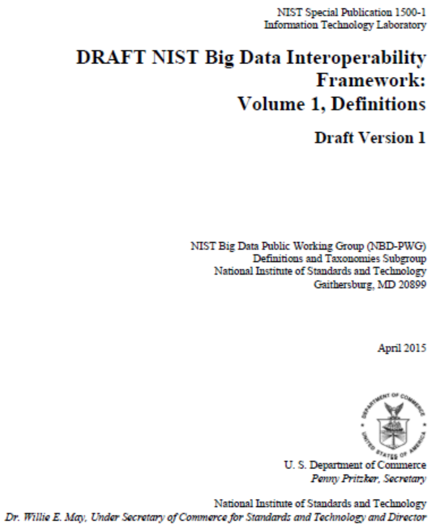
7.4.5 DoD

Big Data is major priority for agencies within the Department of Defense, with many opportunities for contractors working on intelligence gathering, analysis and cyber-security. Deltek forecasts that Defense spending on Big Data will rise steadily for the rest of the decade at a compound annual growth rate (CAGR) of 8.7%. Advanced analytics and technologies like distributed computing are fast becoming integral components of modern, networked weapons systems. This shift reflects not only the growing complexity of weapons but also of the command and control capabilities employed by the military. Faced with dwindling numbers of personnel, all branches of the DoD are turning to networked and unmanned weapons commanded and controlled from a distance. The Department of Defense's 2016 budget request includes $25 million more for Big Data-related research and development than 2015 and all of the military services are funding R&D efforts related to Big Data with a number of projects in the 2016 Defense Research, Development, Test, and Enhancement (RDT&E) budget.

There are also a number of large defence contracts targeting Big Data. The US Cyber Command and the General Services Administration has issued a request for information seeking support for the Cyber National Mission Force [308]. The RFI requests analysis capabilities to fuse "reports from multiple intelligence sources (HUMINT, SIGINT, IMINT, MASINT) to provide intelligence preparation of the battlespace, target development, and early warning of emerging threats." The Defense Information Systems Agency (DISA) [309] is also expected to release a request for proposals for new Joint Management System (JMS) software that will include advanced analytics capabilities. The JMS is critical to the secure functioning of DoD's Joint Regional Security Stacks. The new commercial software should have the ability to "harvest security insights from data that is not intuitively security-related." DISA is also looking for Big Data analytics to add to its Cyber-Security Advanced Analytics Cloud (CSAAC) which defends DoD networks where they connect to the Internet. Contractors must provide software with advanced analytical capabilities that is open source and commercial-off-the-shelf. Northrop Grumman was awarded a $74 million task order in March 2015 for operation of the Acropolis Big Data storage portion of the CSAAC.

Defense officials have also announced the establishment of a Defense Insider Threat Management and Analysis Center (DITMAC) to identify and mitigate the security challenges posed by insider threats. Developed in the aftermath of the 2013 shooting at the Washington Navy Yard it will utilise an array of predictive analytics that facilitate the identification of insider threats before they become a major hazard.

7.4.6 NIST Big Data Public Working Group (NBD-PWG)


NIST Special Publication 1500-1
Information Technology Laboratory

DRAFT NIST Big Data Interoperability Framework:
Volume 1, Definitions

Draft Version 1

NIST Big Data Public Working Group (NBD-PWG)
Definitions and Taxonomies Subgroup
National Institute of Standards and Technology
Gaithersburg, MD 20899

April 2015

U. S. Department of Commerce
Penny Pritzker, Secretary

National Institute of Standards and Technology
Dr. Willie E. May, Under Secretary of Commerce for Standards and Technology and Director

Figure 65. NIST Big Data Interoperability Framework

NIST is leading the development of a Big Data Technology Roadmap [310]. This roadmap will define and prioritise requirements for interoperability, portability, reusability, and extensibility for Big Data analytic techniques and technology infrastructure in order to support secure and effective adoption of Big Data (See Figure 65). To help develop the ideas in the Big Data Technology Roadmap, a community of interest has been formed from industry, academia, and government. The goal is to develop consensus on definitions, taxonomies, secure reference architectures, and the technology roadmap. The aim is to create vendor-neutral, technology and infrastructure agnostic deliverables that enable stakeholders to pick-and-choose the best analytics tools for their processing and visualization requirements on the most suitable computing platforms and clusters. At the same time data needs to flow between the stakeholders in a cohesive and secure manner. A number of deliverables have been defined covering a range of key areas: Big Data Definitions, Big Data Taxonomies, Big Data Requirements, Big Data Security and Privacy Requirements, Big Data Security and Privacy Reference Architectures, Big Data Reference Architectures, and a Big Data Technology Roadmap.

7.5 Rest of the World

7.5.1 Worldwide Open Data Readiness

Country	Rank	Readiness Sub-Index	Implementation Sub-Index	Impact Sub-Index	ODB Overall
United Kingdom	1	100.00	100.00	79.91	100.00
United States	2	95.26	86.67	100.00	93.38
Sweden	3	95.20	83.14	71.95	85.75
New Zealand	4	81.88	65.49	89.81	74.34
Norway	5	91.88	70.98	46.15	71.86
Denmark	5	83.54	70.20	55.73	71.78
Australia	7	87.88	64.71	51.19	67.68
Canada	8	79.11	63.92	51.59	65.87
Germany	9	74.50	63.14	53.81	65.01
France	10	79.39	64.31	39.07	63.92
Netherlands	10	85.92	67.03	21.42	63.86
Korea (Rep. of)	12	77.19	54.90	24.58	54.21
Iceland	13	62.99	52.94	26.45	51.01
Estonia	14	72.38	49.41	24.00	49.45
Finland	14	91.19	41.18	40.87	49.44
Japan	14	76.99	47.08	27.94	49.17
Spain	17	67.48	49.41	21.13	48.19
Austria	18	68.56	39.22	48.62	46.03
Israel	18	61.82	45.88	25.36	45.58
Italy	20	50.39	42.75	45.69	45.30
Russia	20	54.43	40.39	48.86	44.79
Switzerland	22	65.11	41.57	26.80	43.24
Czech Republic	22	61.83	40.00	35.36	43.18
Kenya	22	49.70	45.88	21.55	43.06
Mexico	25	49.10	45.49	8.37	40.30
Chile	25	66.70	30.22	18.27	40.11
Portugal	27	60.38	38.04	19.25	38.63
Brazil	28	66.03	32.16	27.87	36.83
Singapore	29	70.28	35.29	8.97	36.29
Ireland	29	61.61	32.55	23.82	35.76
Thailand	31	38.09	39.22	14.88	35.33
Argentina	31	40.08	30.47	17.29	35.00
Belgium	31	72.01	28.63	25.64	34.80
India	34	57.36	33.73	9.87	33.38

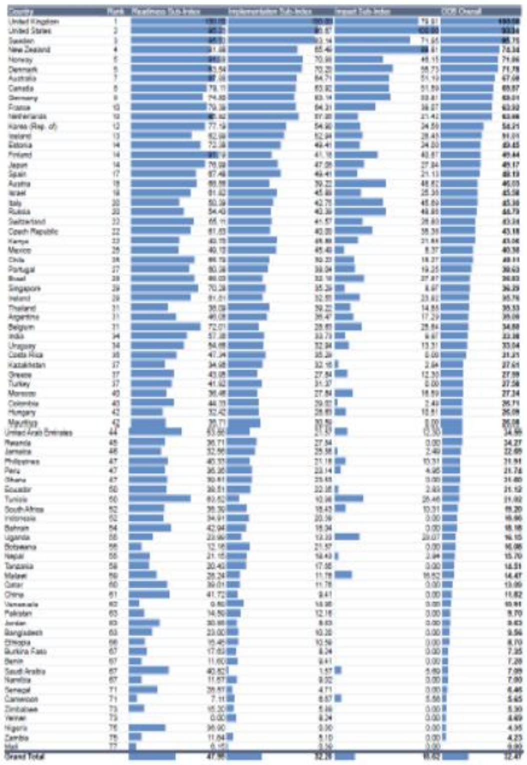


Figure 66. Open Data Barometer 2013 Global Report

The Open Data Barometer [311] published by the Open Data Institute UK and the World Wide Web Foundation is a measure of the prevalence and impact of open data initiatives around the world in 88 countries (See Figure 66). It analyses global trends, and also ranks countries and regions according to an index that considers: readiness to secure the benefits of Open Data; actual levels of implementation; and the impact of such initiatives, combining peer-reviewed expert survey data and secondary indicators. The Barometer ranks the UK as the most advanced country for open data readiness, implementation and impact, scoring above the USA (2nd), Sweden (3rd), New Zealand (4th), Denmark and Norway (joint 5th). The leading developing country is Kenya (21st), ranking higher than rich countries such as Ireland (29th) and Belgium (31st).

The index highlights a number of things. Firstly, Open Data is still in its infancy and many governments are still at the stage of exploring how providing open data can unlock value, stimulate innovation and increase transparency and accountability. The availability of truly open data, however, remains low with only around 10% of the surveyed countries providing published data in bulk machine-readable forms, and under open licenses. Thirty-one countries have at least one open dataset, and just over 50% of the datasets surveyed among the 11 top-ranked countries qualified as being “fully open”. This makes it difficult for users to access, process and work with government data. In particular, entrepreneurs face significant legal uncertainty over their rights to build businesses on top of government datasets. In general, leading countries in the ODB are investing in the creation of “National Data Infrastructures” to provide a foundation for public and private innovation and efficiency. Mid-ranking countries have put in place an open data portal and run competitions and events to catalyse the re-use of data, however, they have often failed to make key datasets available, failed to have strong “Right to Information laws” which prevents citizens from using open data, or have weak or missing Data Protection Laws. Low-ranking countries have not yet started to engage with open data. Many developing countries lack basic foundations such as well-managed and digitised government datasets.

7.6 Need for Regulations

7.6.1 European Data Protection Directive

The European Data Protection Directive (Directive 95/46/EC) [312] protects an individual with respect to processing of personal data and on the free movement of such data. This EU directive which regulates the processing of personal data was adopted in 1995. It is an important part of the European approach to privacy and human rights law. A new draft General Data Protection Regulation was announced by Europe on 25 January 2012 [313] which replaced the Data Protection Directive. In particular the right to privacy is a highly developed area of law in Europe. All the member states of the EU are also signatories of the European Convention on Human Rights (ECHR). Article 8 of the ECHR provides a right to respect "private and family life, his home and his correspondence". The European Court of Human Rights has given this article a very broad interpretation in its jurisprudence. In 1980, in an effort to create a comprehensive data protection system throughout Europe, the Organisation for Economic Cooperation and Development (OECD) issued "Recommendations of the Council Concerning Guidelines Governing the Protection of Privacy and Trans-Border Flows of Personal Data” [314]. This sets out seven key principles governing protection of personal data were:

1. Notice—data subjects should be given notice when their data is being collected;
2. Purpose—data should only be used for the purpose stated and not for any other purposes;
3. Consent—data should not be disclosed without the data subject’s consent;
4. Security—collected data should be kept secure from any potential abuses;
5. Disclosure—data subjects should be informed as to who is collecting their data;
6. Access—data subjects should be allowed to access their data and make corrections to any inaccurate data; and
7. Accountability—data subjects should have a method available to them to hold data collectors accountable for not following the above principles.

Although these guidelines set out a number of key principles for Europe they were not binding and data privacy laws still vary widely across Europe. The US endorsed the OECD recommendations but did not implement them. In 1981 the Convention for the Protection of Individuals with regards to Automatic Processing of Personal Data [315] was negotiated within the Council of Europe. This convention obliges the signatories to enact legislation concerning the automatic processing of personal data. This states that personal data should not be processed at all, except when certain conditions are met with respect to transparency, legitimate purpose and proportionality. The EC also realised that diverging data protection legislation within different EU member states impeded the free flow of data within the EU which resulted in the Data Protection Directive being proposed.

7.6.2 Safe Harbour

The international “Safe Harbour Privacy Principles” have been defined with the aim of enabling some US companies to comply with privacy laws that protect European Union and Swiss citizens [316]. According to the Safe Harbour principles US companies who store European customer data may self-certify that they adhere to 7 key principles that comply with the EU Data Protection Directive and with Swiss requirements. This has been jointly developed by the US Department of Commerce, the European Union and the Federal Data Protection and Information Commissioner of Switzerland. The European Commission came to the conclusion in 2000 that the United States' principles complied with the EU Directive but this was ruled as being invalid in October 2015 by the European Court of Justice when a customer complained that his Facebook data was insufficiently protected. A new framework was agreed for transatlantic data flows on 2nd February 2016, known as the "EU-US Privacy Shield" with the aim of meeting the requirements set out by the European Court of Justice. The new arrangement has stronger obligations on US companies to protect the personal data of Europeans and this is backed up by stronger monitoring and enforcement by the U.S. Department of Commerce and Federal Trade Commission. In particular, the new arrangement includes commitments by the US that public authorities can only gain access to personal data transferred subject to clear conditions, limitations and oversight, preventing generalised access. Europeans will also be able to raise an enquiry or complaint with a new Ombudsperson.

The Commission promised to deliver the full text of the Privacy Shield agreement [317] by the end of February 2016. This was done and in March 2016, the authorities met to decide if the "shield" is strong enough. Here it was highlighted that the UK, France and Hungary might not meet the ECJ's Safe Harbour test. Notably several countries in Europe, e.g. France, Germany, Poland, the UK and the Netherlands permit certain types of surveillance that are not targeted at identified suspected individuals. Here it is possible to apply 'keywords' or 'selectors' to large communications data flows crossing their territory and also intercept external communications that are not targeted at specific individuals. This has been the subject of recent court cases in the UK where it is possible for communications received via submarine cables to be intercepted. In France, a new law passed in 2015 allows spies to use International Mobile Subscriber Identity catchers to intercept telephone data, access emails and online communications of anyone suspected of being linked to terrorism with little or no judicial oversight. There have also been other cases with respect to privacy. For instance the European Court of Human Rights found that Hungary violated Article 8 of the Convention on Human Rights [317] – the right to respect for private and family life was contravened by a law that allows police to carry out secret intelligence gathering on the grounds of national security, which can search and record private electronic communications without consent and without any judicial oversight. In other European countries the surveillance framework is more complex as there are a number of different overlapping laws with varying checks and balances. For instance, in Germany a parliamentary control panel oversees strategic telecommunications surveillance carried out by the Federal Intelligence Service (BND).

A key problem is that while it is possible for lawyers and academics to argue the validity of national laws when it comes to meeting the privacy shield there is no power when it comes to national security.

Google, Facebook and Amazon are keen for EU data regulators to define in specific terms what Europe's new privacy rules demand. As the privacy shield is still under debate for business the safe option at present is to keep European data in Europe. 4500 US companies relied on the old Safe Harbour agreement for legal cover which suddenly became invalid [317]. Currently only 27 % of cloud services provide Europe-only data storage. This has actually increased from 14 % in June 2015 according to a survey of 16,000 cloud services by Skyhigh Networks. Companies are looking at moving their data centres to Europe and particularly Germany where there are some of the strictest data privacy laws. A challenge is that the storage, processing and personnel must also all be local. This presents problems as although some companies such as Adobe already has European data centres, its cloud services may still require transfers to the US to provide some features. It may also be necessary for support personnel to access data stored in Europe from the US (or other countries) which requires the rules on data transfer outside the EU to be met.

Encryption is thought to be a way forward. For instance ID data can be encrypted before it enters the cloud with the data owner keeping the encryption keys. This gets around Safe Harbour and Privacy Shield laws but only 1% of cloud companies offer this as it is expensive. Self-encryption is possible but this often prevents the data being processed by needed cloud services.

7.6.3 US 14th Amendment of the Constitution

The US has no single data protection law comparable to the EU's Data Protection Directive. In the US privacy legislation has developed in an *ad hoc* basis when required by certain sectors and circumstances. Fundamental within the US is "The Fourteenth Amendment (Amendment XIV)" which was adopted in the United States Constitution on July 9, 1868 [318]. The amendment addresses citizenship rights and equal protection of the laws, and was proposed in response to issues related to former slaves following the American Civil War. The first section of The Fourteenth Amendment is one of the most litigated parts of the Constitution and has been used as the basis or a number of landmark decisions, e.g. such as Roe vs. Wade (1973) regarding abortion, Bush vs. Gore (2000) regarding the 2000 presidential election, and Obergefell vs. Hodges (2015) regarding same-sex marriage. Critically the amendment limits the actions of all state and local officials, including those acting on behalf of officials.

The right to privacy or the “right to be left alone” is a key area within the US. While not explicitly stated in the U.S. Constitution, some of the amendments provide a degree of protection. Privacy is most often protected by statutory law in the US, e.g. the Health Information Portability and Accountability Act (HIPAA) protects a person's health information, and the Federal Trade Commission (FTC) enforces the right to privacy in various privacy policies and privacy statements.

There is a challenge, however, when there is a need to balance privacy against the needs of public safety and improving the quality of life, e.g. seat-belt laws and motorcycle helmet requirements. Most Americans accept that government surveillance and collecting of personal information is necessary. The right to privacy often relates to the right to personal autonomy where an individual has the right to choose whether or not to engage in certain acts or have certain experiences. Several amendments to the U.S. Constitution have been used in varying degrees of success in determining a right to personal autonomy:

- The First Amendment protects the privacy of beliefs
- The Third Amendment protects the privacy of the home against the use of it for housing soldiers
- The Fourth Amendment protects privacy against unreasonable searches
- The Fifth Amendment protects against self-incrimination, which in turn protects the privacy of personal information
- The Ninth Amendment says that the "enumeration in the Constitution of certain rights shall not be construed to deny or disparage other rights retained by the people." This has been interpreted as justification for broadly reading the Bill of Rights to protect privacy in ways not specifically provided in the first eight amendments.

However, the right to privacy is most often cited in the Due Process Clause of the 14th Amendment, which states:

> *No state shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any state deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.*

However, the protections have been narrowly defined and are usually interpreted as only applying to family, marriage, motherhood, procreation and child rearing. The controversial *Roe v. Wade* case in 1972 case established the right to privacy as fundamental, and required that any governmental infringement of that right had to be justified by a compelling state interest.

In the US a person has the right to determine what sort of information about them is collected and also how that information is to be used. In the marketplace this is enforced by laws intended to prevent deceptive practices and unfair competition. The Privacy Act of 1974 [319] prevents unauthorised disclosure of personal information held by the federal government. A person has the right to review their own personal information, ask for corrections and be informed of any disclosures. This has also been imposed upon financial institutions in the Financial Monetization Act (1999) which requires financial institutions to provide their customers with an explanation of what kind of information is being collected and how it is being used. Safeguards are also required to protect customer information, e.g. the Fair Credit Reporting Act, protects personal financial information collected by credit reporting agencies. The act puts limits on who can access the information and requires agencies to have simple processes by which consumers can get their information, review it and make corrections.

Online privacy is also important in the US and Internet users can protect their privacy by taking actions that prevent the collection of information, for instance to not allow tracking cookies. Browsers and social media platforms, e.g. Facebook and Twitter, allow user selected privacy settings, from sharing everything to only sharing with friends. At a minimum level this can be only a name, gender and profile picture and increasingly citizens are aware that it is necessary to protect personally identifiable information to prevent identity theft.

Also within the US the Children's Online Privacy Protection Act (COPPA) [320] enforces a parent's right to control what information websites collect about their children. In particular websites that target children younger than 13 or knowingly collect information from children must post information on their privacy policies and also get parental consent before collecting information from children. Parents can thus decide how such information can be collected.

As well as a right to privacy there is also a right to publicity. Here there is a right to control the use of his or her identity for commercial promotion. Unauthorised use of someone’s name or likeness is recognised as an invasion of privacy. This is classed into 4 areas: intrusion, appropriation of name or likeness, unreasonable publicity and false light. For instance, if a company falsely uses a person's photo in an advert claiming that the person endorses a product, the person can file a lawsuit claiming misappropriation.

A challenge is that the Supreme Court in the US approaches the right to privacy and personal autonomy on a case-by-case basis. Notably public opinion is constantly changing regarding relationships and activities. The boundaries of personal privacy are also changing due to social media and a move towards "sharing." As a consequence the definition of the “right to privacy” is constantly changing.

7.7 Standards

7.7.1 ISO/IEC JTC 1 Study Group on Big Data (BD-SG)

The ISO/IEC Joint Technical Committee (JTC) 1 on Information Technology has set up a Working Group (WG) focused on standardization for Big Data. The American National Standards Institute (ANSI) holds the secretariat to JTC 1. The objective is to identify standardization gaps, develop foundational standards for Big Data, develop and maintain liaisons with all relevant JTC 1 entities and raise awareness of standardisation efforts. A Study Group has been mapping the existing landscape for key technologies and relevant standards/models/studies /use cases and scenarios for Big Data from JTC 1, ISO, IEC and other standards setting organizations. It has also identified key terms and definitions commonly used in the area of Big Data. The current status of Big Data standardization considering market requirements, standards gaps, and standardization priorities have been identified and recommendations have been made in 2014 [321].

7.7.2 IEEE Standards Association

The IEEE Standards Association introduced a number of standards related to Big Data applications enabled by the Internet of Things and a specific standard is under development [322]. This standard "IEEE Standard for an Architectural Framework for the Internet of Things (IoT)" defines the relationships between devices used in industries, including transportation and health care. It also provides a blueprint for data privacy, protection, safety, and security, as well as a means to document and mitigate architecture divergence.

7.7.3 ITU

The International Telecommunications Union (ITU) created its first standards for Big Data services, entitled 'Recommendation ITU-T Y.3600 "Big Data - cloud computing based requirements and capabilities" in 2015 [323]. Recommendation Y.3600 provides requirements, capabilities and use cases of cloud computing based Big Data as well as its system context.

.

8 IoT/CPS

8.1 European IoT/CPS Drivers and Policy Initiatives

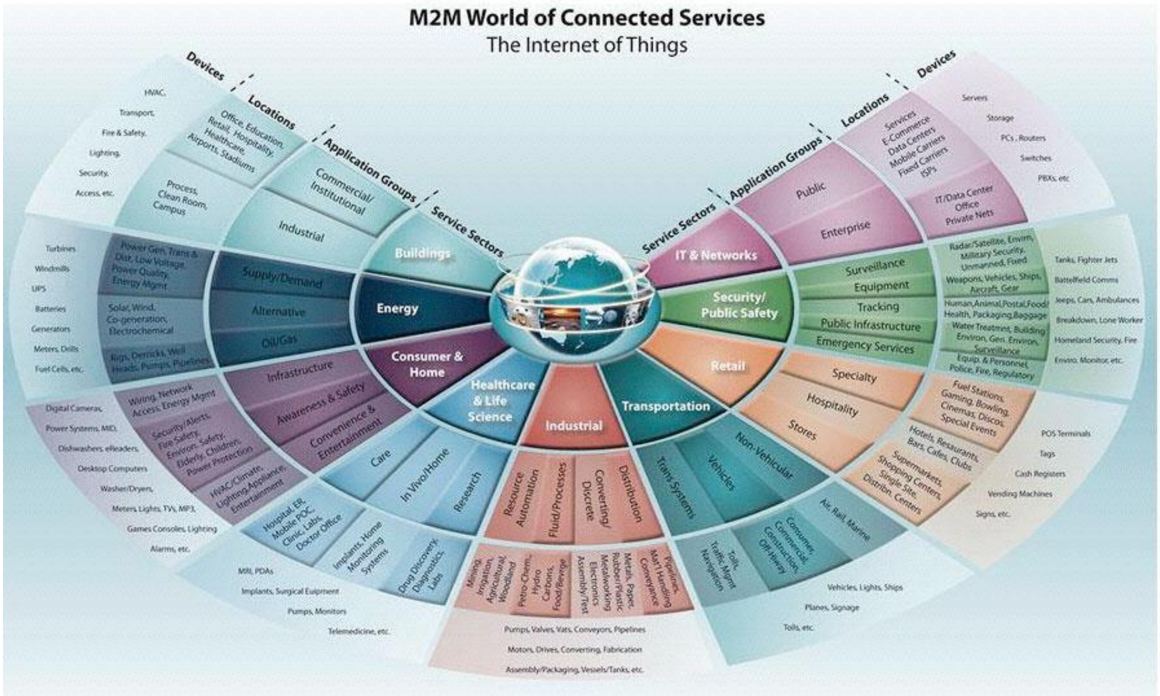


Figure 67. M2M World of Connected Services

Within Europe Support for development and integration of Cyber-Physical Systems and the Internet of Things is seen as essential for the future. As the embedded world meets the Internet world there will be an increasing number of interacting systems with strong connectivity utilised in both society and in industry (See Figure 67). Connectivity between embedded systems and computing devices is predicted to grow massively over the coming years. Gartner [324] for instance, estimates that there will be 26 billion connected devices (excluding PCs, tablets and smartphones) by 2020 world-wide, and even higher predictions of 40-50 billion devices are being made by other analyst companies. This equates to a global market value of $1.9 trillion, of which 80% is expected to come from services.

Europe is a world leader in the area of time-critical and safety-critical systems and to maintain this position there is a need to be able to design, develop and deploy highly distributed and connected digital technologies [325]. Underlying this is a need for the development and introduction of platforms for CPS deployment. This is seen as key to the future but only through building a supporting ecosystem of developers and users can a platform flourish and become successful. Europe has a strong position in ICT market with an ecosystem of world leading suppliers and systems integrators. The embedded systems industry alone creates 50,000 new jobs every year and Europe accounts for 34% of world production of embedded systems with particular strengths in the automotive sector, aerospace and health.

CPS/IoT/digitalisation is expected to change our everyday lives and open up new business opportunities with opportunities to provide efficient, environmentally friendly, autonomous and safe mobility in the automotive, aeronautics, rail, maritime and logistics sectors; greater efficiency in management and operations for process automation, manufacturing, conventional/renewable power plants, energy conversion, smart grids and smart metering; greater benefits to citizens via smart, safe and secure cities, energy efficient buildings and green infrastructure (traffic management, lighting, water and waste management); and smart devices and services for smart home functionality, home monitoring, health services and assisted living.

To support this European Commissioner Gunther Oettinger in his speech on “A Digital Single Market Strategy”, highlights the need for Industrial Leadership in the Digital Economy. The second key pillar of this strategy is to create leadership in platforms for digital industry with the objective to “ensure the availability of state-of-the-art open and interoperable platforms that any business can use to make its products, processes or services ready for the digital age”. The development of such platforms requires collaboration between actors across value chains, including users and the supply industry so it is necessary to bring these together. To achieve this the ECSEL Joint Technology Initiative and PPPs in Factories of the Future, 5G and Big Data have been set up and it is planned to launch at least 5 large-scale platform projects per year until 2018. Investment in platform-building in Horizon 2020 is expected to reach more than €800 million in the next five years with matching investment from industry and government of €3 billion until 2020. To build the Digital Single Market it is also necessary to address the skills gap and legal issues. Collaboration and consensus building on standards and platforms for strategic positioning of European industry is required as it is essential to have common standards and interoperable solutions throughout the products and services life cycles.

Within the European Union the concept of “Smart Everything Everywhere” is very much a key concept for the future. The majority of IoT research and development to date has focused on sensors and on providing connectivity, whereas the real value to users and society is from using the information provided by the sensors and networks in a smart fashion (and in connecting sensing to actuation to create a CPS). Connectivity from the Internet of Things is thus seen as an enabling technology for Cyber-Physical Systems that close the loop from sensors to actually influence users and physical systems. The enormous potential of these technologies has been recognised by the European Union, as CPS and IoT are key pillars of the Europe 2020 initiative Digital Agenda for Europe [326] and of other research and innovation programmes, e.g. the ECSEL Joint Undertaking, EUREKA/ITEA, and the ARTEMIS Industry Association.

Cyber-Physical Systems are an important area for Europe with a 410B€ market, 4 million associated jobs worldwide of which one quarter are in Europe, contributing both to the employment and to the quality of life and the industrial competitiveness across all sectors.

A major initiative launched in April 2016 is the Digitising European Industry initiative [327]. This aims to link national initiatives for the digitisation of industry and related services and to boost investment through strategic partnerships and networks. Although many areas within Europe have embraced the take up digital technologies and processes some sectors are lagging behind such as construction, agro-food, textiles and steel. SMEs are particularly lagging behind in their digital transformation. It is estimated that digitisation of products and services will add more than €110 billion of revenue for industry per year in Europe in the next five years. To support this several EU Member States have launched strategies to support the digitisation of industry, however, the aim of the Commission is to coordinate activities across Europe to avoid fragmentation. Key areas being addressed are standards for 5G, cloud computing, Internet of Things, data technologies and cyber-security. The Commission will also set up a European cloud to give Europe's 1.7 million researchers and 70 million science and technology professionals a virtual environment to store, manage, analyse and re-use a big amount of research data. The Commission intends to:

- help coordinate national and regional initiatives on digitising industry
- focus investments in EU's public-private partnerships

- invest €500 million in a pan-EU network of digital innovation hubs (centres of excellence in technology) where businesses can obtain advice and test digital innovations.
- set up large-scale pilot projects to strengthen Internet of Things, advanced manufacturing and technologies in smart cities and homes, connected cars or mobile health services.
- adopt future-proof legislation that will support the free flow of data and clarify ownership of data generated by sensors and smart devices. The Commission will also review rules on safety and liability of autonomous systems.
- present an EU skills agenda that will help give people the skills needed for jobs in the digital age.

It is expected that €50 billion of public and private investments will be mobilised in support of the digitisation of industry. In order to create critical mass the Commission has an approach of "clustering" projects in key areas. In the following sections key clusters and projects are highlighted in the areas of IoT and CPS.

8.2 IOT Initiatives in Europe

8.2.1 IERC European Research Cluster on the Internet of Things

The IERC (IoT European Research Cluster) [328] is bringing together EU-funded projects with the aim of defining a common vision for IoT technology, identifying research challenges, and coordinating and encouraging the convergence of ongoing work. The cluster includes over 40 European projects including CLOUT, VITAL, SOCIOTAL, RERUM, COSMOS, CITY PULSE, ALMANAC, SMARTIE, SMART-ACTION, FITMAN, ASPIRE, CASCADAS, CONFIDENCE, CuteLoop, DACAR, EPoSS, EU-IFM, EURIDICE, GRIFS, HYDRA, IMS2020, Indisputable Key, iSURF, LEAPFROG, PEARS Feasibility, PrimeLife, RACE networkRFID, SMART, StoLPaN, SToP, TraSer, WALTER, IOT-A, INTREPID, IOT@Work, ELLIOT, SPRINT, NEFFICS, IOT-I and CASAGRAS2. Communication between these projects is seen as essential to create a competitive industry and to provide secure, safe and privacy preserving deployment of IoT. Additionally, the IERC also links to Member States' initiatives and engages with policy activities such as BRIDGE, AITPL, AMI-4-SME, CE-RFID, CoBIS, Dynamite, PRIME, PROMISE and SMMART, and cooperates with other large initiatives such as FIA, 5G, ARTEMIS-IA, AENEAS, EPoSS and the ECSEL JU.

Although the IERC has a European aim to enhance Europe's competitiveness and to drive the development of an information based economy and society it also has a global dimension facilitating knowledge sharing at the global, industrial and organisational level to encourage and exchange best practices and new business models that are emerging in different parts of the world. The main objectives of the IERC are to:

- Establish a cooperation platform and develop a research vision for IoT activities in Europe and become a major entry and contact point for IoT research in the world.
- Define an international strategy for cooperation in the area of IoT research and innovation and have an overview of the research and innovation priorities at the global level.
- Coordinate the cooperation activities with other EC Clusters and ICT projects.
- Coordinate and align the SRIA agenda at the European level with the developments at the global level.
- Organise debates/workshops leading to a better understanding of IoT and Future Internet, 5G, cloud technology, and adoption.

8.2.2 AIOTI

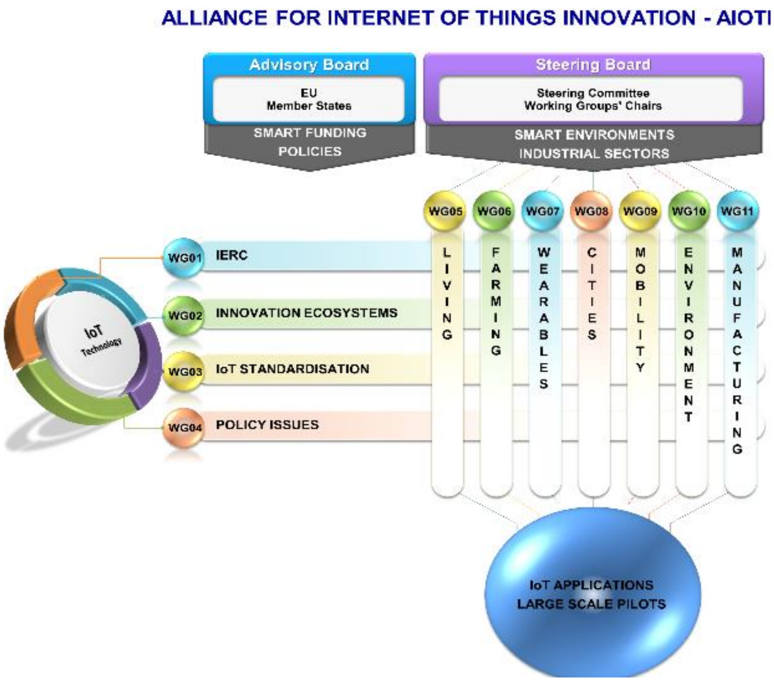


Figure 68. AIOTI

The Alliance for the Internet of Things (AIOTI) [329] was launched by the European Commission with key IoT players with the aim of creating a European IoT ecosystem that can promote dialogue and interaction to give the EU a lead in the technology. The AIOTI is helping the European Commission prepare strategy for future IoT research (See Figure 68) as well as providing input to innovation and standardisation policies. In particular, it has been providing input to the design of IoT Large Scale Pilots to be funded under H2020.

8.2.3 H2020 Pilot Projects

The European Commission call with €100 million on Internet of Things Large Scale Pilots [330] is aimed at promoting IoT take up in Europe and development of open technologies and platforms to support IoT ecosystems. The Large Scale Pilots are goal driven with the aim of using IoT approaches for real-life industrial/societal challenges. The areas include smart living environments for ageing well, smart farming and food security, wearables for smart ecosystems, reference zones in EU cities and autonomous vehicles in a connected environment. The Large Scale Pilots will involve stakeholders from both the supply and demand side and contain all the technology development, testing, integration and innovation activities for use, application and deployment of IoT. Coordination and support actions are also being funded to encourage cooperation and support cross-fertilisation between pilots and use cases. The pilot areas are:

- Pilot 1: Smart living environments for ageing well (EU contr. up to €20 million)
- Pilot 2: Smart Farming and Food Security (EU contr. up to €30 million)
- Pilot 3: Wearables for smart ecosystems (EU contr. up to €15 million)
- Pilot 4: Reference zones in EU cities (EU contr. up to €15 million)

- Pilot 5: Autonomous vehicles in a connected environment (EU contr. up to €20 million)

A further call will be made in 2017 on sophisticated platform architectures for smart objects, embedded intelligence, and smart networks. International joint IoT calls are also being supported with Japan, South Korea, China and Brazil.

8.2.4 BUTLER

Under the EU FP7 call the BUTLER project (uBiquitous, secUre inTernet-of-things with Location and contEx-awaReness) [331] was funded [€14.9 million] to develop secure and smart life assistant IoT applications using a context and location aware, pervasive information system. BUTLER is now finished and BUTLER components are available on the Open Platform Portal [332]**.** BUTLER emphasised pervasiveness**,** context-awareness and security for IoT and developed new technologies to form a “bundle” of applications, platform features and services. This included improving and creating technologies to implement secure, pervasive and context-aware IoT for different scenarios, e.g. Home, Office, Transportation, Health, etc. A new flexible smartDevice**-**centric network architecture was created where platforms (devices) function according to three well-defined categories**:** smartObject (sensors, actuators, gateways), smartMobile (user’s personal device) and smartServers (providers of contents and services), interconnected over IPv6. Several field trials were also performed.

8.2.5 FIWARE

FIWARE [333] is an EU initiative to create a truly open, public and royalty-free architecture and set of open specifications that will allow developers, service providers, enterprises and other organizations develop products. A key aim is to provide an infrastructure to support cost-effective creation and delivery of digital services. These need to provide a high quality of service with guaranteed security. A set of open APIs are being provided to developers to foster innovation and entrepreneurship. FIWARE is seen as a foundation for the Future Internet, cultivating a sustainable ecosystem for both service providers delivering new applications and solutions meeting the requirements of established and emerging use areas, and end users/consumers participating in content and service consumption/creation. The project is developing 16 Future Internet Accelerators including applications on Smart Cities, E-Health, Transport, Energy and Environment, Agrifood, Media and Content, Manufacturing and Logistics, Social and Learning.

8.3 European Initiatives in CPS

8.3.1 ECSEL Joint Undertaking

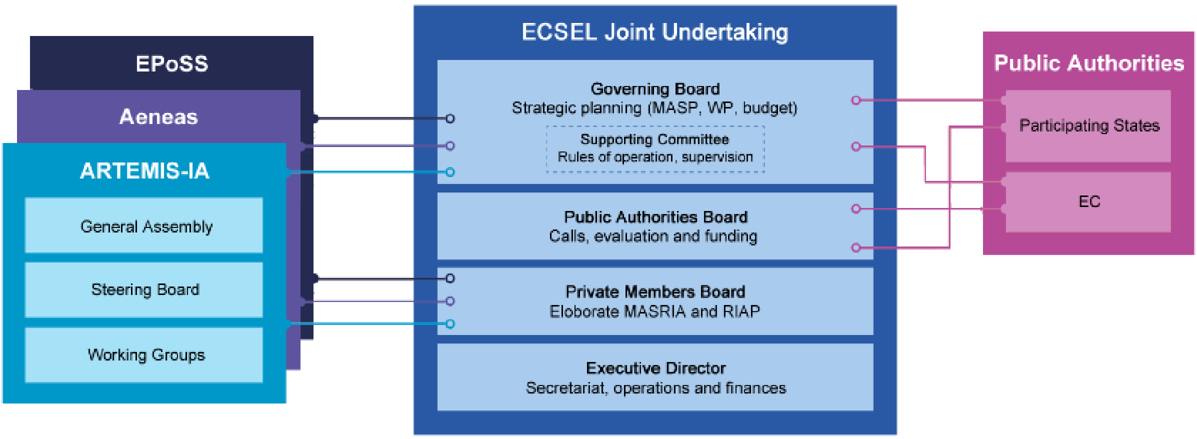


Figure 69. Overview of ECSEL Undertaking

The ECSEL-JU (Electronic Components and Systems for European Leadership) programme [334] was created from a merger of ARTEMIS-JU and the ENIAC-JU in June 2014 and will finish in 2024. ECSEL has coverage from industry in a number of areas including micro-/nanoelectronics, embedded and Cyber-Physical Systems and smart systems as shown in Figure 69. Within the programme projects are funded in several application areas and in key enabling technologies as shown in Figure 70.

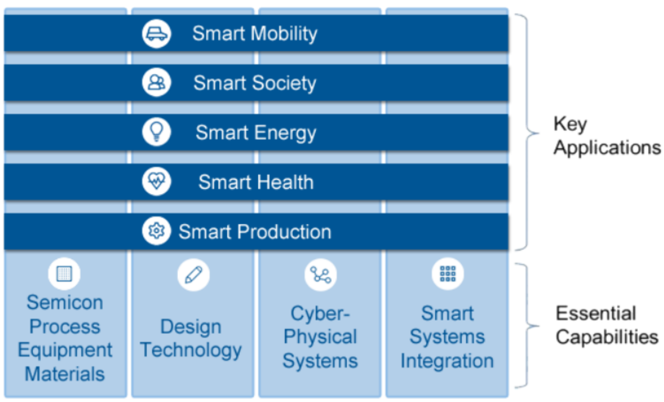


Figure 70. ECSEL Applications and Essential Capabilities

The strategy of ECSEL is decided by a Governing Board which comprises the ARTEMIS Industry Association, AENEAS and EPoSS, participating states and the European Commission. ECSEL makes its own calls to fund R&I projects via the Public Authorities Board. Call topics are agreed by participating states, associated countries and the European Commission.

8.3.2 ARTEMIS-IA

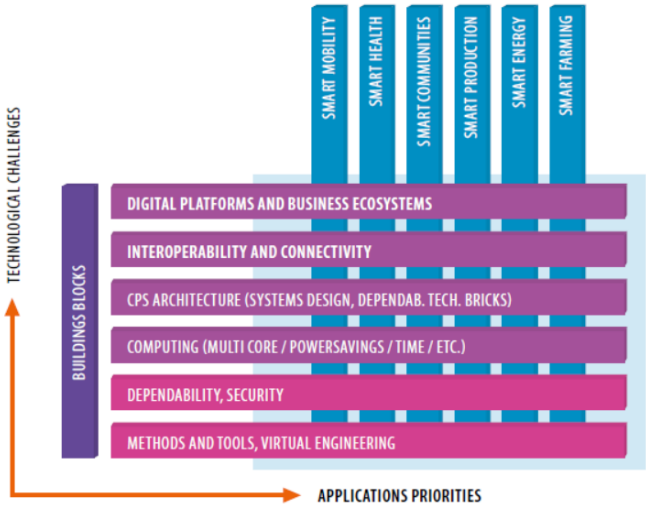


Figure 71. ARTEMIS Strategic Roadmap

The ARTEMIS Industry Association (Advanced Research and Technology for Embedded Intelligence and Systems) [335] is an association of European actors in Embedded & Cyber-Physical Systems. ARTEMIS represents its members (industry, SMEs, universities and research institutes) coordinating strategy and promoting R&I interests to the European Commission and the Public Authorities of the participating states. The Industry Association follows on from the ARTEMIS European Technology Platform and maintains responsibility for the ARTEMIS Strategic Research Agenda (SRA) on Embedded & Cyber-Physical Systems (See Figure 71). There are more than 180 members and associates from all over Europe with a wide range of backgrounds and disciplines being represented.

8.3.3 CRYSTAL - CRitical sYSTem engineering AcceLeration

Funded by the former ARTEMIS Joint Undertaking there are still ongoing projects. The CRYSTAL project [336] is developing an Interoperability Specification (IOS) and a Reference Technology Platform (RTP) as a European standard for safety-critical systems. The project has a budget of €82 million with 71 partners from 10 different European countries including OEMs, suppliers, tool vendors and academia. Driven by industry, CRYSTAL aims to provide mature, ready-to-use integrated tool chains with TRL of 7. A loose coupling between tools is being used to enable sharing and interlinking of data via standardised and open Web technologies. A key aim is to provide common interoperability across various life cycle domains. Real-world industrial use cases from the automotive, aerospace, rail and health sectors are being used as a focus for the work. This builds upon previous successful projects like CEASAR, SAFE, iFEST and MBAT.

8.3.4 EMC2 Embedded Multi-Core Systems for Mixed Criticality Applications in Dynamic and Changeable Real-time Environments

EMC² [337] is another ARTEMIS Joint Undertaking project of note that is an Innovation Pilot Programme on computing platforms for embedded systems. The objective of EMC² is to enable the use of multi-core technology across several embedded systems domains. There are 6 Technology Work Packages and a number of Living Labs in the Automotive, Aerospace, Space, Shipping, Railway and Logistics domains. The aim is to allow cost efficient integration of different applications with different levels of safety and security on a single computing platform in an open context. EMC² is addressing dynamic adaptability in open systems, handling of mixed-criticality applications under real-time conditions, scalability and flexibility, full-scale deployment and management of integrated tool chains, through the entire lifecycle. The project has 99 industry and research partners from 19 European countries with an effort of around 800 person years and a total budget of about €100 million.

8.3.5 I4MS and Competence Centres

Figure 72. I4MS

The EC I4MS initiative (ICT Innovation for Manufacturing SMEs) was launched in 2013 (See Figure 72) with a budget of €77 million [338]. The aim of the initiative is to help SMEs and mid-caps in the manufacturing sector by providing access to competences that can help in assessing, planning and mastering digital transformation and in providing access to innovation networks and best practice examples. Financial support is also directly provided for digital transformation. Underlying this is the idea to foster collaboration across manufacturing value chains through the help of the competence centres and innovation hubs (HPC centres, universities, application oriented research organisations) across Europe. Short duration experiments are funded to transfer know-how and technology from the innovation hubs to SMEs and mid-caps bridging the competence gap and providing financial means to adopt leading edge technology that can be used to bring innovative new products and services to market. Here cross-border experiments are particularly supported with an intention to broaden the field of the application and open up new markets. Participating competence centres benefit as they extend their research oriented activities with industrial projects.

8.3.6 SmartAnythingEverywhere

Figure 73. Smart Everything Everywhere

The Smart Everything Everywhere (SAE) initiative [339] is combining efforts from a number of initiatives (See Figure 73). €25 million of funding is being used to support around 100 experiments with the aim of involving more than 200 SMEs and midcaps in the field of Cyber-Physical Systems (CPS), Internet of Things (IoT) and Smart Systems Integration (SSI). Open calls are made from 4 different initiatives within SAE.

- EuroCPS - A network of design centres boosting and initiating synergies between SMEs, major CPS-platforms and competency providers to capture the emerging markets of IoT products. 30 SME led experiments have been initiated.
- CPSeLabs – A CPS engineering infrastructure, knowledge and tools for realising novel CPS-based products and services. The CPSELabs marketplace provides an open forum for sharing platforms, architectures and SW tools for the engineering of dependable and trustworthy CPS. This innovation initiative funds focused 12-18 month experiments with 3-6 partners.
- Gateone – Is an innovation service for European "smartisation" by SMEs with a focus on bioelectronics technologies. 50 small scale experiments have been funded to deliver innovation concept demonstrators with SMEs engaged in the testing phase.
- Smarter–SI – This initiative provides smart access to manufacturing for systems integration. It utilises a Community Foundry Model (CFM) to accelerate Smart Systems Integration for SMEs to exploit in niche markets (low volume high value) by providing access to design facilities and manufacturing capabilities for prototyping, early validation and first production. A test bed to realise 10 application experiments has been set up.

8.3.7 European H2020 CPS Projects

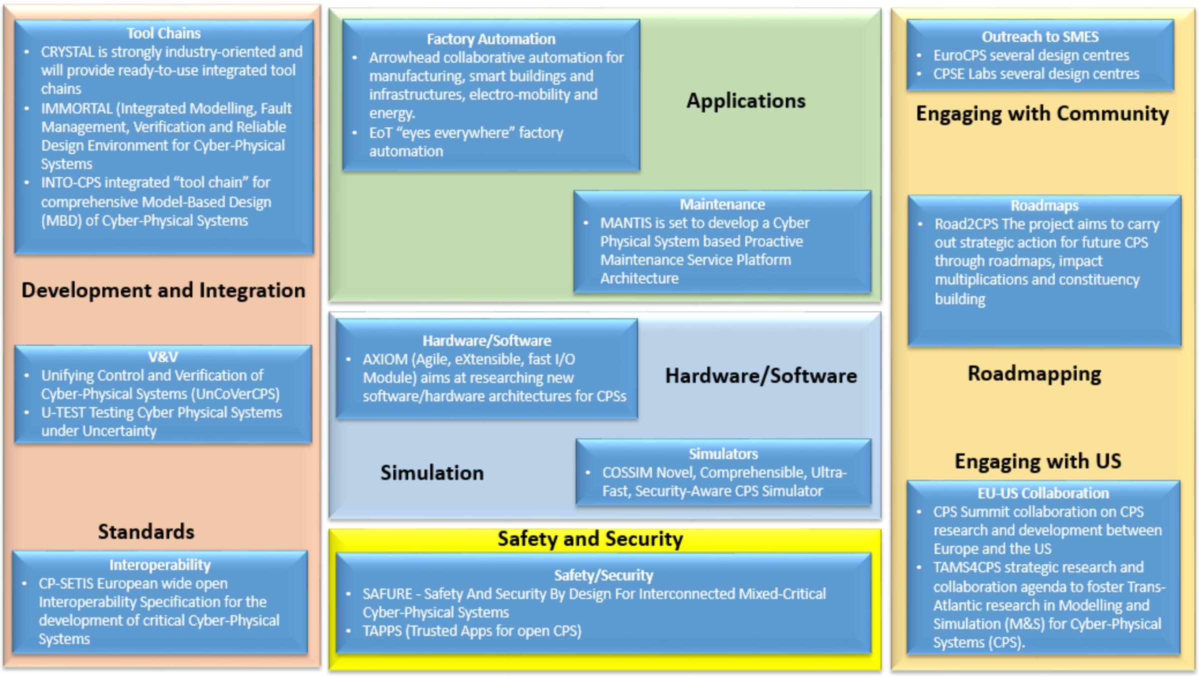


Figure 74. Overview of EU H2020 CPS Projects

In the CPS Cluster of projects the EC has funded 8 Research and Innovation Actions, 4 Innovation Actions and 3 Coordination and Support Actions under H2020 (See Figure 74) to provide coverage across a number of areas. The 8 Research and Innovation Actions are addressing areas such as co-simulation/modelling of all of system levels, including circuits, communication networks, firmware, operating system, system architecture and software layers with a view to providing model-based design. The aim is to reduce development cost and time by managing and reducing complexity. There are a broad range of use cases and the projects support more fundamental and longer term research issues than projects supported by ARTEMIS or ECSEL. The different projects and their coverage are briefly outlined below.

Research and Innovation Actions

- TAPPS: Trusted Apps for open CPS developing a platform for open CPS Apps with high security standards
- SAFURE: SAFety and secURity by design for interconnected mixed-critical Cyber-Physical Systems providing safety and security by construction for mixed–critical systems at design and run-time
- UnCoVerCPS: Unifying Control and Verification of Cyber-Physical Systems within the tool chain through modelling, verification, conformance testing and code generation
- U-TEST: Testing Cyber-Physical Systems under Uncertainty by developing systematic, extensible, and configurable model-based and search-based testing methodologies for building dependable CPS, and testing for uncertainty
- AXIOM: Agile, eXtensible, fast I/O Module to allow easy programmability of multi-core multiboard systems

- IMMORTAL: Integrated Modelling, Fault Management, Verification and Reliable Design Environment for Cyber-Physical Systems addressing reliable design and real time fault management in multi-core CPS
- INTO-CPS: INtegrated TOol chain for model-based design of CPSs providing an integrated tool chain for comprehensive model-based design of CPS
- COSSIM: A Novel, Comprehensible, Ultra-Fast, Security-Aware CPS Simulator providing an open-source framework to simulate the networking and processing parts of a CPS more accurately, more quickly and while considering security

Innovation Actions

In addition to EUROCPS and CPSELabs already described in the previous section two further projects are being funded:

- EOT: Eyes of Things building an ultra-low power and low cost vision platform for surveillance, augmented reality/wearable, cloud computing and perceptual computing
- CP-SETIS: Towards Cyber-Physical Systems Engineering Tools Interoperability Standardisation to produce an International Open Standard for development tools

Coordination and Support Actions

Three Coordination and Support Actions (CSAs) are being funded. One is developing strategic roadmaps for CPS and the other two are addressing EU-US collaboration.

- Road2CPS: Strategic action for future CPS through roadmaps, impact multiplication and constituency building
- TAMS4CPS: Trans-Atlantic Modelling and Simulation For Cyber-Physical Systems
- CPS-SUMMIT: Bringing experts together to identify research areas where the EU and US may work together, e.g. trustworthy systems

8.3.8 European Cluster of CPSoS Projects

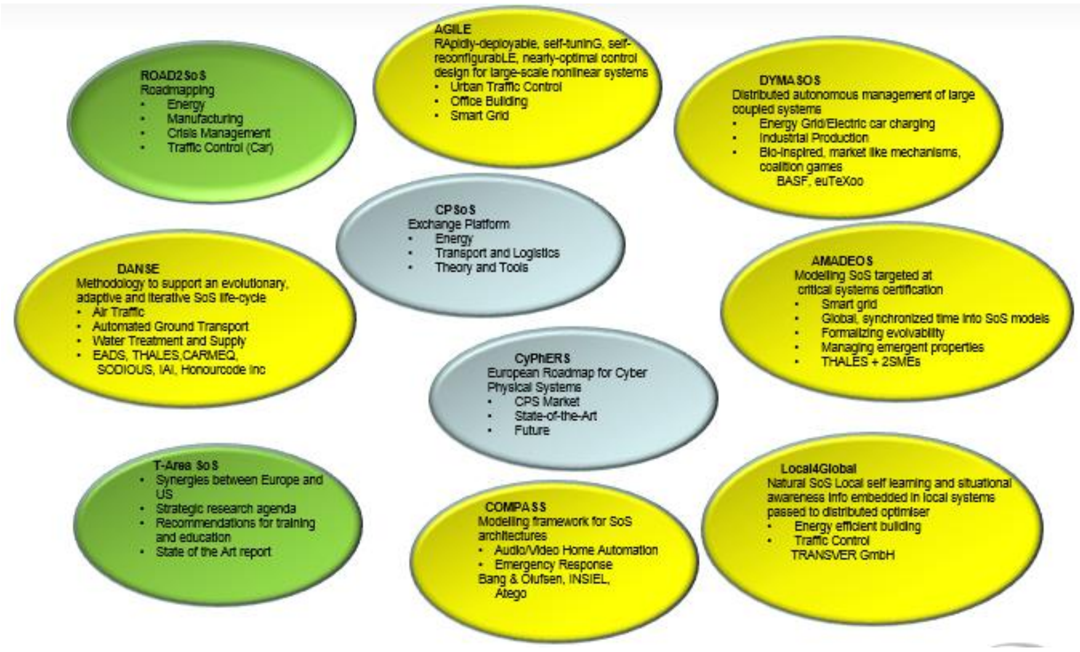


Figure 75. European Cluster of CPSoS Projects

At the Systems of Systems level the European Commission has also funded a number of Integrated Projects and Research and Innovation Projects to create a cluster of Cyber-Physical Systems of Systems projects (CPSoS). These are shown in Figure 75.

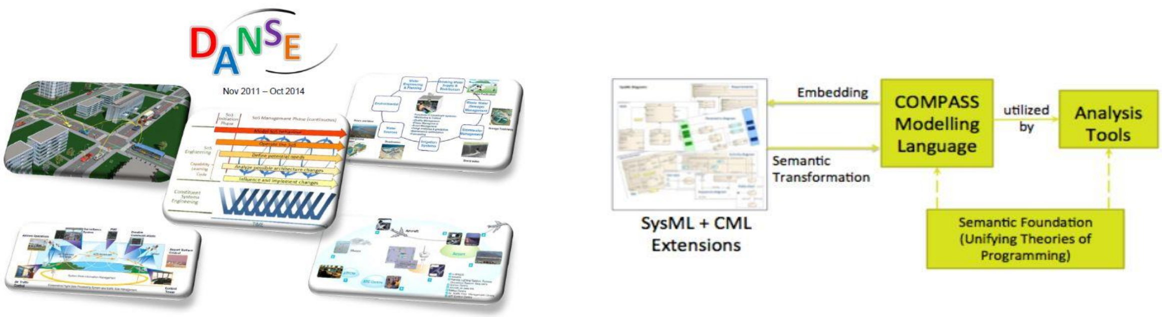


Figure 76. DANSE and COMPASS Integrated Projects

Two large Integrated projects (See Figure 76) were funded DANSE (Designing for Adaptation and evolutionN in System of systems Engineering) which addressed approaches for SoS engineering considering the need to deal with an evolving, adaptive and iterative life cycle, and COMPASS (Comprehensive Modelling for Advanced

System of Systems) which addressed model-based techniques for developing and maintaining Systems of Systems based on augmenting SysML with CML supported by proof and model checking tools. These projects addressed a diverse number of applications, including aerospace, autonomous vehicles, water management, emergency management and audio/video/home automation.

Other smaller projects within the cluster are:

- AMADEOS: Architecture for Multi-criticality Agile Dependable Evolutionary Open System-of-Systems
- AGILE: RApidly-deployable, self-tuninG, self-reconfIgurabLE, nearly-optimal control design for large-scale nonlinear systems for Urban Traffic Control (UTC) targeting a road network of Chania, Greece, and control of an Office Building
- LOCAL4GLOBAL: System-of-Systems that act LOCALly For optimising GLOBALly to create a plug-and-play control mechanism for the constituent systems of a SoS
- DYMASOS: Dynamic Management of Physically Coupled Systems of Systems addressing the management of large physically coupled systems of systems.

There have also been a number of roadmapping activities.

- CyPhERS: Cyber-Physical European Roadmap. The CyPhERS project developed a European strategic research and innovation agenda for Cyber-Physical Systems (CPS) identifying and prioritising research areas, support measures for both horizontal and vertical cooperation, research funding policies, training and standardization.
- ROAD2SOS: A Roadmap for Innovation in SoS. The project defined roadmaps for the domains of distributed energy generation and smart grids, integrated multi-site industrial production, emergency and crisis management, and multi-modal traffic control.
- T-AreaSoS: Trans-Atlantic Research and Education Agenda on Systems of Systems with the aim to create a commonly agreed EU-US Systems of Systems (SoS) research agenda.
- CPSoS: Cyber-Physical Systems of Systems. CPSoS provides a forum and an exchange platform for systems of systems related communities/projects and has also produced a strategic research and innovation agenda.

The CPSoS project is particularly relevant to the areas of smart cities, smart transportation and smart energy and led to the formation of the PICASSO project. The project analysed a number of key domains to formulate a view on the key challenges of CPSoS and to put forward recommendations for future research. The project highlighted that CPSoS are large, often spatially distributed physical systems with complex dynamics as shown in Figure 77. They are socio-technical systems with user/operator interaction but rely on distributed control, supervision and management. Partial autonomy of the subsystems is common to manage complexity of operation, and as the systems are in place for many decades, they dynamically evolve and reconfigure on different time-scales. Due to the complexity of interactions there is also a possibility of emerging behaviors.

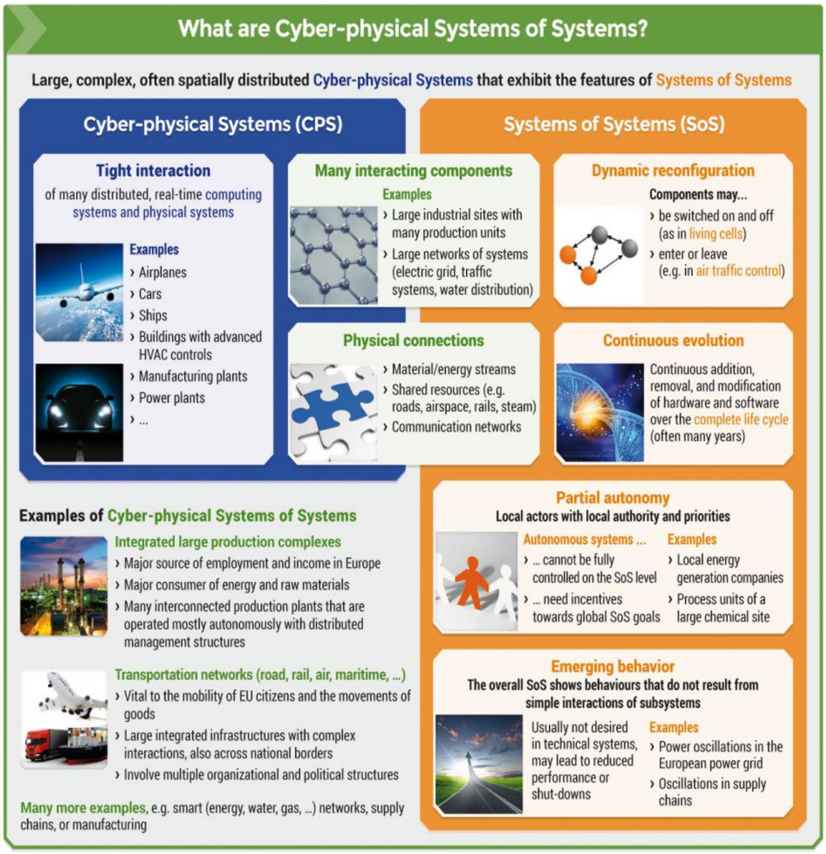


Figure 77. Cyber-Physical Systems of Systems

Three key research areas identified in CPSoS are the need for:

- Distributed, reliable and efficient management of Cyber-Physical Systems of Systems
- Cognitive Cyber-Physical Systems of Systems
- Engineering support for the design-operation continuum of Cyber-Physical Systems of Systems.

The project also identified needs for collaboration at a world-wide level, the needs for standards for interoperability, regulatory support and legal support to cover areas such as liability.

8.4 National CPS Initiatives

There are many national projects across Europe that address aspects of the development and use of CPS and IoT and it is not possible to cover them all in this report. However, a key notable initiative within the manufacturing sector that has gained considerable traction across Europe and is influencing developments at a European level, is Industry 4.0.

8.4.1 Germany

8.4.1.1 Industrie 4.0

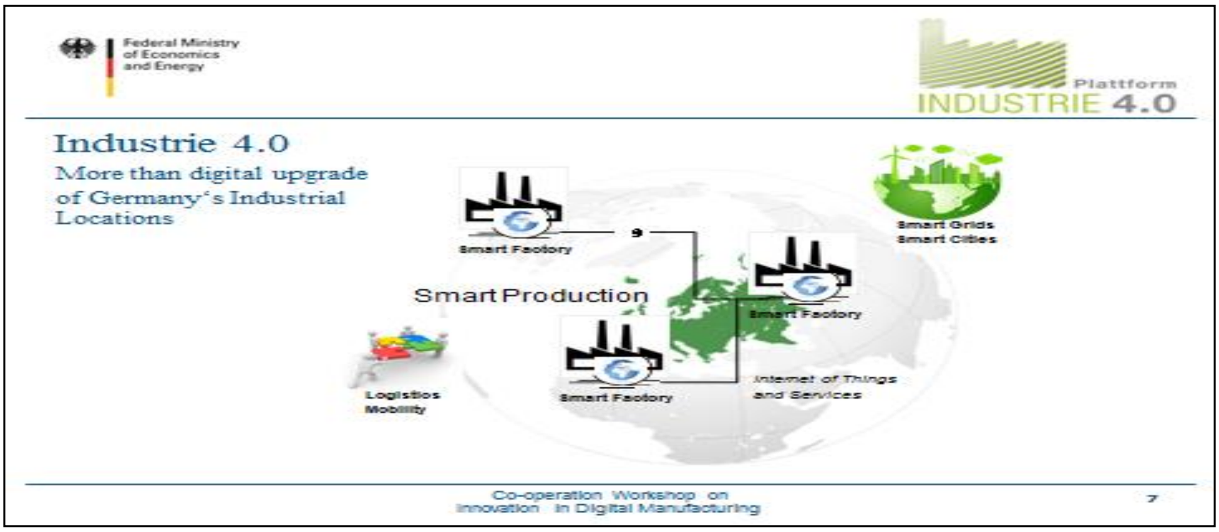


Figure 78. Industrie 4.0

A major initiative in Germany that is addressing the CPS area within manufacturing and automation is Industrie 4.0 [340] (See Figure 78). This is being supported under the German Government's HighTech Strategy 2020 which covers health, nutrition, communication, safety, climate, energy and mobility. There are 10 forward looking projects, two of which are in communications, with €200 million of funding from government. The working group identified 8 action areas: standardisation, reference architecture, mastery of complex systems, national broadband infrastructure, security, organisation and structure, vocational and further education, legal framework and resource efficiency. The aim is to achieve a quantum leap in organisational management through the entire value chain and product life-cycles. Key technologies identified are dynamic self-organising systems, real-time availability of information, Big Data management and optimisation. An aim is to realise research in the real world and BITKOM, VDMA and ZVEI are working together in the Platform Industrie 4.0. A scientific advisory board has been formed and 4 working groups on strategy, standardisation, research and security have been created. This initiative although German is also driving activities across Europe.

8.5 US activities

8.5.1 US IoT Drivers and Policy Initiatives

Developments in IoT in the US are being largely driven by companies with major players Google, Cisco, etc. dominating the marketplace. Various consortia and alliances have been formed to promote the uptake of IoT in the US. The vital importance of IoT has also been acknowledged by the Department of Commerce and it has been made a top priority as part of the Department's Digital Economy Agenda. An inquiry is being initiated by the National Telecommunications and Information Administration (NTIA) to review the current technological and policy landscape taking input from industry, researchers, academia, and society. The aim is to issue a green paper on key potential benefits and challenges of IoT, and identify possible roles for the federal government in fostering the advancement of IoT technologies in partnership with the private sector.

8.5.2 US IoT Initiatives

8.5.2.1 Industrial Internet Consortium

The US Industrial Internet Consortium (IIC) [341] although originally American is now attracting members from all over the world. It is a non-profit organisation with 14 staff (See Figure 79). Launched by AT&T, CISCO, GE, IBM and Intel, it is strongly tied to Object Management Group (OMG). There are currently 130 members (20 from the EU) and it is growing quickly. There are 20 working groups. The consortium has developed use cases in healthcare, transportation, manufacturing and smart grid and 3 approved testbeds have been developed.

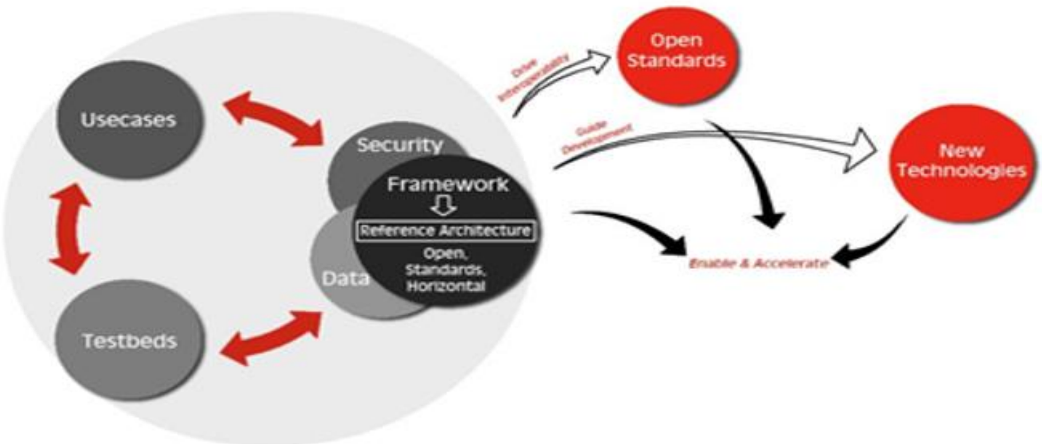


Figure 79. Industrial Internet Consortium

The initiative is addressing revenue generation, new operational efficiencies to drive down costs and improvements in customer satisfaction. This will create new markets and new working styles. Workforce productivity gains will be gained from digitalisation of tasks and reduced maintenance costs will result from use of predictive maintenance. Material and energy saving is also a key aim from reduced waste by precision monitoring to predict and control machines. There are working groups on security, technology, legal issues (e.g. who does data belong to), marketing and testbeds. Existing technology is used to identify research topics which are then investigated with the industrial members. It is not a standards organisation but it does evaluate and influence standards, e.g., ISO/IEC OMG and World Wide Web Consortium (W3C).

8.5.2.2 Allseen Alliance

Figure 80. AllSeen Alliance Structure

The Allseen Alliance [342] is a non-profit consortium which is dedicated to providing an open environment for the Internet of Things to allow the widespread adoption of billions of products, systems and services (See Figure 80). The aim is to create a vibrant ecosystem and thriving technical community of hardware manufacturers and software developers who can create interoperable products that can discover, connect, communicate and interact directly with other devices, systems and services regardless of brand. This has been developed as part of the collaborative AllJoyn open source project, which provides an industry-supported software and service framework for smart connected products. The AllSeen Alliance organization is open and inclusive.

8.5.2.3 Open Interconnect Foundation

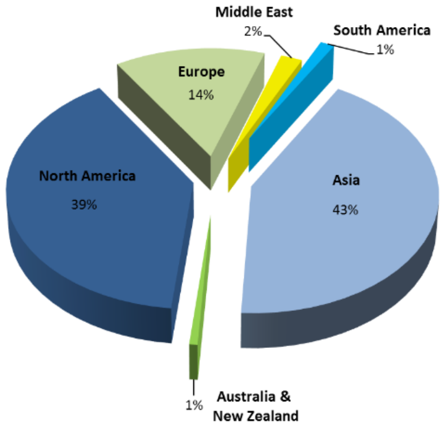


Figure 81. Open Interconnect Consortium Membership by Country

The Open Interconnect Foundation [343] (See Figure 81) was founded by major companies working on IoT chips, software, platforms and products for the Internet of Things to work together towards a single standard for IoT. It includes Intel, Microsoft, Samsung, Qualcomm, GE Digital and Cisco Systems. The aim is to come up with a single specification, or at least an open source common set of protocols and projects, for wearables, home appliances, industrial equipment, etc. and will provide OCF-certified products. This should allow billions of connected devices (devices, phones, computers and sensors) to communicate with one another regardless of manufacturer, operating system, chipset or physical transport. The aim is to accelerate industry innovation and help developers and companies create solutions that map to a single open specification providing interoperability for consumers, business, and industry. The OCF brings together members of the two major rival organizations: the Open Interconnect Consortium (OIC) and the AllSeen Alliance. Both groups have been promoting their own ways for connected devices to discover each other and determine what they can do together. All OIC members, including Intel and Samsung, are now part of OCF and two key members of AllSeen, Qualcomm and Microsoft, are now in the OCF. Microsoft has already stated that all Windows devices will in future interoperate with the OCF standard. Cisco and GE Digital have also joined OCF along with CableLabs, home appliance maker Electrolux and video and broadband company Arris Group. The OCF is promoting the IoTivity open source project [344].

8.6 US CPS Drivers and Policy Initiatives

8.6.1 NSF Cyber-Physical Systems Programme

The NSF Cyber-Physical Systems programme [345] has so far funded over 300 projects. The goal of the CPS programme is to develop the core system science needed to engineer complex Cyber-Physical Systems. The programme aims to create a research community committed to advancing research and education in CPS and to transitioning CPS science and technology into engineering practice. The programme addresses in particular cross-cutting fundamental scientific and engineering principles that underpin the integration of cyber and physical elements across all application sectors. Additionally, the programme supports the development of methods, tools, hardware and software components, prototypes and testbeds.

In the US there is interest in addressing basic CPS research that is required across multiple application domains. NSF is thus working closely with multiple agencies of the federal government within the US, including the U.S. Department of Homeland Security (DHS) Science and Technology Directorate (S&T); the U.S. Department of Transportation (DOT) Federal Highway Administration (FHWA), and through FHWA, the U.S. DOT Intelligent Transportation Systems (ITS) Joint Program Office (JPO); the National Aeronautics and Space Administration (NASA) Aeronautics Research Mission Directorate (ARMD); several National Institutes of Health (NIH) institutes and centres [including the National Institute of Biomedical Imaging and Bioengineering (NIBIB), Office of Behavioral and Social Sciences Research (OBSSR), National Cancer Institute (NCI), and National Center for Advancing Translational Sciences (NCATS)]; and the U.S. Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA). This highlights the wide breadth of areas where CPS has a role. There are three classes of research and education projects that are funded by NSF:

- **Breakthrough** projects giving a significant advance in fundamental CPS science, engineering and/or technology that has the potential to change the field. This category focuses on new approaches to bridge computing, communication, and control. Funding for Breakthrough projects may be requested for a total of up to $500,000 for a period of up to 3 years.
- **Synergy** projects to demonstrate innovation at the intersection of multiple disciplines, to accomplish a clear goal that requires an integrated perspective spanning the disciplines. Funding for Synergy projects may be requested for a total of $500,001 to $1 million for a period of 3 to 4 years.
- **Frontier** projects that address clearly identified critical CPS challenges that cannot be achieved by a set of smaller projects. Funding may be requested for a total of $1 million to $7 million for a period of 4 to 5 years.

8.6.2 NITRD - The Networking and Information Technology Research and Development Programme

The NITRD CPS SSG [346] is coordinating programmes, budgets, and policy recommendations for Cyber-Physical Systems (CPS) research and development (R&D). This includes identifying and integrating requirements, conducting joint programme planning, and developing joint strategies for the CPS R&D programmes conducted by agency members of the NITRD Subcommittee (Federal IoT/Cyber-Physical Systems) The Cyber-Physical Systems Vision Statement document produced by NITRD provides a summary of the drivers, e.g. building controls, energy, transportation, healthcare and crosscutting strategic challenges such as cyber-security, privacy and social-technical aspects of CPS. The report also lists the technologies being explored and the multi-agency plans for CPS. Critically the report identified that although a number of federal agencies have independent research efforts to address CPS science and engineering challenges, there are still many gaps in the federal R&D portfolio.

R&D Gaps	DARPA	DHS	DOD/services	DOE	DOE/ARPA-E	DOT	FDA	NIH	NASA	NIFA	NSA	NIST	NSF	OASD (R&E)	Others
Mission R&D: Crosscutting Research and Development															
Core CPS Science and Technology – control, real-time computing, communication concepts, modeling, hardware, and software platforms. Advanced engineered systems: manufacturing, energy, medical devices, transportation		X							X			X	X		
Science of Security for CPS				X							X	X	X		
CPS Virtual Organization (CPS-VO)		X				X					X		X		X
Complex systems, cascading failure, engineered resilient systems, fault identification, diagnosis and recovery		X	X						X			X		X	
Mission R&D: Sector-Specific Challenges															
Aviation safety, certification, enabling bold, visionary aviation systems and technology for a safe, efficient Next Generation airspace									X						
Intelligent transportation infrastructure systems, enabling technology for high confidence next-generation transportation (NextGen, automotive autonomy, intelligent vehicles)		X	X			X			X			X	X		
New control architectures/algorithms and power electronics for distributed generation, storage, and managed consumption			X		X				X			X			
Smart food systems that support safety, logistic efficiencies, cold-chain integrity, and traceability										X					
Time-critical systems, mixed criticality architectures, verification, aviation autonomy		X	X						X						
Rapid design and manufacturing of advanced CPS technologies. Rapid verification and real-time health monitoring and reconfiguration/re-verification. Application to autonomous systems	X		X						X						
Real-time physiological sensing, modeling, control, and feedback; advanced medical devices and system interoperability, integration, and certification							X	X	X			X			
Mission R&D: Crosscutting Standards-Based Platform Technologies															
Cyber-Physical Systems Engineering Testbed		X										X			
Measurement Science and Standards for Model-Based Diagnostics & Prognostics, Time Synchronization, Industrial Cybersecurity												X			
Measurement Science and Standards for Quality Measurement Systems for CPS, Wireless Networking for CPS, Advanced Battery Technology for CPS, Multi-Physics Modeling and Optimization, Adaptive and Predictive Control in CPS												X			
Standards-Based Integrated Architectures and Prototype Platform for CPS		X							X			X			
Education and Crosscutting Research Centers															
Future of skills development and instructor resources for CPS including online CPS training and educational infrastructure resources (e.g., CPS virtual laboratory)		X							X				X		
Research and Infrastructure for innovation in Medical CPS Pilot							X	X				X	X		
CPS Outreach Centers (CPS Government, Industry, Academia cooperative research model)													X		X
Industrial Internet Consortium		X										X			
Transportation CPS Pilot (NSF Engineering Research Centers model)						X			X				X		X

Figure 82. Technical Gaps Identified by NITRD

There are many technical barriers to rapid, predictable development and deployment of CPS as shown in Figure 82. It is not possible to address the gaps agency by agency, sector by sector, or company by company so a multi-agency, multi-sector comprehensive focus on crosscutting R&D challenges in CPS is advocated. This should allow developments for unmanned aerial vehicles or self-driving cars to be applied in other domains. Likewise technology developed for secure reprogrammable networked medical devices could be used in the smart grid. Thus collaboration between industry, university and government contributors in private-public partnerships is being promoted and a number of funding schemes have been proposed.

8.6.3 NIST CPS Initiative

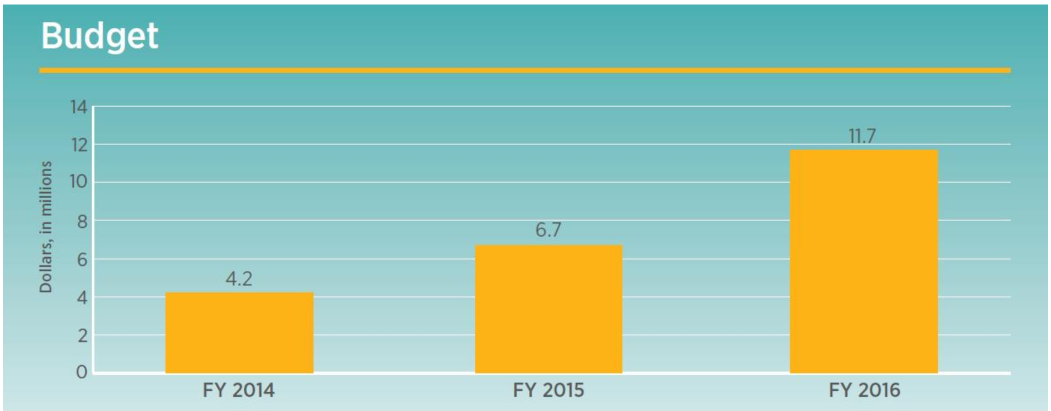


Figure 83. NIST Budget for CPS

The NIST Engineering Laboratory, through the Cyber-Physical Systems and Smart Grid Program Office, is leading NISTs activities on Cyber-Physical Systems. Notably the budget in this area for NIST is increasing as shown in Figure 83 highlighting the importance of the area.

In 2014 NIST formed the Cyber-Physical Systems Public Working Group (CPS PWG). This brings together experts to help define and shape key aspects of CPS to accelerate its development and implementation within multiple sectors. The CPS PWG has prepared a draft CPS Framework as shown in Figure 84.

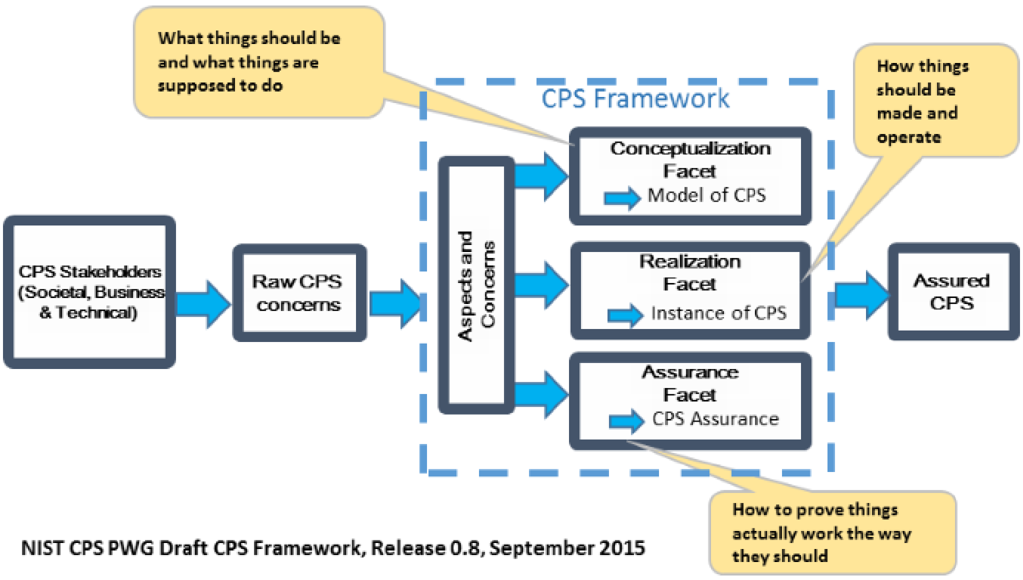


Figure 84. Draft NIST Framework for CPS

The CPS Framework [347] was developed in partnership with industry, academic and government experts and provides a methodology for understanding, designing and building CPS including those with multiple applications. As CPS are multidisciplinary CPS research and standards development is carried out in multiple NIST Laboratories. This includes advanced manufacturing, cyber-security, buildings and structures, disaster resilience, and smart grid. A key goal is to design and develop a CPS testbed to characterise CPS equipment, systems, performance, and standards.

NIST has also investigated the strategic importance of CPS and produced a report [348] "Strategic R&D opportunities for 21st Century Cyber-Physical Systems" which highlights the importance of the technology across a number of sectors including autonomous cars, robotic surgery, intelligent buildings, smart electric grid, smart manufacturing, and implanted medical devices. Key needs identified were:

- Robust, effective design and construction of systems and infrastructure - key to designing dependable systems from the ground up and reducing cost and time to market;
- Improved performance and quality assurance - essential for spurring future investment, acceptance, and use of innovative systems that promise to provide revolutionary improvements to conventional practice;
- Effective and reliable system integration and interoperability - required for highly connected and networked components to work together effectively as a total system; and
- Dynamic, multi-disciplinary education and training - will make possible sustained growth and innovation and spawn a new generation of entrepreneurs, as well as the next generation of cyber-physical systems.

8.6.4 US IoT Global City Teams Challenge

Figure 85. US IoT Global City Teams Challenge

NIST [349] lead the Global City Teams Challenge (See Figure 85) to help communities around the world work together to address issues ranging from air quality to traffic management to emergency services coordination. Under the challenge communities and innovators are brought together to create teams addressing smart city issues utilising networked technologies to better manage resources and improve quality of life in nine-month projects.

This challenge follows on from the SmartAmerica Challenge, which ran from December 2013 to June 2014. This challenge successfully brought together more than 100 companies, universities and other organizations to form teams that developed and applied networked technologies. In the first round of the Global Cities Challenge more than 40 teams or "action clusters" performed projects related to energy, transportation, public safety, and other key sectors. Partners in the Global Cities Challenge include IBM, AT&T, Intel, NSF, ITA, DoT, State Department, GSA, FIWARE, World e-Governments (WeGO), IIC, The Kingdom of Netherlands and the Republic of South Korea. Participating Organizations include Qualcomm, Bosch, Siemens, CH2M Hill, Mathworks, Pecan Street, Inc., Yeti Analytics, MIT, Vanderbilt, UT Dallas, University of North Texas, Ohio State University and Columbia University.

8.7 Rest of the World

8.7.1 Worldwide Alliances in IoT

Figure 86. Worldwide Alliances in IoT

In the IoT domain there are many world-wide alliances as shown in Figure 86. Many companies are promoting their ideas for the Internet of Things. One example is Intel who have developed an Intel IoT platform [350]. This is an end-to-end architecture based on secure horizontal and interoperable building blocks that functions as an IoT platform that can be deployed across industry sectors, e.g. manufacturing, utilities, healthcare, and public safety, and smart cities.

8.7.2 South Korea and EU

The South Korean Government is investing $350 million over the next 4 years in 300 companies it thinks can compete globally in the IoT ecosystem. There is also a Joint Research initiative with Europe under H2020 to create IoT ecosystems with open and interoperable platforms. This is addressing harmonised IoT architectures and reference implementations, security and privacy mechanisms for devices, architectures, service and network platforms, including characteristics such as openness, dynamic expandability, interoperability, dependability, cognitive capabilities and distributed decision making, cost and energy efficiency, ergonomic and user friendliness. A key aim is to develop joint IoT infrastructure reference implementation models and IoT standardisation with pilot projects in smart cities, health care, smart factories and smart logistics with cross regional demonstrations.

8.8 ICT Regulations

With the explosion of interest in IoT it is increasingly apparent that policy and regulation have a crucial role. IoT deployments at scale have implications for privacy and security, business models and market access, standards and interoperability, and national and supranational priorities. Here a harmonization between the US and Europe is not only advantageous but strongly needed. Engineering trustable, reliable, evolvable and affordable cyber-physical systems connected by the Internet of Things is a scientific and technological challenge that requires huge efforts. Joining forces will help to advance more quickly, meet societal challenges and help both US and European companies compete in world markets.

8.8.1 CPS and IoT Security

Security is a key issue in both CPS and IoT systems. In the US this is being addressed by NIST's Cyber-security and Privacy Subgroup [351]. Here cyber-security and privacy considerations are being considered in all phases of the system development lifecycle, from design through implementation, maintenance, and disposition. The primary aim is to develop a cyber-security and privacy strategy for the common elements of CPS that includes the identification, protection, detection, response and recovery of CPS elements. The subgroup is identifying areas for further CPS cyber-security research and development.

Supporting this the NITRD's Cyber Security and Information Assurance (CSIA) Interagency Working Group is performing research and development to prevent, resist, detect, respond to, and/or recover from actions that compromise or threaten to compromise the availability, integrity, or confidentiality of computer- and network-based systems. This includes applications in power grids, emergency communications systems, financial systems, air-traffic-control networks, national defence, national and homeland security. Research areas include Internet and network security; confidentiality, availability, and integrity of information and computer-based systems; new approaches to achieving hardware and software security; testing and assessment of computer-based systems security; and reconstitution and recovery of computer-based systems and data [352].

Cyber-security has become an increasingly important topic and the President announced a Cyber-security National Action Plan (CNAP) and the Administration released the 2016 Federal Cyber-security Research and Development Strategic Plan [353] which is an update of the Trustworthy Cyberspace: Strategic Plan for the Federal Cyber-security Research and Development Program which was launched in 2011. The plan is aimed at providing methods and tools for deterring, protecting, detecting, and adapting to malicious cyber activities [354].

In 2014, NIST published the Framework for Improving Critical Infrastructure Cyber-security. The NIST Cyber-security Framework Core defines five functions (Identify, Protect, Detect, Respond, Recover). The new plan defines four elements (Deter, Protect, Detect, and Adapt).

Figure 87. IoT Security

Within Europe the area of cyber-security is more fragmented. The Internet of Things Security Foundation (IoTSF) is a group set up by the UK's National Measurement Institute. This is trying to put together guidelines for developers to improve overall security [355]. Elsewhere in Europe there are a number of other national initiatives. Urgent action is needed, however, as the applications being developed and put into service (See Figure 87) are increasingly vulnerable to attack.

9 Key Points Identified, Analysis and Opportunities

9.1 Smart Cities

Within the smart cities domain the report highlights the wide range of areas where "smartness" can be exploited including government, economic and financial systems, building management, manufacturing, education, community and social services, healthcare, transportation, utilities and infrastructure and communications. In Europe a number of initiatives, e.g. the European Initiative on Smart Cities and the European Innovation Partnership for Smart Cities and Communities is investing in sustainable development in as many cities as possible. Key to this is partnership and the aim is to create equal partnerships between cities and companies. The EC is supporting this by creating "Lighthouse Projects" with the intention of these signalling numerous follow-up projects across Europe. Smaller cities are also being aided by the Small Giants Initiative. To support development there is a need for "at scale" experimental research facilities such as SmartSantander and partnerships between companies, governments and knowledge institutions such as Amsterdam Smart City (ASC) (100 partners). Notably there are different levels of "smartness" and there are a large number of cities in Europe (468) which display some attributes of smart cities.

In the US the White House Smart Cities Initiative is investing over $160 million in federal research. NITRD has set up a framework to coordinate Federal Smart Cities activities, agency investments and outside collaborations. A number of initiatives are being supported by NIST including the Global City Teams Challenge. Underlying this is foundational research supported by NSF that supports design and management of Smart and Connected Communities. There are a number of notable smart cities in the US including Boston, New York, Seattle, San Francisco, Washington, San Jose and Chicago.

The smart cities industry is valued at more than $400 billion globally by 2020. Providing technology and services for Smart Cities around the world is thus a major business opportunity and this is a major driver for some national programmes, e.g. in the UK, Japan, South Korea and China. There are many deployments around the rest of the world where best practice could be gleaned.

There are a number of barriers to deployment. There is a critical need for regulation in the area of privacy and in allowing sharing of data to provide services. Smart cities will constantly register and process private data from individuals leading to questions about gathering data with consent and reconciling the value of services with privacy. Anonymising data, encryption and processing in encrypted domains may all be needed. The very wide scope of smart cities, which covers not only interactions with citizens and use of their data, but also control of the energy, waste, transportation systems and social interactions with government, education and e-health, leads to many areas were regulation may be required. Safety is of underlying importance to citizens.

There are several developing standards for smart cities notably in the UK covering a variety of smart city topics, ETSI smart city standards targeted at mobility, transportation, M2M, energy efficiency, security, and ITU-T on Smart Sustainable Cities. China is also very active in standardisation with a number of Chinese bodies.

A growing problem is susceptibility to cyber-attack and here lessons can be learned from countries such as Estonia. In the US NIST is working with the National Cyber-security Center of Excellence (NCCOE) to provide cyber-security solutions based on commercially available technologies for smart city applications.

9.2 Smart Energy and Smart Grid

Energy and Smart Grids are key topics in both Europe and the US driven by national and European green initiatives, e.g. in Europe to reduce greenhouse gas emissions by 40%. Urban areas consume 70% of energy, and account for 75% of the EU's greenhouse gas emissions. Thus buildings, transportation systems, water supply and treatment, and sewage management are an area where most energy savings could be made. Notably nearly 50% of European Smart City initiatives address environmental problems. The European Strategic Energy Technology Plan (SET Plan) has identified the need for development of energy technologies to combat climate change and the need for securing energy supply at the European and global level. To support this the European Energy Research Alliance (EERA) has been set up by leading European research institutes and the SETIS Initiative has been created to support cities and regions in sustainable use and production of energy. An ICT Roadmap for Energy Efficient Neighbourhoods has been created and KIC InnoEnergy has been set up as a commercial company dedicated to promoting innovation, entrepreneurship and education in the sustainable energy field.

At a national level policies are also driving change, e.g. in Germany to phase out nuclear power and in the UK cutting greenhouse gas emissions by 80% by 2050. The low carbon economy strategy in the UK is driving the setup of public private research partnerships and a range of initiatives for electivity and heating. Every home will be supplied with a smart meter helping consumers to understand their energy consumption and make savings.

In the US as part of the Reinvestment and Recovery Act there are a number of government initiatives and policies including totalling $3.4 billion of investment grants for Smart Grid projects. This includes funding to promote energy-saving choices for consumers, increasing efficiency, and fostering the growth of renewable energy sources such as wind and solar. The Energy Independence and Security Act of 2007 (EISA) also made it policy to modernise the nation's electricity transmission and distribution system to create a smart electric grid. This is supported by the Administration's commitment in the "Blueprint for a Secure Energy Future" and "A Policy Framework for the 21st Century Grid: Enabling Our Secure Energy Future."

Around the world different approaches are being adopted and a wide variety of technologies and services are being demonstrated driven by national and regional business drivers. In the US peak load reduction technology and dynamic pricing tariffs are being pursued. In Europe emphasis is on improving energy efficiency and reducing emissions through the decentralised production. In the Asia-Pacific region China is modernising and improving grid reliability and Australia and New Zealand are exploring new techniques for load management. There are major investments in China ($128 billion) to reduce carbon intensity by investment in renewable power and to create grid interconnectivity with neighbouring countries such as Russia, Mongolia, Kazakhstan, Pakistan, Myanmar, Laos, Nepal and Thailand. Since 1992 China has relied heavily on electricity it purchases from Russia. Other countries also actively pursuing smart grids are Brazil, Mexico South Korea and Japan.

The smart grid market is led by regulation and reductions in emissions, consumer choice and energy security are driving adoption of smart grid technologies. Regulation in the UK and Germany is introducing smart metering and tariffs. A number of pricing structures are being explored: tiered pricing rates that reflect system capacity and time-of-use pricing (off-peak/on-peak) schemes. Introducing "Critical Peak" prices has been found in a US pilot to be the most effective technique to trigger load reduction.

Across Europe grid regulation varies considerably making smart grid investments difficult. In Europe the Electricity Directive and the Energy Services Directive provide a mix of obligations and incentives to Member States to establish a common regulatory framework. Other bodies also need to be involved such as the Agency for the Cooperation of Energy Regulators (ACER) that fosters cooperation among European energy regulators to ensure market integration and the harmonisation of regulatory frameworks within the framework of the EU's energy policy objectives and the Council of European Energy Regulators (CEER) that represent national regulators.

In the smart grid area there is a need for standards for interoperability and safety. Standards are voluntary in Europe and are developed by industry and market actors. The European Commission and EFTA have issued the Smart Grid Mandate M/490 which was accepted by CEN, CENELEC and ETSI. In the US EISA asked NIST and FERC to facilitate the development and adoption of interoperability standards. NIST is leading this coordinating the development of a framework that includes protocols and model standards for information management to achieve interoperability of Smart Grid devices and systems.

Security is another key concern and smarter grids lead to increased vulnerabilities from intrusions, error-caused disruptions, malicious attacks, destruction, and other threats. As the electric grid network is key to the operation of a country, cyber-security is a key topic on both sides of the Atlantic. The European Commission has put together a multi-stakeholder and multidisciplinary group of experts to discuss and work on relevant matters regarding the security and resilience of communication networks and information systems for Smart Grids across Europe. Although standards for smart grid cyber-security are already available these need to be maintained and enhanced as it and technology evolves. In Europe Alstom Grid, Intel, and McAfee produced a white paper on smart grid cyber-security. In the US the Administration has proposed specific cyber-security legislation to ensure that grid operators and all stakeholders have access to actionable threat information and provide support for research, development, and demonstration of cyber-security systems. The aim is to identify and prioritise relevant cyber risks - including malware, compromised devices, insider threats, and hijacked systems - and develop standards and guidelines that enable the design of effective plans for mitigating those risks. A number of threat warning bodies have been set up in the US, Electricity Sector - Information Sharing and Analysis Center, the United States Computer Emergency Readiness Team, and the National Electric Sector Cyber-security Organization. The NIST Information Technology Laboratory (ITL), Computer Security Division leads the Smart Grid Interoperability Panel (SGIP) Cyber-security Committee (SGCC) which has produced the NISTIR 7628 Guidelines for Cyber-Security (Volumes 1, 2, and 3) which is widely used by utilities, vendors, and regulators in the US.

9.3 Smart Transportation

North America and Europe are expected to be the largest markets for Intelligent Traffic Systems. Within Europe sustainability is a key issue with a dramatic increase in both freight and passenger transport and associated emissions. There is also an aim to halve casualties with respect to 2001 levels in road transport. The European Commission supports a number of transport Technology Platforms that includes ERRAC (Rail), ERTRAC (Road), ACARE (Aerospace) and WATERBORNE (Marine) that sets the research agenda for each domain. At a national level there are several programmes implementing Intelligent Transport Systems for road users. To coordinate the planning of infrastructure projects across Europe the Trans-European transport networks (TEN-T) policy has an investment programme of EUR 400 billion. The European Commission's aim is to accelerate and coordinate the deployment of ITS applications on the Trans European Road Network (TERN) across Member States in a consistent and harmonised way.

The drivers in the US for intelligent transportation are similar to those in Europe, however, another key driver is homeland security. There is a desire to provide surveillance of roadways and also a means for mass evacuation of people in urban areas as a result of natural disaster or threat. In the US the ITS Joint Program Office (ITS JPO), coordinates across the Federal Highway Administration, Federal Motor Carrier Safety Administration, Federal Railroad Administration, Federal Transit Administration, Maritime Administration, and the National Highway Traffic Safety Administration to plan, programme, and execute the ITS Research Program. This is being supported by a guidance handbook from the SSTI and Smart Growth America and stimulation challenges, e.g. the Smart City Challenge. Looking longer term Beyond Traffic is looking at trends and needs over the next three decades.

9.3.1 Road

The ERTRAC Strategic Research Agenda covers mobility, transport and infrastructure, safety and security, environment, energy and resources, design and production. It highlights a number of key research topics including traffic management, integration of vehicle and infrastructure systems, traffic management using ITS, data collection and processing, business models, optimisation of road space to ensure that vehicles (particularly HGVs) adopt routing systems that minimise adverse impacts, systems for segregating traffic with dedicated infrastructure and prioritised traffic management and methods to assist the booking of optimised slots for freight vehicles. Multi-modal door-to-door mobility is also being considered in the New Mobility Services project. Autonomous driving features to improve safety are being pursued in both Europe and the US with projects like HAVEit and the Swedish Drive Me project and the Google car in the US.

The adoption of electric vehicles and the decarbonisation of transport is a priority in both Europe and the US. In Europe there are a number of “green initiatives” at European level to promote the take up the use of electric vehicles and to reduce pollution across a range of transport modes such as the European Green Cars Initiative and the Smart, Green and Integrated Transport programme and EV4SCC with €6339m funding. In the US the Recovery Act is providing large investments in advanced vehicle and fuel technologies, public transit, and high speed rail and tax credits are being used to encourage uptake of electric cars. Urbanisation is a key challenge and air quality directives such as Euro 6 are driving new truck and powerplant design in Europe. Likewise stricter standards are being introduced in the USA to raise average fuel economy to 35.5 miles per gallon for cars and trucks by 2016. Within trucking there are particular problems for Europe such as a myriad of automatic tolling systems which would benefit from harmonisation but on both sides of the Atlantic telematics is being exploited to reduce carbon footprint and optimise deliveries by major logistics companies such as DHL and UPS.

Increased connections to cars is being driven by regulation in the EU for the introduction of car emergency vehicle notification systems (eCall). Standards for car-to-car and car-to-infrastructure communication need to be global to allow automotive companies to sell their products around the world. The ITS market is global and there are many opportunities in the Asia Pacific region and India. In the EU the CAR 2 CAR Communication Consortium has been developing and testing standards driven by the EC M/453 mandate for European standardisation organisations ETSI, TC, ITS, and CEN to produce a minimum set of standards that ensure interoperability. Likewise in the US there is a focus on vehicle-to-vehicle and vehicle-to-infrastructure connectivity through the application of advanced wireless technologies. The ITS Research Program is developing and testing the underlying technology and applications. The ISO/TC 204 standard particularly addresses intelligent transport systems, with a focus on standardization of information, communication and control systems for urban and rural surface transportation.

9.3.2 Rail

The interoperability regulations and the 2011 Transport White Paper require that the European railway system behaves as a single network. The commercial drivers in the industry are for 24/7 operation, high availability, low cost, safety, increased capacity, recovery from disturbance, low carbon emissions and customer satisfaction. In Europe the Strategy for European Rail Research – Vision 2020, the Strategic Rail Research Agenda, and “Railroute 2050” vision highlights the need for the European Railway Traffic Management System (ERTMS)to replace the existing 20 train control systems utilised across the European Union. Key initiatives are the H2020 supported FOSTER-RAIL and SHIFT^{2}RAIL joint technology initiatives to focus research and innovation (R&I). In the US there is a major need to modernise the rail system and although there has been investment of $11 billion on High Speed Rail initiatives this has been largely spent on upgrading the existing Amtrak service which is limited to 110 miles per hour. Although a further $10 billion has been requested by Congress to support high-speed initiatives there is considerable opposition from republican governors and community opposition for projects that are seen as too expensive and unnecessary. Some high speed rail projects are going ahead, e.g. the controversial Los Angeles - San Francisco route, and privately funded initiatives from All Aboard Florida with a $1.5 billion loan from the Federal Railroad Administration and the Texas Central Railway company, which plans to introduce Japanese bullet trains between Houston and Dallas.

9.3.3 Air Transportation

Air passenger volume is predicted to double air traffic density over the next two decades in an already congested airspace. The Single European Sky ATM research programme SESAR is reforming the architecture of European Air Traffic Control to meet future capacity and safety needs. SESAR aims at developing the new generation air traffic management system capable of ensuring the safety and fluidity of air transport worldwide over the next 30 years. The equivalent of SESAR in the US is NEXTGEN. A concern here from a global perspective is that both systems adopt fundamentally different approaches to air traffic management. Autonomous Unmanned Vehicles are very active area on both sides of the Atlantic with many military programmes. Considering civilian programmes there is only one major national programme in the UK ASTRAEA. In the US there is some activity from Amazon on developing drones for delivering parcels which raises some certification issues with the FAA.

9.3.4 Marine

A key driver in the maritime industry is improving safety of waterborne operations as accidents come with high costs in terms of loss of life, environmental damage and with high economic impact. Traffic management is key for safer and more secure operations. The technology can also be used for optimised shipping operations and voyage optimisation, condition based maintenance, reducing costs and reducing emissions which is driven by strict legislation in Europe at a national level and also at local level in ports. In Europe the WATERBORNE European Technology Platform has defined the Marine Vision 2020 and Strategic Research Agenda which drives funding for projects. There is a major e-Maritime initiative to exploit advanced information technologies within the maritime sector. Unmanned navigation and autonomous ships are also being researched but there are considerable hurdles to adoption coming from regulators concerned about safety and unions who are concerned about job losses. Current regulations dictate minimum crew levels by international conventions.

In the US the Maritime Administration of the US Department of Transportation has highlighted that policy reforms are needed to address international shipping trade. Offices have been created at major US gateway ports to interact with key stakeholders to identify Federal and state funding and cooperate on projects. Public private partnerships are being used to identify bottlenecks and ways of improving freight movement, and to fund redevelopment of port infrastructure, e.g. berths, piers, container cranes, on-site rail and railroad trailers. This is being strongly supported by the Transportation Infrastructure Finance and Innovation Act (TIFIA) programme that provides direct loans, loan guarantees and credit with $1.435 billion in capital over five years. A strong domestic maritime industry is seen as being critical for America's economic, national, and homeland security. The maritime industry is strongly represented by the American Maritime Partnership (AMP) with 450 members and the Jones Act requires that any vessel transporting goods or passengers between two points in the United States or engaging in activities in US waters must be US owned, US built, and US crewed.

9.4 5G

5G extends the cellular network from content delivery to a 'Control Network' that allows new control applications. The EC has committed €50 million for research to deliver 5G mobile technology by 2020. This includes initiatives like the METIS project and the 5G Infrastructure Public Private Partnership to bring together several telecommunications companies. A portfolio of first phase projects are defining key functionalities for 5G. Notably the work on 5G within Europe and the US has largely based at a few key Universities, e.g. 5G Innovation Centre (University of Surrey), Kings College, TU Dresden, Chalmers and Lund, largely in partnership with companies such as Ericsson and Vodafone. A similar situation exists in the US where 5G research has been mainly being funded by NSF at institutions such as University of Texas at Austin, Stamford, Berkeley, New York University, and Rutgers. However, with the announcement in July 2016 of the $400 million Advanced Wireless Research Initiative in the US there will now be a significant boost in research for 5G with the establishment of four city-sized testing grounds for 5G wireless services from October 2017. This is also being supported by a number of leading American telecommunications companies.

There is a key need for spectrum harmonisation for 5G so that the same frequencies are used worldwide and companies who operate on a world-wide basis can produce appropriate equipment. Within the US the MOBILE NOW Act has been introduced to boost the development of next-generation 5G wireless by ensuring more spectrum is made available. Lessons have been learned from 4G LTE where harmonisation was not possible. The US has a dominant lead in 4G LTE but it is not widely deployed in Europe and the rest of the world where the concentration is very much on 5G. Supporting development of 5G in the US the Federal Communications Commission has voted to make a large block of spectrum available for permission-less 5G development and use. This US strategy of providing spectrum first with flexible-use has been very successful in promoting 4G development and is likely to have a similar impact on 5G uptake.

The uptake of 5G depends upon standardisation to be in place, however, the roll out of 5G is expected to be gradual allowing equipment upgrades to occur before some of the key 5G standards are formalised in 2018 and 2019. There are many pre-standards activities going on around the world in 5G with demonstrators from T-Mobile US, Ericsson, Verizon, and large investment from South Korean companies and China with pilots in European countries. Docomo is publicly committed to having a 5G service up and running for the Tokyo Olympics in 2020. To get joint agreement on technical fundamentals and 5G spectrum bands globally by 2018 NTT Docomo (Japan), KT and SK Telecom (South Korea), and Verizon(US) are forming the 5G Open Trial Specification Alliance to drive technology and standards forward. Pre-standard "5G-ready" equipment using software defined network (SDN) technology will allow network operators and enterprise customers to move to upgrade to full 5G once standardisation is in place for spectrum allocation and licences are issued. Other bodies that are likely to influence 5G standards are the Next Generation Mobile Networks NGMN 5G roadmap targeting 2020 for 5G launch, the GSMA to bring together operator, and the ITU for harmonisation.

9.4 Big Data

In Europe the Digital Single Market (DSM) and data driven economy is driving many activities on data. The European Big Data strategy has funded over 150 research and innovation projects and the Big Data Public Private Partnership has been set up with many key players. Similarly the NSF Big Data Research Initiative is driving activities in the US with the Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21) fostering development and implementation of a national cyberinfrastructure for researchers in science and engineering by making data available within different research communities. Several agencies such as NSF, NIH, DARPA and the Department of Defence (DoD) have their own Big Data Initiatives. Policy in the US focuses more on the actual uses of Big Data and less on its collection and analysis. The US wants to lead both in the international arena and at home. The BD2K Centres of Excellence programme has established 11 Centres of Excellence for Big Data Computing. DARPA and the DoD are also actively engaged. DARPA is creating and providing open source the XDATA software toolkits for applied mathematics, data visualization and computer science. The DoD is increasing Big Data-related research and development with all military services funding R&D in the area. The Big Data Technology roadmap is being led by NIST and this is defining requirements for interoperability, portability, reusability, and extensibility for Big Data analytic techniques, and technology infrastructure in order to support secure and effective adoption of Big Data.

Critical to maximising the benefits from Big Data is the ability to access and share data. In Europe the Open Cities project and Commons for Europe Project are demonstrating the power of open data as well as the Citadel on the Move project that makes it easier for citizens and application developers from across Europe to use Open Data. At a national level a number of projects such as Open Data Helsinki are showing the promise of open data usage. A key barrier is ability to share information effectively. There is a need to rationalise Information Governance regimes across public services and to address the semantics of information to allow sharing. There is also a general lack of openness with respect to data around the world as indicated by the Open Data Barometer published by the Open Data Institute (ODI) which covers 88 countries.

The area of privacy is a key topic on both sides of the Atlantic. In the US there is also an accent on policies to strengthen and stimulate US research in practical privacy-related technologies. This is covered by the 14th Amendment of the US Constitution which gives a person the right to determine what sort of information about them is collected and also how that information is to be used. In the marketplace this is enforced by laws intended to prevent deceptive practices and unfair competition. The Privacy Act of 1974 prevents unauthorised disclosure of personal information held by the federal government. In Europe there are different attitudes and regulation with respect to privacy in different member states. In Germany, for instance privacy is a key concern and privacy is strictly controlled. To try and harmonise the area of privacy the European Data Protection Directive (Directive 95/46/EC) was introduced to protect an individual with respect to processing of personal data and on the free movement of such data. At an international level the “Safe Harbour Privacy Principles” were introduced so that US companies can comply with privacy laws that protect European Union and Swiss citizens. US companies who store European customer data may self-certify that they adhere to 7 key principles. This was challenged and ruled inadequate so a Privacy Shield agreement has been put forward. This is still the subject of intense discussion. In particular, even if US companies have European data centres, cloud services may still require transfers of EU data to the US (or other countries) to provide some features or provide technical support. Encryption is thought to be a way forward. For instance ID data can be encrypted before it enters the cloud with the data owner keeping the encryption keys. This gets around Safe Harbour and Privacy Shield laws, however, it is expensive and only 1% of cloud companies offer encryption.

Standardisation is a key enabler in the field of Big Data. There are already a number of standardisation initiatives at a world-wide level such as the ISO/IEC Joint Technical Committee (JTC) 1 Working Group (WG) on Big Data, the IEEE Standards Association standards related to Big-Data applications and specifically IEEE P2413, and the ITU “Recommendation ITU-T Y.3600” for Big Data services.

9.5 CPS/IoT

The Digital Single Market Strategy and the Digitising European Industry initiative are key drivers in Europe for IoT and CPS with the concept of “Smart Everything Everywhere”. There are major initiatives to cluster activities together and support development of platforms. The IERC (IoT European Research Cluster) brings together 40 EU-funded projects with the aim of defining a common vision, identifying common research challenges and coordinating and encouraging the convergence of ongoing work. Likewise the Alliance for the Internet of Things (AIOTI) has the aim of creating a European IoT ecosystem with further IoT Large Scale Pilots being funded to promote IoT take up. There are also major initiatives such as BUTLER and FIWARE which are providing open, public and royalty-free architectures and specifications to allow developers, service providers, enterprises and other organizations to develop products. This is cross sectoral with 16 Future Internet Accelerators addressing Smart Cities, E-Health, Transport, Energy and Environment, Agrifood, Media and Content, Manufacturing and Logistics, Social and Learning.

In the US Developments in IoT are largely being driven by companies with major players Google, Cisco, etc. dominating the marketplace. The Department of Commerce is promoting growth of the digital economy and as part of the Digital Economy Agenda, the National Telecommunications and Information Administration is initiating an inquiry to review the current technological and policy landscape for IoT and issue a “green paper”. This will highlight potential benefits and challenges, and possible roles for the federal government in fostering the advancement of IoT technologies in partnership with the private sector. Various consortia and alliances have been formed to promote the uptake of IoT. These include the US Industrial Internet Consortium (IIC), the Allseen Alliance (dedicated to providing and open environment for the Internet of Things) and the Open Interconnect Foundation founded by major companies (Intel, Microsoft, Samsung, Qualcomm, GE Digital and Cisco Systems) working on IoT chips, software, platforms and products with the aim of working together towards a single standard for IoT.

The importance of IoT is recognised world-wide and large investments are being made in a number of countries, e.g. the Korean Government is investing $350 million in 300 companies it thinks can compete globally in the next four years, to develop an IoT ecosystem.

Within Europe there are many projects on CPS and the EC strategy has been to cluster these with clusters being formed on CPS and Systems of Systems. To bring industry together in the areas of micro-/nanoelectronics, embedded and Cyber-Physical Systems and smart systems the ECSEL-JU has been set up. Supporting this the ARTEMIS Industry Association (170 members) is also funding large projects such as CRYSTAL on interoperability and EMC on mixed-criticality systems. Competence centres have also been set up to engage with SMEs. At a national level the Industrie 4.0 programme is driving work on CPS in manufacturing.

In the US work on CPS is being driven by the NSF Cyber-Physical Systems programme which has funded over 300 projects. Here basic CPS research is being addressed that can be used across multiple application domains and NSF is working closely with multiple federal government agencies. The NITRD CPS SSG is responsible for coordinating programmes, budgets, and policy recommendations for Cyber-Physical Systems (CPS) research and development (R&D). Although a number of federal agencies have independent research efforts it has been identified that there are still many gaps in the federal R&D portfolio. The NIST Cyber-Physical Systems and Smart Grid Program Office is leading NISTs activities on Cyber-Physical Systems. A Public Working Group (CPS

PWG) has been formed and a CPS Framework has been developed in partnership with industry, academic and government experts. This provides a methodology for understanding, designing and building CPS. NIST also lead the Global City Teams Challenge to help communities around the world work together to address issues including air quality, traffic management and emergency services coordination.

Key challenges are interoperability of systems to allow easier integration of highly complex systems. There is a need to provide standards and regulations that support the creation of an ecosystem of developers and users of CPS and IoT systems. Here a harmonisation between the US and Europe is not only advantageous but strongly needed. Regulation also has a crucial role in the development of CPS and IoT. IoT deployments at scale have many implications including implications for privacy (applications rely on collecting and utilising data from a myraid of sensors) and security where the increasing interconnectedness of systems leads to vulnerabilities to unintentional errors and cyber-attacks. There is also a need for business models and regulation to support market access.

10 Concluding Remarks

This report provides a panorama of ICT policies, regulations, programmes and networks in smart cities, smart transportation and smart energy and also an overview of industry-driven programmes, priorities, networks, major projects in EU and US in the areas of 5G Networks, Big Data, Internet of Things and Cyber-Physical Systems. Additionally, these areas are also actively being pursued at the world level and relevant major programmes around the world have also been identified. Key work with respect to regulations and standards is highlighted for each of the domains.

It is clear that there are many opportunities for joint collaborations between the EU and US. Based on the work in this report a number of suggested collaboration opportunities have been put forward for further discussion in the PICASSO Expert Groups. This includes 15 areas where it may be possible to collaborate on research and policy, 16 areas where there is an opportunity to work together on regulations and 9 areas where it would be beneficial to work together on standards. These will be discussed and refined in future work to identify the key recommendations for future collaboration.

There are many areas where it would be possible to work together. There are key opportunities in the areas of smart cities and IoT/CPS which are rapidly developing areas and where there are common research, regulatory and standardisation needs. There are also great opportunities in the areas of smart energy and smart transportation, however, here there is existing regulation and legislation which needs to be harmonised.

For the underlying technologies which are the basic building blocks of future applications, e.g. 5G, Big Data and IoT/CPS, there are many opportuntiies to work together which would allow bilateral access to EU and US markets and would allow technology and products to be sold on the world stage increasing the competitiveness of EU and US companies in existing and developing markets.

11 References

[1] IBM - Survey on smart cities. A vision of smarter cities. How cities can lead the way into a prosperous and sustainable future. By Susanne Dirks and Mary Keelin, IBM Global Business Services Executive Report, IBM Institute for Business Value.

[2] http://www.ierc.ie/wp-content/uploads/2014/07/IERC-Resident-Engagement-Whitepaper.pdf

[3] *Smarter cities for smarter growth. How cities can optimize their systems for the talent-based economy.* By Susanne Dirks, Constantin Gurdgiev and Mary Keeling. IBM Global Business Services Executive Report, IBM Institute for Business Value.

[4] Mapping Smart Cities in the EU, Directorate General for Internal Policies Economic and Scientific Policy, Jan 2014.

[5] http://www.smart-cities.eu/

[6] European Commission (2011a) 'eGovernment Benchmark Pilot on Open Government and Transparency', prepared by Capgemini, IDC, Rand Europe, Sogeti and DTi. http://ec.europa.eu/information_society/newsroom/cf/dae/itemdetail.cfm?item_id=10145, http://ec.europa.eu/information_society/activities/sustainable_growth/docs/smartcities/smart-cities-adv-group_report.pdf

[7] Communication from the Commission to the European Parliament, The Council and the European Economic and Social Committee 2013, Commission Communication on Smart Cities, October 2012

[8] Smart Cities and Communities Initiative, launched in 2011, http://ec.europa.eu/dgs/connect/en/content/smart-cities

[9] http://setis.ec.europa.eu/implementation/technology-roadmap/european-initiative-on-smart-cities

[10] European Innovation Partnership (EIP) on Smart Cities and Communities (EIP-SCC).

[11] European Commission, 2012, Communication from the Commission – Smart Cities and Communities – European Innovation Partnership.

[12] European Innovation Partnership on Smart Cities and Communities Operational Implementation Plan: First Public Draft Sherpa Group Feb 2014.

[13] European Innovation Partnership on Smart Cities and Communities Strategic Implementation Plan October 2013.

[14] Smart Cities Stakeholder Platform Integrated Action Plan – Report Process & Guidelines Dec 2013.

[15] Smart Cities Stakeholders Platform Finance Working Group Guidance Document using EU Funding Mechanisms for Smart Cities Nov 2013.

[16] https://eu-smartcities.eu/content/small-giants-0

[17] Smart Cities: Background paper, Department for Business, Innovation and Skills, UK, October 2013.

[18] "Preparing the way for smart cities" BSI: https://www.bsigroup.com/Documents/standards/case-studies/BSI-supporting-innovation-smart-cities-case-studies-UK-EN.pdf

[19] BIS Research Paper No. 135, "International Case Studies on Smart Cities", Department for Business, Innovation and Skills, UK, October 2013.

[20] BIS Research Paper No.136, "The Smart City Market: Opportunities for the UK", Department for Business, Innovation and Skills, UK, October 2013.

[21] https://www.gov.uk/government/news/new-initiative-to-support-40-billion-smart-cities-in-the-uk

[22] http://www.bis.gov.uk/foresight/our-work/projects/current-projects/future-of-cities

[23] http://liveablecities.org.uk/

[24] Smart Cities Transforming the 21st century city via the creative use of technology, ARUP 2010.

[25] HyperCat consortium: http://www.hypercat.io/

[26] https://www.gov.uk/government/news/launch-of-smart-cities-initiative

[27] http://www.smartsantander.eu/

[28] http://www.spiegel.de/international/world/santander-a-digital-smart-city-prototype-in-spain-a-888480.html

[29] http://spain-lab.net/project/smart-citizen-fab-lab-barcelonaiaachangar/

[30] http://cityclimateleadershipawards.com/barcelona-barcelona-smart-city/

[31] http://amsterdamsmartcity.com/

[32] http://www.spiegel.de/international/business/amsterdam-as-smart-city-going-green-fast-in-holland-a613503.html

[33] http://www.npr.org/sections/parallels/2013/06/12/190693661/tallinn-the-former-soviet-city-that-gave-birth-to-skype

[34] http://www.river-cities.net/pages/show/Stockholm_2013

[35] https://www.whitehouse.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help

[36] https://www.nitrd.gov/sccc/

[37] Citizen-Centred Governance: The Mayor's Office of New Urban Mechanics and the Evolutions of CRM in Boston, Case Study, By Susan Crawford and Dana Walters, July 30, 2013.

[38] Portney, K. 2005. Civic Engagement and Sustainable Cities in the United States. Public Administration Review, 65(5): 579-591.

[39] http://www.cisco.com/c/dam/en_us/about/ac79/docs/ps/motm/City-24x7_PoV.pdf

[40] http://venturebeat.com/2014/05/01/nyc-wants-to-turn-all-of-its-payphones-into-a-massive-citywide-wifi-network/

[41] http://www.hudsonyardsnewyork.com/

[42] http://bottomline.seattle.gov/2013/07/09/high-performance-building-pilot/

[43] [http://www.forbes.com/sites/ptc/2014/08/15/5-u-s-cities-using-technology-to-become-smart-and-connected/#39f350b33539

[44] DC Geographic Information System Program: http://octo.dc.gov/node/718302

[45] http://www.smartcommunitieschicago.org/index.html

[46] http://www.fastcoexist.com/1680967/the-top-10-smartest-cities-in-north-america#1

[47] http://smartcities.ieee.org/home/ieee-smart-cities-initiative.html

[48] http://www.scmp.com/comment/insight-opinion/article/1898359/how-hong-kong-can-become-smart-city-connected-future

[49] Research Report on Smart city, Central Policy Unit, The Government of the Hong Kong Special Administrative Region. September 2015. http://www.cpu.gov.hk/doc/en/research_reports/CPU%20research%20report%20-%20Smart%20City(en).pdf

[50] "Comparative Study of Smart Cities in Europe and China" prepared for Ministry of Industry and Information Technology (MIIT), DG CNECT EU Commission with the China Academy of Telecommunications Research (CATR), White paper, June 2014.

[51] http://symposium.net.au/australia-its-time-to-be-the-smart-country/

[52] http://www.computerweekly.com/news/4500256861/Australian-cities-are-getting-smarter

[53] http://www.adelaidecitycouncil.com/assets/documents/NATGEO-adelaide-smart-city.pdf?utm_source=media%20release&utm_medium=PDF-download&utm_content=smart-city&utm_campaign=NatGeo

[54] http://www.citysmart.com.au/

[55] http://www.cityofsound.com/files/c40_melbourne_report_final_email.pdf

[56] http://www.zdnet.com/article/city-of-melbourne-pushes-for-smart-cities-built-on-it/

[57] http://newsroom.cisco.com/feature-content?type=webcontent&articleId=1738492

[58] http://www.nesta.org.uk/news/10-people-centred-smart-city-initiatives/better-reykjavik

[59] http://searchsoftwarequality.techtarget.com/news/4500245158/Data-privacy-lawyer-explains-data-by-design

[60] Privacy and snooping in Smart cities: utopia or reality? vol 11, issue 01, Engineering and Technology Magazine, 18 January 2016, By Katia Moskvitch (http://eandt.theiet.org/magazine/2016/01/privacy-and-smart-cities.cfm?utm_source=Adestra&utm_campaign=E%26T%20News%20Trial%20Programme%20Daily%20Members%20A1%20-%20Day7&utm_medium=Newsletters%20-%20E%26T%20News&utm_content=E%26T%20News%20-%20Frequency%20Trial%20Programme%20%28Daily%29&utm_term=http%3A%2F%2Feandt.theiet.org%2Fmagazine%2F2016%2F01%2Fprivacy-and-smart-cities.cfm&utm_contact=32547089)

[61] http://www.bsigroup.com/Documents/standards/case-studies/BSI-supportinginnovation-smart-cities-case-studies-UK-EN.pdf

[62] Smart city framework – Guide customer service to establishing strategies for smart cities and communities, BSI Standards Publication, PAS 181:2014, Department for Business Innovation and Skills, February 2014.

[63] www.etsi.org

[64] European Commission (2012a) 'Energy Efficient Buildings'. http://ec.europa.eu/research/industrial_technologies/energy-efficient-buildings_en.html

[65] European Commission (2011b) 'Report of the Meeting of Advisory Group ICT Infrastructure for Energy-Efficient Buildings and Neighbourhoods for Carbon-Neutral Cities: Strategic Priorities for the New Framework Programme for Research and Innovation Covering the Period 2014–2020', 16 September.

[66] http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/507480/IPOLITRE_ET(2014)507480_EN.pdf

[67] http://climate-adapt.eea.europa.eu/cities

[68] https://setis.ec.europa.eu/set-plan-implementation/technology-roadmaps/european-initiative-smart-cities

[69] The ICT Roadmap for Energy Efficient Neighbourhoods (http://www.ireenproject.eu/)

[70] http://www.smartgrids.eu/KIC-InnoEnergy

[71] https://eu-smartcities.eu/content/positive-energy-blocks

[72] http://www.world-nuclear.org/information-library/country-profiles/countries-a-f/france.aspx (Updated 31 March 2016)

[73] http://www.metering.com/frances-smart-grid-project-sogrid-rolls-out-to-1000-homes/

[74] The future of Heating: Meeting the challenge, published in March 2013, UK https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190149/16_04-DECC-The_Future_of_Heating_Accessible-10.pdf

[75] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/276656/smart_meter_roll_out_for_the_domestic_and_small_and_medium_and_non_domestic_sectors.pdf

[76] Low Carbon Pioneer Cities Heat Networks Project: a process evaluation. Department of Energy & Climate Change, January 2015 (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/457057/Pioneer_Cities_Evaluation_27_08.pdf)

[77] http://www.lowcarboninnovation.co.uk/working_together/strategic_framework/overview/

[78] https://www.gov.uk/government/news/20m-up-for-grabs-for-energy-storage-innovation

[79] https://www.ofgem.gov.uk/sites/default/files/docs/2010/11/lcnbro_0.pdf

[80] http://www.ifhp.org/sites/default/files/staff/PreConferenceReport_FINAL080915.pdf

[81] https://www.epa.gov/laws-regulations/summary-energy-independence-and-security-act

[82] https://www.treasury.gov/initiatives/recovery/Pages/recovery-act.aspx

[83] Blueprint for a Secure Energy Future, March 30, 2011 (http://www.whitehouse.gov/sites/default/files/blueprint_secure_energy_future.pdf)

[84] "A Policy Framework for the 21st Century Grid: Enabling Our Secure Energy Future", National Science and Technology Council, June 2011.

[85] www.SmartGrid.gov

[86] http://www.whitehouse.gov/sites/default/files/microsites/ostp/nstc-smart-grid-june2011.pdf

[87] Global Inventory and Analysis of Smart Grid Demonstration Projects, DNV KEMA Energy & Sustainability, October 2012.

[88] http://smartgrid.ieee.org/about-ieee-smart-grid

[89] Evolution of the smart grid in China, Xu et al., McKinsey on Smart Grid, Summer 2010.

[90] http://www.metering.com/china-to-invest-us31bn-in-provincial-smart-grid-by-2020/

[91] http://www.metering.com/canadas-ontario-seeks-new-projects-for-smart-grid-fund/

[92] http://www.metering.com/smart-grid-trends-in-japan-7-things-to-know/

[93] http://www.metering.com/smart-grid-equipment-market-to-grow-in-south-korea/

[94] http://www.metering.com/brazils-power-outages-push-case-for-smart-grid-technology/

[95] http://www.metering.com/smart-grid-india-government-approves-us210m-investment/

[96] Smart Grid, Smart City: Shaping Australia's Energy Future, Executive Report, July 2014. By Arup, Energeia, Frontier Economics, Institute for Sustainable Futures (University of Technology Sydney), http://industry.gov.au/Energy/Programmes/SmartGridSmartCity/Documents/SGSC-Executive-Report-National-Cost-Benefit.pdf

[97] [http://www.metering.com/smart-grid-mexico-smart-meters/

[98] http://magrid.raabassociates.org/Articles/DNVKEMAReportGlobalInventorySmartGrid%20Projects.pdf

[99] http://www.smartgrids.eu/node/70

[100] https://ec.europa.eu/energy/en/topics/markets-and-consumers/smart-grids-and-meters/smart-grids-task-force

[101] http://www.smartgrids.eu/ACER

[102] http://www.smartgrids.eu/CEER

[103] Smarter Grids: the opportunity, Department Energy and Climate Change, December 2009 (file:///C:/Users/uos/Downloads/smartergridsopportunity.pdf)

[104] A Smart Grid Routemap, Electricity Networks Strategy Group, , February 2010 (http://webarchive.nationalarchives.gov.uk/20100919181607/http:/www.ensg.gov.uk/index.php?article=126)

[105] https://www.ofgem.gov.uk/electricity/distribution-networks/network-innovation/low-carbon-networks-fund

[106] https://www.gov.uk/government/publications/2010-to-2015-government-policy-uk-energy-security/2010-to-2015-government-policy-uk-energy-security

[107] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/285417/Smart_Grid_Vision_and_RoutemapFINAL.pdf

[108] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/486362/Towards_a_smart_energy_system.pdf

[109] http://www.smartgrids.eu/sites/default/files/private/2011_03_01_mandate_m490_en%5B1%5D.pdf

[110] http://www.smartgrids.eu/standards

[111] http://www.cencenelec.eu/standards/Sectors/SustainableEnergy/SmartGrids/Pages/default.aspx

[112] NIST Special Publication 1108R2, NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0

[113] http://www.nist.gov/smartgrid/upload/NIST_Framework_Release_2-0_corr.pdf

[114] Cyber Security of the Smart Grids, Summary Report. The paper was produced by the European Commission using input and comments from the Expert Group on the Security and Resilience of Communication Networks and Information Systems for Smart Grids (file:///C:/Users/uos/Downloads/SummaryReport.pdf).

[115] http://www.mcafee.com/uk/resources/white-papers/wp-smart-grid-cyber-security.pdf

[116] https://www.whitehouse.gov/sites/default/files/microsites/ostp/nstc-smart-grid-june2011.pdf

[117] http://www.nist.gov/el/smartgrid/cybersg.cfm

[118] “Roadmap for Cross-Modal Transport Infrastructure Innovation - Towards a Performing Infrastructure", ERTRAC, ERRAC, Waterborne, ACARE, ECTP, 2013.

[119] "A Sustainable Future for Transport", Directorate-General for Energy and Transport European Commission, 2009.

[120] “European Transport Policy for 2010 - Time to Decide”, CEC, 2001.

[121] http://www.errac.org/

[122] http://www.ertrac.org/

[123] http://www.acare4europe.com/

[124] http://www.waterborne-tp.org/

[125] “Keep Europe Moving - Sustainable Transport for our Continent”, CEC, 2006.

[126] "Traffic Management for Land Transport", Transport research Knowledge Centre, Directorate-General for Energy and Transport, European Commission, 2009.

[127] www.car-2-car.org

[128] Thematic Research Summary, Intelligent transport systems, 2014 (http://www.transport-research.info/sites/default/files/thematic-analysis/20150430_174628_50849_TRS22_fin.pdf)

[129] http://www.transport-research.info/sites/default/files/thematic-analysis/20150430_174628_50849_TRS22_fin.pdf

[130] Policy Paper: Sustainable Urban Mobility and the Smart City. Scope – Finance – Community Building. POLIS | Cities and regions networking for innovative transport solutions, January 2016, http://www.polisnetwork.eu/uploads/Modules/PublicDocuments/2016-polis-smart-city-policy-paper.pdf

[131] http://ec.europa.eu/transport/themes/its/index_en.htm

[132] “Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010 on the framework and deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport", 6th August 2010.

[133] “Towards a 50% More Efficient Road Transport System by 2030”, ERRAC, 2010.

[134] “Vision 2020 and Challenges”, White Paper, ERTRAC, 2004

[135] http://www.transportncps.net/index.php?option=com_k2&view=item&layout=item&id=249&Itemid=344]

[136] https://eu-smartcities.eu/content/new-mobility-services

[137] http://www.drive-c2x.eu/project

[138] http://www.ertico.com/prevent

[139] http://www.cvisproject.org/cvisproject/en/about_cooperative_systems/introduction/index.html

[140] http://www.safespot-eu.org/
[141] http://ercim-news.ercim.eu/en67/special-theme-embedded-intelligence/coopers-automotive-visions-beyond-in-car-driver-assistance
[142] http://www.cvisproject.org/en/links/pre-drive_c2x.htm
[143] http://www.ecomove-project.eu/links/comesafety/
[144] http://www.haveit-eu.org/
[145] https://www.media.volvocars.com/global/en-gb/media/pressreleases/136182/volvo-car-group-initiates-world-unique-swedish-pilot-project-with-self-driving-cars-on-public-roads
[146] http://ec.europa.eu/research/transport/road/green_cars/index_en.htm
[147] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/smart-green-and-integrated-transport
[148] http://urbanforesight.org/projects/projectsites/ev4scc/about-ev4scc/
[149] https://eu-smartcities.eu/content/electromobility
[150] U. Clausen, "A European Perspective of Current and Future Research on Sustainability in Transportation", TRB Conference, Washington 2013.
[151] http://europa.eu/legislation_summaries/environment/air_pollution/l28186_en.htm
[152] http://www.dhl.com/en/about_us/green_solutions/gogreen_products_services.html
[153] http://www.cofret-project.eu/
[154] http://www.greenfreighteurope.eu/
[155] https://www.government.nl/latest/news/2015/12/02/government-authorities-investing-70-million-in-intelligent-transport-systems
[156] http://www-03.ibm.com/press/us/en/pressrelease/40212.wss
[157] [http://www.verkeersnet.nl/3728/enschede-krijgt-vip-reistijdinformatiesysteem/
[158] [http://meet.barcelona.cat/en/visit-barcelona/get-around-the-city/electric-vehicles
[159] http://w41.bcn.cat/mou-te/
[160] http://www.catalonia.com/en/trade-with-catalonia/sustainablemobility.jsp
[161] http://www.zaragoza.es/ciudad/viapublica/movilidad/trafico/trafico.htm
[162] http://www.zaragoza.es/contenidos/sectores/tecnologia/Estrategia-Ciencia-Tecnologia-en.pdf
[163] https://www.ibm.com/midmarket/ie/en/att/pdf/IBM_DCC_130715.pdf
[164] http://www.mobithess.gr/
[165] https://ts.catapult.org.uk/
[166] http://www.mira.co.uk/our-services/intelligent-transport-systems/
[167] http://www.ieee802.org/11/Reports/tgp_update.htm
[168] "Roadmap to a Single European Transport Area - Towards a competitive and resource efficient transport system", European Commission, 2011
[169] "Joint Strategy for European Rail Research -Vision 2020", 2001.
[170] "Strategic Rail Research Agenda 2020 European Rail Advisory Council", May 2007.
[171] http://errac.uic.org/IMG/pdf/Suburban_and_Regional_Railways_Landscape_in_Europe.pdf
[172] http://www.uitp.org/sites/default/files/cck-focus-papers/files/errac_metrolr_tramsystemsineurope.pdf
[173] http://www.railway-research.org/Rail-Research-in-Europe
[174] "RailRoute 2050 - The Sustainable Backbone of the Single European Transport Area", ERRAC.
[175] http://www.errac.org/foster-rail/
[176] http://errac.uic.org/spip.php?article13
[177] http://www.ertms.net/

[178] http://www.shift2rail.org/shift%C2%B2rail-officially-launched-alongside-other-horizon2020-jtis-at-the-european-commissions-launch-event-in-brussels/

[179] "Modernising the European Sky", SESAR. 2011.

[180] "Today's Partners for Tomorrow's Aviation", SESAR,

[181] http://www.gsa.europa.eu/galileo-0

[182] https://www.atmmasterplan.eu/

[183] "European ATM Master Plan - The Roadmap for Sustainable Air Traffic Management", Edition 2, SESAR, October 2012.

[184] "European ATM Master Plan Edition 2", Factsheet, SESAR 2012.

[185] http://www.cleansky.eu/

[186] https://www.thalesgroup.com/en/united-kingdom/defence/watchkeeper

[187] http://www.baesystems.com/enhancedarticle/BAES_157659/taranis-unmanned

[188] http://astraea.aero/

[189] http://www.imo.org/OurWork/Environment/PollutionPrevention/AirPollution/Pages/Nitrogen-oxides-(NOx)-%E2%80%93-Regulation-13.aspx

[190] http://www.vsat-systems.com/

[191] http://www.waterborne-tp.org/

[192] "Waterborne Transport Thematic Research Summary", Transport Research Knowledge Centre, Directorate-General for Mobility and Transport, European Commission, 2010.

[193] "Waterborne Vision 2025 - Waterborne Transport and Operations Meeting the Challenges through Ambitious Innovation", December 2012.

[194] "Waterborne Strategic Research Agenda Implementation", May 2011.

[195] "Waterborne Strategic Research Agenda Overview Issue II", May 2011.

[196] http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/2634-mg-4.2-2014.html

[197] http://www.e-maritime.com/

[198] http://ec.europa.eu/maritimeaffairs/documentation/multimedia/videos/items/video3_en.htm

[199] http://www.economist.com/news/technology-quarterly/21598318-autonomous-cargo-vessels-could-set-sail-without-crew-under-watchful-eye

[200] http://www.its.dot.gov/its_program/about_its.htm#sthash.JFHy2v82.dpuf

[201] [http://www.ssti.us/

[202] The Innovative DOT: A Handbook of Policy and Practice. State Smart Transportation Initiative (SSTI) & Smart Growth America (SGA), January 2015. (http://www.ssti.us/2015/01/the-innovative-dot-a-handbook-of-policy-and-practice-ssti-sga/)

[203] http://www.smartgrowthamerica.org/issues/transportation

[204] https://www.transportation.gov/smartcity

[205] http://arc.applause.com/2015/12/08/smart-cities-challenge/

[206] (https://www.transportation.gov/sites/dot.gov/files/docs/Draft_Beyond_Traffic_Framework.pdf

[207] http://traffic.berkeley.edu/project

[208] http://www.umtri.umich.edu/

[209] J. Parker, "Applying a Systems of Systems Approach for Improved Transportation", SAPIENS, Vol. 3, Issue 2, 2010.

[210] http://www.apple.com/uk/ios/carplay/?cid=wwa-uk-kwg-features-com

[211] http://www.openautoalliance.net/

[212] http://www.inrix.com/companyoverview.asp

[213] https://www.whitehouse.gov/sites/default/files/blueprint_secure_energy_future.pdf

[214] http://en.wikipedia.org/wiki/Google_driverless_car

[215] http://www.forbes.com/sites/alexkonrad/2013/11/01/meet-orion-software-that-will-save-ups-millions-by-improving-drivers-routes/

[216] http://www.nytimes.com/2014/08/07/us/delays-persist-for-us-high-speed-rail.html?_r=0 (August 06, 2014)

[217] http://www.faa.gov/nextgen/

[218] Press Release “Lockheed Martin Plans "System of Systems" Approach For UAVS”, Farnborough, England, July 23rd, 2002

[219] http://www.lockheedmartin.co.uk/us/what-we-do/aerospace-defense/unmanned-systems.html

[220] http://www.amazon.com/b?node=8037720011

[221] http://www.ihsglobalinsight.com/Highlight/HighlightDetail15614.htm

[222] *The Maritime Administration and the U.S. Marine Transportation System: A Vision for the 21st Century.* United States Department of Transportation Maritime Administration, November 2007.

[223] http://www.maritime-executive.com/article/us-announces-2016-funding-for-infrastructure-projects]

[224] http://www.americanmaritimepartnership.com/about-amp/

[225] http://www.marketsandmarkets.com/PressReleases/smart-transportation.asp

[226] http://www.itsnetwork.org/

[227] http://www.ertico.com/ertico-its-europe/

[228] http://itsworldcongress.org/

[229] http://www.ewg.apec.org/keypillars.html

[230] http://www.itsinternational.com/categories/utc/news/us-to-offer-smart-transportation-solutions-for-three-indian-cities/

[231] http://ec.europa.eu/digital-agenda/en/ecall-time-saved-lives-saved

[232] http://www.safetrip.eu/

[233] http://www.cvisproject.org/en/links/c2c-cc.htm

[234] http://ec.europa.eu/transport/themes/its/road/action_plan/

[235] M/453 http://ec.europa.eu/enterprise/sectors/ict/files/standardisation_mandate_en.pdf

[236] ec.europa.eu/digital-agenda/en/towards-5g

[237] ANALYSIS, Understanding 5G: Perspectives on future technological advancements in mobile, GSMA Intelligence, December 2014.

[238] 5G Infrastructure Association: Vision White Paper, February 2016.

[239] https://www.metis2020.com/

[240] https://metis-ii.5g-ppp.eu/

[241] http://www.eitdigital.eu/news-events/news/article/toward-green-5g-mobile-networks-5green-new-project-launched/#gridView

[242] http://www.ict-ijoin.eu/

[243] http://www.ict-crowd.eu/

[244] http://networld2020.eu/overview/

[245] http://www.surrey.ac.uk/mediacentre/press/2015/5g-innovation-centre-officially-opens-university-surrey

[246] http://www.surrey.ac.uk/5gic

[247] http://www.ctr.kcl.ac.uk/5G2015/

[248] http://eandt.theiet.org/magazine/2015/03/tactile-internet-5g.cfm
[249] https://5g.co.uk/news/ericsson-king%E2%80%99s-college-london-5g-research/4066/
[250] http://www.fast-zwanzig20.de/EN/
[251] http://5glab.de/
[252] http://www.5g-berlin.org/
[253] https://www.telkom.com/5ghaus
[254] http://5gtnf.fi/overview/
[255] www.tekes.fi/5thGear
[256] http://www.5gamericas.org/en/newsroom/industry-news/commerce-approves-mobile-now-act-and-nominations/
[257] https://www.whitehouse.gov/the-press-office/2016/07/15/fact-sheet-administration-announces-advanced-wireless-research
[258] http://www.4gamericas.org/en/newsroom/press-releases/4g-americas-evolves-name-5g-americas/
[259] http://www.fiercewireless.com/tech/special-reports/top-5-academic-institutions-leading-5g-research
[260] https://wncg.org/tags/5g-wireless-networks
[261] http://wireless.engineering.nyu.edu/
[262] http://www.5gamericas.org/en/newsroom/member-news/t-mobile-and-ericsson-partner-5g-pre-standard-trials/
[263] https://www.washingtonpost.com/news/innovations/wp/2015/10/26/what-is-5g-and-why-should-lawmakers-care/
[264] http://www.cnet.com/uk/news/verizon-to-hold-worlds-first-crazy-fast-5g-wireless-field-tests-next-year/
[265] https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf
[266] http://www.fiercewireless.com/tech/story/china-south-korea-commit-5g-leadership-while-japan-and-us-rely-private-effo/2014-06-08
[267] http://about.keysight.com/en/newsroom/pr/2014/11aug-em14115.shtml
[268] http://www.5gforum.org/#!eng/cvb1
[269] https://www.etri.re.kr/eng/sub6/sub6_0101.etri?departCode=6
[270] http://www.timesofmalta.com/articles/view/20150712/local/pm-thanks-sai-mizzi-as-chinese-telecoms-giant-prepares-to-test-5g-in.576179
[271] http://www.huawei.com/minisite/5g/en/huawei-5gppp.html
[272] http://www.rtvnoord.nl/nieuws/152709/Noord-Groningen-krijgt-onvoorstelbaar-snel-mobiel-internet
[273] http://www.gsma.com/newsroom/press-release/gsma-publishes-new-report-outlining-5g-future/
[274] http://www.fiercewireless.com/story/verizons-mcadam-5g-speeds-will-be-1-gbps-and-will-be-live-verizon-hq-januar/2015-12-08
[275] https://www.nttdocomo.co.jp/english/corporate/technology/whitepaper_5g/index.html
[276] http://spectrum.ieee.org/tech-talk/telecom/wireless/5g-coming-sooner-not-later
[277] https://uk.news.yahoo.com/nokia-5g-pickup-could-begin-early-2017-chief-193049841--finance.html
[278] http://www.itu.int/en/ITU-T/focusgroups/imt-2020/Pages/default.aspx
[279] http://spectrum.ieee.org/telecom/wireless/telecom-experts-plot-a-path-to-5g
[280] https://www.washingtonpost.com/news/innovations/wp/2015/10/26/what-is-5g-and-why-should-lawmakers-care/
[281] http://www.cnet.com/news/how-5g-will-push-a-supercharged-network-to-your-phone-home-and-car/

[282] http://www.forbes.com/sites/larrydownes/2015/09/09/judgment-day-for-the-fccs-latest-net-neutrality-folly/#14b16b046f92
[283] http://www.forbes.com/sites/mergermarket/2015/12/04/fcc-facing-identity-crisis-over-net-neutrality-challenge/#5a8f3b9b2908
[284] https://ec.europa.eu/digital-single-market/en/digital-single-market
[285] Towards a thriving data-driven economy. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels, July 2014.
[286] http://ec.europa.eu/digital-agenda/en/open-data
[287] https://ec.europa.eu/digital-single-market/en/big-data-value-public-private-partnership
[288] http://ec.europa.eu/digital-agenda/en/blog/open-and-smart-cities-common-future
[289] http://opencities.net/
[290] https://ec.europa.eu/digital-single-market/en/blog/open-and-smart-cities-common-future
[291] https://www.govdata.de/
[292] http://data.amsterdam.nl/
[293] http://commonsforeurope.net/
[294] http://www.esade.edu/web/eng
[295] http://codeforeurope.net/]
[296] https://www.codeforamerica.org/
[297] http://guifi.net/
[298] http://www.citadelonthemove.eu/en-us/about/aims.aspx
[299] http://www.forumvirium.fi/en/project-areas/smart-city/open-helsinki-hack-at-home
[300] https://www.gov.uk/government/publications/e-infrastructure-strategy-roadmap-for-development-of-advanced-computing-data-and-networks
[301] https://www.gov.uk/public-data-group
[302] http://theodi.org/
[303] Report To The President Big Data And Privacy: A Technological Perspective", Executive Office of the President President's Council of Advisors on Science and Technology May 2014 (https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf)
[304] http://www.nsf.gov/cise/news/bigdata.jsp
[305] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504730, http://www.nsf.gov/cise/aci/cif21/CIF21Vision2012current.pdf
[306] https://datascience.nih.gov/bd2k/funded-programs
[307] http://www.darpa.mil/program/xdata
[308] https://defensesystems.com/articles/2015/07/06/us-cyber-forces-private-sector-help.aspx
[309] http://www.disa.mil/
[310] http://www.nist.gov/itl/bigdata/20150406_big_data_framework.cfm
[311] http://theodi.org/news/open-data-barometer-the-odi-publishes-key-research-as-challenger-initiatives-gather-speed
[312] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=URISERV%3Al14012
[313] http://ec.europa.eu/justice/data-protection/
[314] http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm
[315] http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/108

[316] http://www.export.gov/safeharbor/index.asp

[317] Politico, Vol. 2 No.6, February 11-16 2016. Special Report Privacy Shield pp.13-16

[318] https://www.law.cornell.edu/constitution/amendmentxiv

[319] https://www.justice.gov/opcl/privacy-act-1974

[320] http://www.inc.com/encyclopedia/childrens-online-privacy-protection-act-COPPA.html

[321] ISO/IEC JTC 1, Information Technology, Big Data, Preliminary report 2014, ISO, IEC. http://www.iso.org/iso/big_data_report-jtc1.pdf

[322] http://standards.ieee.org/develop/project/2413.html

[323] http://www.itu.int/rec/T-REC-Y.3600-201511-I

[324] Gartner, "Top Strategic Predictions for 2016 and Beyond: The Future Is a Digital Thing"

[325] H.A. Thompson, "Cyber-Physical Systems: Uplifting Europe's Innovation Capacity", Report from the EC Workshop in Brussels, Belgium, October 2013

[326] https://ec.europa.eu/digital-single-market/en/europe-2020-strategy

[327] http://europa.eu/rapid/press-release_IP-16-1407_en.htm

[328] http://www.internet-of-things-research.eu/

[329] https://ec.europa.eu/digital-single-market/en/alliance-internet-things-innovation-aioti

[330] https://ec.europa.eu/digital-single-market/en/news/horizon-2020-work-programme-2016-2017-internet-things-large-scale-pilots

[331] http://www.iot-butler.eu/

[332] http://open-platforms.eu/?s=butler

[333] https://www.fiware.org/

[334] https://artemis-ia.eu/ecsel-joint-undertaking.html

[335] https://artemis-ia.eu/

[336] https://artemis-ia.eu/project/46-crystal.html

[337] http://www.artemis-emc2.eu/

[338] https://ec.europa.eu/digital-single-market/en/news/innovation-manufacturing-smes-i4ms-initiative-enhancing-digital-transformation-european

[339] https://smartanythingeverywhere.eu/

[340] http://www.plattform-i40.de/I40/Navigation/DE/Industrie40/industrie40.html

[341] http://www.iiconsortium.org/

[342] https://allseenalliance.org/

[343] http://openconnectivity.org/

[344] http://www.computerworld.com/article/3035084/internet-of-things/another-iot-group-ocf-may-really-make-it-all-work.html

[345] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503286

[346] https://www.nitrd.gov/nitrdgroups/index.php?title=Cyber_Physical_Systems_(CPS_SSG)#title

[347] DRAFT, Framework for Cyber-Physical Systems, Release 0.8, September 2015, Cyber Physical Systems Public Working Group (http://www.cpspwg.org/)

[348] Strategic R&D Opportunities for 21st Century Cyber-Physical Systems. Connecting computer and information systems with the physical world. January 2013. (http://www.nist.gov/el/upload/12-Cyber-Physical-Systems020113_final.pdf)

[349] https://www.us-ignite.org/globalcityteams/about/#sthash.e4XmCrwo.dpuf

[350] http://www.intel.co.uk/content/dam/www/public/us/en/documents/corporate-information/policy-iot-framework.pdf

[351] http://www.nist.gov/cps/cpswpg_security.cfm

[352] https://www.nitrd.gov/nitrdgroups/index.php?title=Cyber_Security_and_Information_Assurance_Interagency_Working_Group_(CSIA_IWG)#title

[353] Federal Cybersecurity Research and Development Strategic Plan, *Ensuring Prosperity and National Security*, National Science and Technology Council, Networking and Information Technology Research and Development Program. February 2016

[354] https://www.nitrd.gov/Publications/PublicationDetail.aspx?pubid=61

[355] E&T Engineering and Technology Magazine vol 11, issue 02, IoT security threat, as embedded systems struggle, 15 February 2016, By Chris Edwards.

12 APPENDIX A – 5G Projects

To be supplied.

13 APPENDIX B – Big Data Projects

No	Acronym	Full Title	Website reference	Funding Body	Abstract
1	SUMMA	Scalable Understanding of Multilingual Media	http://cordis.europa.eu/project/rcn/199857_en.html	European Commission H2020 Call:ICT-16-2015	Media monitoring enables the global news media to be viewed in terms of emerging trends, people in the news, and the evolution of story-lines. The massive growth in the number of broadcast and Internet media channels means that current approaches can no longer cope with the scale of the problem. The aim of SUMMA is to significantly improve media monitoring by creating a platform to automate the analysis of media streams across many languages, to aggregate and distil the content, to automatically create rich knowledge bases, and to provide visualisations to cope with this deluge of data. SUMMA has six objectives: (1) Development of a scalable and extensible media monitoring platform; (2) Development of high-quality and richer tools for analysts and journalists; (3) Extensible automated knowledge base construction; (4) Multilingual and cross-lingual capabilities; (5) Sustainable, maintainable platform and services; (6) Dissemination and communication of project results to stakeholders and user group. Achieving these aims will require advancing the state of the art in a number of technologies: multilingual stream processing including speech recognition, machine translation, and story identification; entity and relation extraction; natural language understanding including deep semantic parsing, summarisation, and sentiment detection; and rich visualisations based on multiple views and dealing with many data streams. The project will focus on three use cases: (1) External media monitoring - intelligent tools to address the dramatically increased scale of the global news monitoring problem; (2) Internal media monitoring - managing content creation in several languages efficiently by ensuring content created in one language is reusable by all other languages; (3) Data journalism. The outputs of the project will be field-tested at partners BBC and DW, and the platform will be further validated through innovation intensives such as the BBC NewsHack.

2	STREAMLINE	STREAMLINE	http://cordis.europa.eu/project/rcn/199862_en.html	European Commission H2020 Call:ICT-16-2015	STREAMLINE will address the competitive advantage needs of European online media businesses (EOMB) by delivering fast reactive analytics suitable in solving a wide array of problems, including addressing customer retention, personalised recommendation, and more broadly targeted services. STREAMLINE will develop cross-sectorial analytics drawing on multi-source data originating from online media consumption, online games, telecommunications services, and multilingual web content. STREAMLINE partners face big and fast data challenges. They serve over 100 million users, offer services that produce billions of events, yielding over 10 TB of data daily, and possess over a PB of data at rest. Their business use-cases are representative of EOMB, which cannot be handled efficiently & effectively by state-of-the-art technologies, as a consequence of system and human latencies. System latency issues arise due to the lack of appropriate (data) stream-oriented analytics tools and more importantly the added complexity, cost, and burden associated with jointly supporting analytics for both "data at rest" and "data in motion." Human latency results from the heterogeneity of existing tools and the low level programming languages required for development using an inordinate number of boilerplate codes that are system specific (e.g., Hadoop, SolR, Esper, Storm, and databases) and a plethora of scripts required to glue systems together. Our research and innovation actions, include addressing the challenges brought on by system and human latencies. In this regard, STREAMLINE will: 1. Develop a high level declarative language and user-interface, and corresponding automatic optimisation, parallelisation, and system adaptation technologies that reduce the programming expertise required by data scientists, thereby enabling them to more freely focus on domain specific matters. 2. Overcome the complexity of the so-called 'lambda architecture' by delivering simplified operations that jointly support "data at rest" and "data in motion" in a single system and is compatible with the Hadoop ecosystem. 3. Develop fast reactive machine learning technologies based on distributed parameter servers and fully distributed asynchronous and approximate algorithms for fast results at high input rates. The impact of developing a European open source tool for analysing "data at rest" and "data in motion" in a single system featuring a high level declarative language and a fast reactive machine learning library is much wider than just the recommender, ad targeting, and customer retention applications that the industrial partners in STREAMLINE will use to demonstrate the business value of our work for the data economy. Our open source tools will help Europe, in general, since they lower the big data analytics skills barrier, broaden the reach of data analytics tools, and are applicable to diverse market sectors, including healthcare, manufacturing, and transportation. Thereby, enabling a broad number of European SMEs in other markets to explore and integrate these technologies into their businesses. At the same time, STREAMLINE will provide a solid foundation for big data leadership in Europe, by providing an open-source platform ready to be used by millions of stakeholders in companies, households, and government.

					The STREAMLINE consortium comprises world-renowned scientists and innovators in the areas of database systems (DFKI), distributed systems (SICS), and machine learning (SZTAKI) who have won many international awards, hold 18 patents collectively, and have founded and advised nine startups. Complementing the research excellence are four leading European enterprises in the data economy, in the areas of global telecommunication services (e.g., Internet, IPTV, mobile, and landline networks) (PT), games and entertainment (Rovio), media content streaming (NMusic), and web-scale data extraction and business analytics (IMR), with Petab.
3	PROTEUS	Scalable online machine learning for predictive analytics and real-time interactive visualization	http://cordis.europa.eu/project/rcn/199839_en.html	European Commission H2020 Call:ICT-16-2015	PROTEUS mission is to investigate and develop ready-to-use scalable online machine learning algorithms and interactive visualization techniques for real-time predictive analytics to deal with extremely large data sets and data streams. The developed algorithms and techniques will form a library to be integrated into an enhanced version of Apache Flink, the EU Big Data platform. PROTEUS will contribute to the EU Big Data area by addressing fundamental challenges related to the scalability and responsiveness of analytics capabilities. The requirements are defined by a steelmaking industrial use case. The techniques developed in PROTEUS are however, general, flexible and portable to all data stream-based domains. In particular, the project will go beyond the current state-of-art technology by making the following specific original contributions: i) Real-time scalable machine learning for massive, high-velocity and complex data streams analytics; ii) Real-time hybrid computation, batch data and data streams; iii) Real-time interactive visual analytics for Big Data; iv) Enhancement of Apache Flink, the EU Big Data platform; and v) Real-world industrial validation of the technology developed The PROTEUS impact is manifold: i) strategic, by reducing the gap and dependency from the US technology, empowering the EU Big Data industry through the enrichment of the EU platform Apache Flink; ii) economic, by fostering the development of new skills and new job positions and opportunities towards economic growth; iii) industrial, by considering real-world requirements from industry and by validating the outcome on an operational setting, and iv) scientific, by developing original hybrid and streaming analytic architectures that enable scalable online machine learning strategies and advanced interactive visualisation techniques that are applicable for general data streams in other domains.

4	Cloud-LSVA	Cloud Large Scale Video Analysis	http://cordis.europa.eu/project/rcn/199579_en.html	European Commission H2020 Call:ICT-16-2015	Cloud-LSVA will create Big Data Technologies to address the open problem of a lack of software tools, and hardware platforms, to annotate petabyte scale video datasets. The problem is of particular importance to the automotive industry. CMOS Image Sensors for Vehicles are the primary area of innovation for camera manufactures at present. They are the sensor that offers the most functionality for the price in a cost sensitive industry. By 2020 the typical mid-range car will have 10 cameras, be connected, and generate 10TB per day, without considering other sensors. Customer demand is for Advanced Driver Assistance Systems (ADAS) which are a step on the path to Autonomous Vehicles. The European automotive industry is the world leader and dominant in the market for ADAS. The technologies depend upon the analysis of video and other vehicle sensor data. Annotations of road traffic objects, events and scenes are critical for training and testing computer vision techniques that are the heart of modern ADAS and Navigation systems. Thus, building ADAS algorithms using machine learning techniques require annotated data sets. Human annotation is an expensive and error-prone task that has only been tackled on small scale to date. Currently no commercial tool exists that addresses the need for semi-automated annotation or that leverages the elasticity of Cloud computing in order to reduce the cost of the task. Providing this capability will establish a sustainable basis to drive forward automotive Big Data Technologies. Furthermore, the computer is set to become the central hub of a connected car and this provides the opportunity to investigate how these Big Data Technologies can be scaled to perform lightweight analysis on board, with results sent back to a Cloud Crowdsourcing platform, further reducing the complexity of the challenge faced by the Industry. Car manufacturers can then in turn cyclically update the ADAS and Mapping software on the vehicle benefiting the consumer.
5	HOBBIT	Holistic Benchmarking of Big Linked Data	http://cordis.europa.eu/project/rcn/199489_en.html	European Commission H2020 Call:ICT-16-2015	Linked Data has gained significant momentum over the last years. It is now used at industrial scale in many sectors in which an increasingly large amount of rapidly changing data needs to be processed. HOBBIT is an ambitious project that aims to push the development of Big Linked Data (BLD) processing solutions by providing a family of industry-relevant benchmarks for the BLD value chain through a generic evaluation platform. We aim to make open deterministic benchmarks available to test the performance of existing systems and push the development of innovative industry-relevant solutions. The underlying data will mimic real industrial data assembled during the course of the project. At the beginning of the project, HOBBIT will work on roughly 1PB of real industry-relevant data from 4 different domains. The data will be extended through collaborations during the project. To push the use of the benchmarks, we will organize or join challenges that aim to measure the performance of technologies for the different steps of the BLD lifecycle. In contrast to existing benchmarks, we will provide modular and easily extensible benchmarks for all industry-relevant BLD processing steps that allow to assess whole suites of software that cover more than one step. The infrastructure necessary to run the evaluation campaigns will be made available. Our architecture will rely on web interfaces and cloud infrastructures to ensure scalability. The open HOBBIT platform will make human- and machine-readable, public periodic reports available. As exit strategy, the project will create an association after the second project year that will be sustained by the means of subscriptions from industry and academia and

					associated with existing benchmarking associations. The clear portfolio of added value for the members will be defined in the early project stages and disseminated throughout the evaluation campaigns.
6	VaVeL	Variety, Veracity, VaLue: Handling the Multiplicity of Urban Sensors	http://cordis.europa.eu/project/rcn/199181_en.html	European Commission H2020 Call:ICT-16-2015	Urban environments are awash with data from fixed and mobile sensors and monitoring infrastructures from public, private, or industry sources. Making such data useful would enable developing novel big data applications to benefit the citizens of Europe in areas such as transportation, infrastructures, and crime prevention. Urban data is heterogeneous, noisy, and unlabeled, which severely reduces its usability. Succinctly stated, urban data are difficult to understand. The goal of the VaVel project is to radically advance our ability to use urban data in applications that can identify and address citizen needs and improve urban life. Our motivation comes from problems in urban transportation. This project will develop a general purpose framework for managing and mining multiple heterogeneous urban data streams for cities become more efficient, productive and resilient. The framework will be able to solve major issues that arise with urban transportation related data and are currently not dealt by existing stream management technologies. The project brings together two European cities that provide diverse large scale data of cross-country origin and real application needs, three major European companies in this space, and a strong group of researchers that have uniquely strong expertise in analyzing real-life urban data. VaVel aims at making fundamental advances in addressing the most critical inefficiencies of current (big) data management and stream frameworks to cope with emerging urban sensor data thus making European urban data more accessible and easy to use and enhancing European industries that use big data management and analytics. The consortium develops end-user driven concrete scenaria that are addressing real, important problems with the potential of enormous impact, and a large spectrum of technology requirements, thus enabling the realization of the fundamental capabilities required and the realistic evaluation of the success of our methods.
7	TOREADOR	TrustwOrthy model-awaRE Analytics Data platfORm	http://cordis.europa.eu/project/rcn/200253_en.html	European Commission H2020 Call:ICT-16-2015	The TOREADOR project is aimed at overcoming some major hurdles that until now have prevented many European companies from reaping the full benefits of Big Data analytics (BDA). Companies and organisations in Europe have become aware of the potential competitive advantage they could get by timely and accurate Big Data analytics, but lack the IT expertise and budget to fully exploit BDA. To overcome this hurdle, TOREADOR takes a model-based BDA-as-a-service (MBDAaaS) approach, providing models of the entire Big Data analysis process and of its artefacts. TOREADOR open, suitable-for-standardisation models will support substantial automation and commoditisation of Big Data analytics, while enabling it to be easily tailored to domain-specific customer requirements. Besides models for representing BDA, TOREADOR will deliver an architectural framework and a set of components for model-driven set-up and management of Big Data analytics processes. Once TOREADOR MBDAaaS will become widespread, price competition on Big Data services will ensue, driving costs of Big Data analytics well within reach of EU organizations (including SMEs) that do not have either in-house Big Data expertise or budget for expensive data consultancy. Activities supported and automatised by TOREADOR will include (i) planning Big Data sources preparation (ii)

					negotiating machine-readable Service Level Agreements for BDA detailing privacy, timing, and accuracy needs (iii) choosing data management and algorithm parallelisation strategies (iv) ensuring auditing and assessment of legal compliance (for example, to privacy regulations) of BDA enactment. TOREADOR framework will address automatically all major problems of on-demand data preparation, including handling Big Data opacity, diversity, security, and privacy compliance, and will support abstract modelling of the BDA life cycle from distributed data acquisition/storage to the design and parallel deployment of analytics and presentation of results.
8	SEE.4C	SpatiotEmporal ForEcasting: Coopetition to meet Current Cross-modal Challenges	http://cordis.europa.eu/project/rcn/199868_en.html	European Commission H2020 Call:ICT-16-2015	Fast, accurate forecasting of spatiotemporal data is needed in critical industrial domains such as energy (prediction of spatiotemporal patterns in renewable generation, usage and traffic) as well as in public policy. The task is so challenging in scale and scope however as to have been confined mainly to research, while past prize competitions have been limited to forecasts of single dimensional values. Building on our proven success in numerous prize-driven past data challenges, attracting hundreds of participants, we aim to compile and test data grounded on large-scale open European datasets and including specially prepared grid traffic data from Europe's largest Transportation System Operator. The competition evaluates forecasting algorithms on a cloud platform, tracking accuracy and computational efficiency. Emphasizing cross-specialization knowledge transfer and openness to novel technologies which may spring from different subsectors, we aim to build a platform allowing for coopetitions: the ad-hoc coalescence of competing teams during a challenge aimed at forming sustainable partnerships past the prize scheme itself. We will provide comprehensive documentation for a freely extensible open-source cloud-based specialized computing platform (assembling existing, well tested tools) allowing automated evaluation and feedback as in our latest competitions, but scaled to big data needs. We aim to test this platform and provide baseline results in a smaller scale mini-competition (hackathon). Thus we shall lay the groundwork for a larger prize competition in which evaluation data for predictions may arrive in real or near-real time. We also aim to use our wide contacts with industry, domain and data experts and past participants and winners in order to organize focused meetings of panels to refine value chains in data and algorithms as well as conference workshops, talks and newsletters dedicated to widely advertising challenges to past and new participants. Expand / Contract(-) Coordinator

9	SETA	An open, sustainable, ubiquitous data and service ecosystem for efficient, effective, safe, resilient mobility in metropolitan areas	http://cordis.europa.eu/project/rcn/199852_en.html	European Commission H2020 Call:ICT-16-2015	Around 50% of the global population lives in metropolitan areas, and this is expected to grow to 75% by 2050. Mobility within these areas is complex as it involves multiple modalities of transport, multiple managing authorities, as well as several millions of citizens. The cost of inefficiency in transport and mobility are enormous. For example, inefficiency costs the UK economy €5.8 billion each year. €583 million is wasted on fuel (e.g. traffic congestion) alone each year, which attributes to increased urban pollution and CO2. Hold-ups to business or freight vehicles amounts to €1.5bn annually. Mobility generates huge amounts of data thought thousands of sensors, city cameras, and connected cars, as well as millions of citizens connected through their mobile devices. If properly managed, this data can be used to understand, optimise and manage mobility and make it more efficient, sustainable and resilient. SETA will address this challenge, creating a technology and methodology able to use this wealth of data to change the way mobility is organised, monitored and planned in large metropolitan areas. The solution will be able to collect, process, link and fuse high-volume, high-velocity, multi-dimensional, heterogeneous, cross-media, cross-sectorial data and to use it to model mobility with a precision, granularity and dynamicity that is impossible with today's technologies. Such models will be the basis of pervasive services to citizens and business, as well as decision makers to support safe, sustainable, effective, efficient and resilient mobility. The project has the potential to impact the everyday lives of millions of people, their health and the environment with enormous financial and social impact. SETA's solution will be evaluated rigorously by citizens, business and decision makers in 3 cities across Europe. The proposal includes a commercialisation plan and describes the economy of managing the SETA ecosystem in a metropolitan area.
10	datACRON	Big Data Analytics for Time Critical Mobility Forecasting	http://cordis.europa.eu/project/rcn/199835_en.html	European Commission H2020 Call:ICT-16-2015	datACRON is a research and innovation collaborative project introducing novel methods for threat and abnormal activity detection in very large fleets of moving entities spread across large geographical areas.Specifically, datACRON aims to develop novel methods for real-time detection and prediction of trajectories and important events related to moving entities, together with advanced visual analytics methods, over multiple heterogeneous, voluminous, fluctuating, and noisy data streams from moving entities, correlating them with archived data expressing, among others, entities' characteristics, geographical information, mobility patterns, regulations and intentional data (e.g. planned routes), in a timely manner. Technological developments are validated and evaluated in user-defined challenges focusing on increasing the safety, efficiency and economy of operations concerning moving entities in the Air-Traffic Management and Maritime domains. The datACRON project brings together partners from academia and industry to develop the aforementioned novel methods, together with user and data-provision partners from the two domains, in close relation to user-interest groups, focusing on real-life, industrial and user-defined challenges concerning operations (e.g. surveillance, forecasting of trajectories, characterization, etc.) regarding moving entities in sea and air.

11	EuDEco	Modelling the European data economy	http://cordis.europa.eu/project/rcn/194301_en.html	European Commission H2020 Call:ICT-15-2014	The Modelling the European Data Economy (EuDEco) project will assist European science and industry in understanding and exploiting the potentials of data reuse in the context of Big and Open Data big data. The aim isto establish a self sustaining data market and thereby increase the competitiveness of Europe. To be able to extract the benefits of data reuse, it is crucial to know how to understand the underlying economic, societal, legal, and technological framework conditions and challenges to build useful applications and services. Despite the amount of activities in this domain, an effort is missing to develop use cases and business models that are economic viable, legally certain and taking societal needs and concerns into account. EuDEco will accomplish this by leveraging the engagement of other projects conducting pilots on data reuse as well as by the engagement of external experts and stakeholders. EuDEco moves beyond the classical approaches by applying the approach of complex adaptive systems to model the data economy in order to indentify value networks, use cases and business models for data reuse. In the course of the project will develop and refine the data economy model in several steps further by case studies on previous pilots on data reuse, by in-depth analysis from legal, socioeconomic and technological point of view, and by extensive tests of use cases and business models with other projects. Therefore it will analyse framework conditions relevant and challenges related to data reuse and the emergence of a self-sustaining data market. Finally, EuDEco will deliver a model of the data economy including viable use cases and business models as well as suggestions and recommendations addressing the main legal, contractual, societal, and technological concerns and challenges such as contractual framework or data protection. Above that EuDEco will develop an observatory for policy makers enabling them to track the development of the data economy.
12	SSIX	Social Sentiment analysis financial IndeXes	http://cordis.europa.eu/project/rcn/194322_en.html	European Commission H2020 Call:ICT-15-2014	Social Sentiment Indices powered by X-Scores (SSIX) aims to provide European SMEs with a collection of easy to interpret tools to analyse and understand social media users attitudes for any given subject; these sentiment characteristics can be exploited to help SMEs to operate more efficiently resulting in increased revenues. Social media data represents a collective barometer of thoughts and ideas touching every facet of society. SSIX will search and index conversations taking place on social network services, such as Twitter, StockTwits, and Facebook including the most reliable and authoritative Newswires, online newspapers, trade publications and blogs. SSIX will classify and score content using a framework of qualitative and quantitative parameters called X-Scores, regardless of language, locale or data architecture. The X-Scores framework will interpret economically significant sentiment signals in social media conversations producing sentiment metrics, such as momentum, awareness scope, topic density and historical comparison. These metrics will create commercially viable social sentiment indices, which can be tailored to any domain of interest. By enabling European SMEs to analyse and leverage social sentiment in their discipline, SSIX will facilitate the creation of innovative products and services by enhancing the investment decision making process, thus increasing revenue while also minimising risk.

13	AEGLE	An analytics framework for integrated and personalized healthcare services in Europe	http://cordis.europa.eu/project/rcn/194261_en.html	European Commission H2020 Call:ICT-15-2014	The data generated in the health domain is coming from heterogeneous, multi-modal, multi-lingual, dynamic and fast evolving medical technologies. Today we are found in a big health landscape characterized by large volume, versatility and velocity (3Vs) which has led to the evolution of the informatics in the big biodata domain. AEGLE project will build an innovative ICT solution addressing the whole data value chain for health based on: cloud computing enabling dynamic resource allocation, HPC infrastructures for computational acceleration and advanced visualization techniques. AEGLE will: - Realize a multiparametric platform using algorithms for analysing big biodata including features such as volume properties, communication metrics and bottlenecks, estimation of related computational resources needed, handling data versatility and managing velocity - Address the systemic health big bio-data in terms of the 3V multidimensional space, using analytics based on PCA techniques - Demonstrate AEGLE's efficiency through the provision of aggregated services covering the 3V space of big bio-data. Specifically it will be evaluated in: a)big biostreams where the decision speed is critical and needs non-linear and multi-parametric estimators for clinical decision support within limited time, b)big-data from non-malignant diseases where the need for NGS and molecular data analytics requires the combination of cloud located resources, coupled with local demands for data and visualization, and finally c)big-data from chronic diseases including EHRs and medication, with needs for quantified estimates of important clinical parameters, semantics' extraction and regulatory issues for integrated care. - Bring together all related stakeholders, leading to integration with existing open databases, increasing the speed of AEGLE adaptation - Build a business ecosystem for the wider exploitation and targeting on cross-border production of custom multi-lingual solutions based on AEGLE.
14	FREME	Open Framework of E-Services for Multilingual and Semantic Enrichment of Digital Content	http://cordis.europa.eu/project/rcn/194243_en.html	European Commission H2020 Call:ICT-15-2014	The aim of the FREME innovative action is to establish an "Open Framework of E-Services for Multilingual and Semantic Enrichment of Digital Content". Six enrichment services will be designed, piloted, and validated during the action. Their innovation, usability, and robustness will be shaped by four real world business cases that will bring FREME data innovation and technology transfer directly to the market: (1) authoring and publishing multilingually and semantically enriched eBooks; (2) integrating semantic enrichment into multilingual content during localisation; (3) enhancing cross-language sharing and access to open data and (4) empowering personalised content recommendations. FREME innovation will lead to new business models for digital content and Big Data markets and will contribute to EU competiveness and new job profiles in the digital content management realm. FREME addresses specific challenges of topic ICT 15-2014: 1) FREME will improve the ability of EU companies to build innovative multilingual data products and services by providing six innovative, user-friendly, and robust enrichment services that are accessible both via APIs and GUIs; 2) FREME will remove technological barriers for multilingual content technologies by allowing to use data assets in an interoperable, reusable, and aggregatable

					way across sector, borders, and languages; 3) FREME will transfer existing mature multilingual and semantic technologies and cloud-based infrastructures previously developed by partners and available on their 7, 8, or 9 technology readiness level for action as value-adding components to content and data value chains during; 4) FREME business cases will validate the action concept, approach, and ambition within the entire content and data value chain. The core of the FREME consortium are companies who define business cases, partners who shape data innovation and technology transfer including their expertise in market validation and business modelling and planning.
15	MixedEmotions	Social Semantic Emotion Analysis for Innovative Multilingual Big Data Analytics Markets	http://cordis.europa.eu/project/rcn/194226_en.html	European Commission H2020 Call:ICT-15-2014	MixedEmotions will develop innovative multilingual multi-modal Big Data analytics applications that will analyze a more complete emotional profile of user behavior using data from mixed input channels: multilingual text data sources, A/V signal input (multilingual speech, audio, video), social media (social network, comments), and structured data. Commercial applications (implemented as pilot projects) will be in Social TV, Brand Reputation Management and Call Centre Operations. Making sense of accumulated user interaction from different data sources, modalities and languages is challenging and has not yet been explored in fullness in an industrial context. Commercial solutions exist but do not address the multilingual aspect in a robust and large-scale setting and do not scale up to huge data volumes that need to be processed, or the integration of emotion analysis observations across data sources and/or modalities on a meaningful level. MixedEmotions will implement an integrated Big Linked Data platform for emotion analysis across heterogeneous data sources, different languages and modalities, building on existing state of the art tools, services and approaches that will enable the tracking of emotional aspects of user interaction and feedback on an entity level. The MixedEmotions platform will provide an integrated solution for: large-scale emotion analysis and fusion on heterogeneous, multilingual, text, speech, video and social media data streams, leveraging open access and proprietary data sources, and exploiting social context by leveraging social network graphs; semantic-level emotion information aggregation and integration through robust extraction of social semantic knowledge graphs for emotion analysis along multidimensional clusters.
16	ODINE	Open Data INcubator for Europe	http://cordis.europa.eu/project/rcn/194234_en.html	European Commission H2020 Call:ICT-15-2014	The Open Data INcubator for Europe (ODINE) project will set up an environment to support and advice SMEs and start-ups in creating commercial added value from open data. Drawing on the experience from key players in the consortium including Wayra (an incubator/accelerator), Fraunhofer IAIS and the ODI (both data facilitators and incubators), Telefonica (providing computing infrastructure and strong links to data protection and privacy stakeholders), and the University of Southampton (renowned for its open data research and home of Sir Tim Berners-Lee and Sir Nigel Shadbolt), we plan to establish a EU-wide, industry-focused network of open data companies around Europe, promoted and guided by the trusted authorities and commentators at the Guardian and

					OKFN. To achieve this we shall take an approach that follows 'best in class' practice for incubators and which makes it as a simple and smooth as possible for SMEs to apply for help. We will offer: (i) a transparent, fair, and efficient proposal process for SMEs and start-ups, based around Wayra, IAIS and ODI's existing and successful processes; (ii) a network of European open data SMEs and start-ups; (iii) initial investment in these companies through an open call; (iv) incubation and mentoring by ODI, Fraunhofer and Wayra academies, with the possibility of acceleration via Wayra, Telefonica Ventures, and Fraunhofer Venture; (v) access to data services and computing infrastructure from IAIS, ODI, OKFN and Telefonica; (vi) expert support from serial entrepreneurs and intrapreneurs, and coaching from business schools; as well as (vii) high visibility through promotion of success stories by the Guardian, ODI and the OKFN. A successful open data incubator will significantly help Europe to regain leadership in technical innovation, balanced with the ethical treatment of data for which the public are crying out, and to create skilled jobs that help Europe become more productive and competitive in the 21st Century.
17	BigDataEurope	Integrating Big Data, Software and Communities for Addressing Europe's Societal Challenges	http://cordis.europa.eu/project/rcn/194216_en.html	European Commission H2020 Call:ICT-15-2014	BigDataEurope will provide support mechanisms for all the major aspects of a data value chain, in terms of the employed data and technology assets, the participating roles and the established or evolving processes. The effectiveness of the provided support mechanisms will be assessed in different domains pertaining to Europe's major societal challenges with respect to the needs and requirements of the related communities. To this end, BigDataEurope focuses on providing an integrated stack of tools to manipulate, publish and use large-scale data resources; tools that can be installed and used freely in a customised data processing chain with minimal knowledge of the technologies involved and integrating and industrially hardening key open-source Big Data technologies and European research prototypes into a Big Data Integrator Platform, i.e. an ecosystem of specifications and reference implementations that are both attractive to current players from all parts of the data value chain while also lowering the entry barrier for new businesses. In order to realise its objectives, Big Data Europe will focus on two clearly defined coordination and support measures: 1. Coordination: Engaging with a diverse range of stakeholder groups representing particularly the Horizon 2020 societal challenges Health, Food & Agriculture, Energy, Transport, Climate, Social Sciences and Security; Collecting requirements for the ICT infrastructure needed by data-intensive science practitioners tackling a wide range of societal challenges; covering all aspects of publishing and consuming semantically interoperable, large-scale data and knowledge assets; 2. Support: Designing, realizing and evaluating a Big Data Aggregator platform infrastructure that meets requirements, minimises the disruption to current workflows, and maximises the opportunities to take advantage of the latest European RTD developments, including multilingual data harvesting, data analytics, and data visualisation. BigDataEurope will implement and apply two main instruments to successfully realize these coordination and support measures: a) Build Societal Big Data Interest Groups in the W3C interest group scheme and involving a large number of

					stakeholders from the Horizon 2020 societal challenges as well as technical Big Data experts; b) Design, integrate and deploy a cloud-deployment-ready Big Data aggregator platform comprising key open-source Big Data technologies for real-time and batch processing, such as Hadoop, Cassandra and Storm. BigDataEurope aims to provide an adaptable, easy to deploy and use solution, which will allow the interest-ed user groups and stakeholders to extend their Big Data solutions or introduce Big Data technology to their business processes, based on a concrete methodology for producing a technically sound solution and maximizing its outreach to the relevant communities.
18	proDataMarket	Enabling the property Data Marketplace for Novel Data-driven Business Models	http://cordis.europa.eu/project/rcn/194214_en.html	European Commission H2020 Call:ICT-15-2014	Property data are one of the most valuable datasets managed by governments worldwide and extensively used in various domains by private and public organizations. Unfortunately these data are not always easy to access. House and property data is used in a variety of ways to produce value added information within and across several business sectors, including real estate and debt collection. Such sectors suffer from a lack of innovation due to a fragmented data ecosystem which makes it difficult to access relevant datasets. This hampers innovation, protects incumbents and promotes rent-seeking business models. The difficulty in creating a single, open data market partly depends on the fact that some governmental agencies are currently making significant revenues on selling data to a restricted number of business players in the private sector. However, several studies, demonstrated that the transaction costs for government agencies tend to be very high, and often make selling the data non-profitable. proDataMarket aims to disrupt the property data market and demonstrate innovation across sectors where property data is relevant, by integrating a technical framework for effective publishing and consumption of property-related data, showcasing novel data-driven business products and services based on property data. proDataMarket will provide a digital data marketplace for open and non-open property and related contextual data, making it easier for data providers to publish and distribute their data (for free or for a fee) and for data consumers to easily access the data they need for their businesses. The consortium is formed by large companies ensuring high-impact business cases, technology transfer providers ensuring technologies for the creation and maintenance of the data market platform, and large data providers providing data for the business cases. With strong industry involvement proDataMarket will cover a wide range of data value chains related to property data.
19	AquaSmart	Aquaculture Smart and Open Data Analytics as a Service	http://cordis.europa.eu/project/rcn/194237_en.html	European Commission H2020 Call:ICT-15-2014	AQUASMART's objective is to enhance innovation capacity to the aquaculture sector, by addressing the problem of global knowledge access and data exchanges between aquaculture companies and its related stakeholders. Offering aquaculture production companies the tools to access and share global open data and strong data analytics in a multi-lingual, multi-sector and cross-border setting strengthens their competitiveness and growth potential. Experienced research institutes that participate in the consortium as technology suppliers and will transfer their solutions to the aquaculture stakeholders in the consortium. The data collected in the AQUASMART open data cloud is suitable to be reused in other industrial domain if needed, (e.g., environmental or transportation data), providing a cross-sectorial setting to the provided solution. The AQUASMART multi-lingual adaptive eTraining

					program, assures that staff receive the proper training and assures the transfer of the AQUASMART innovations are sustainable. AQUASMART will have a very positive impact on the environment by helping companies to better estimate daily biomass, optimize feeding rates and management practices. This will improve the FCR (Feed Conversion Rate), which means less feed will be provided to the fish and therefore, less organic material and energy are consumed for the production of the feed. AQUASMART also helps the companies to reduce mortalities, which will have a further positive impact on environment. On the social level, AQUASMART contributes to the development of highly skilled workforce through online training programs. The improvement of the efficiency and profitability of the businesses, together with the reduction of the environmental impact will contribute to the increase of the production and the generation of new jobs in the sector.
20	BISON	BIg Speech data analytics for cONtact centres	http://cordis.europa.eu/project/rcn/194308_en.html	European Commission H2020 Call:ICT-15-2014	Contact centers (CC) are an important business for Europe: 35,000 contact centers generate 3.2 Million jobs (~1% of Europe's active population). A typical CC produces a wealth of multilingual spoken data that is nowadays mined by humans (CC agents and supervisors) or by rudimentary technical means. BISON consortium plans to bring significant innovations in three areas: (1) basic speech data mining technologies (systems quickly adaptable to new languages, domains and CC campaigns), (2) business outcome mining from speech (translated into improvement of CCs' Key Performance Indicators) and (3) CC support systems integrating both speech and business outcome mining in user-friendly way. The project will produce two prototypes: smallBison (end of the 1st year) will be a functioning system for real, though limited, deployment and user feedback collection. bigBison (end of the project) will include full range of capabilities and be fully integrated with CC hardware and software infrastructure. Generation of business outputs will be demonstrated on real data. Business indicators and values for the market were instrumental for the definition of the project and will be crucial for project execution. BISON consortium is composed of eight players with complementary skills. Two end users running large CC operations (EBOS, Atento) are generating user requirements and are ready to deploy the prototypes immediately in real scenarios. Phonexia (the coordinator), Brno University of Technology and Telefónica I+D are experts in speech data mining - from R&D, data processing to developing products placed on the market. Telefónica Móviles is an expert in business outcome mining and MyForce is a skilled Contact Center hardware and software integrator. CC data involve a number of legal issues, therefore, the University of Bologna (with significant experience in regulatory and legal aspects) complements the consortium.

21	KConnect	Khresmoi Multilingual Medical Text Analysis, Search and Machine Translation Connected in a Thriving Data-Value Chain	http://cordis.europa.eu/project/rcn/194242_en.html	European Commission H2020 Call:ICT-15-2014	The overall objective of the KConnect project is to create a medical text Data-Value Chain with a critical mass of participating companies using cutting-edge commercial cloud-based services for multilingual Semantic Annotation, Semantic Search and Machine Translation of Electronic Health Records and medical publications. The commercial cloud-based services will be the result of productisation of the multilingual medical text processing tools developed in the Khresmoi FP7 project, allowing wide adoption of these tools by industry. The critical mass will be created by the KConnect Professional Services Community, which will consist of at least 30 companies by the end of the project. These companies will be trained to build solutions based on the KConnect Services, hence serving as multipliers for commercial exploitation of the KConnect services. The KConnect project will facilitate the straightforward adaptation of the commercialised services to new languages by providing toolkits enabling the adaptation to be done by by people having a software engineering skillset, as opposed to the rarer language engineering skillset. The KConnect services will also be adapted to handle text in Electronic Health Records, which is particularly challenging due to misspellings, neologisms, organisation-specific acronyms, and heavy use of negation and hedging. The consortium is driven by a core group of four innovative SMEs following complementary business perspectives related to medical text analysis and search. These companies will build solutions for their customers based on KConnect technology. Two partners from the medical domain will use KConnect services to solve their medical record analysis challenges. Two highly-used medical search portal providers will implement the KConnect services to innovate the services offered by their search portals. Through these search portals, the KConnect technologies will be used by over 1 million European citizens before the end of the project
22	EDSA	European Data Science Academy	http://cordis.europa.eu/project/rcn/194116_en.html	European Commission H2020 Call:ICT-15-2014	Data explosion on the web, fuelled by social networking, micro-blogging, as well as crowdsourcing, has led to the Big Data phenomenon. This is characterized by increasing volumes of structured, semi-structured and unstructured data, originating from sources that generate them at an increasing rate. This wealth of data provides numerous new analytic and business intelligence opportunities to various industry sectors. Therefore, more and more industry sectors are in need of innovative data management services, creating a demand for Data Scientists possessing skills and detailed knowledge in this area. Ensuring the availability of such expertise will prove crucial if businesses are to reap the full benefits of these advanced data management technologies, and the know-how accumulated over the past years by researchers, technology enthusiasts and early adopters. The European Data Science Academy (EDSA) will establish a virtuous learning production cycle whereby we: a) analyse the required sector specific skillsets for data analysts across the main industrial sectors in Europe; b) develop modular and adaptable data science curricula to meet these needs; and c) deliver training supported by multiplatform and multilingual learning resources based on our curricula. The curricula and learning resources will be continuously evaluated by pedagogical and data science experts during both development and deployment.

23	AutoMat	Automotive Big Data Marketplace for Innovative Cross-sectorial Vehicle Data Services	http://cordis.europa.eu/project/rcn/194228_en.html	European Commission H2020 Call:ICT-15-2014	Inside today's vehicles ~4000 CAN-Bus signals/sec are processed in comparison to very few signals in smart phones and alike. This large amount of continuously gathered vehicle data represents major big data business potentials, not only for the automotive industry but in particular for cross-sectorial industries with interdisciplinary applications. With today's proprietary approaches focusing on bringing services into vehicles and the applied ignorance of customer privacy concerns, this major business potential is still locked because the automotive industry was not yet able to establish an open service ecosystem equivalent to the ones in the smart phone industry. The core intention of the AutoMat project is to innovate an open ecosystem for Vehicle Big Data, materializing in the form of a cross-border Vehicle Big Data Marketplace that leverages currently unused information gathered from connected vehicles. The interface to the marketplace is derived from a Common Vehicle Information Model that makes mined and anonymous vehicle data from various OEMs accessible to cross-sectorial service providers. With the huge amount of volatile data from vehicles, the AutoMat ecosystem heavily builds upon current trends in Big Data. Exemplary service scenarios, driven by service providers dedicated to generate concrete businesses from the AutoMat ecosystem, are developed in the context of meteorological data based hyper local and extended innovative enterprise service domains. By defining an open value chain, the proposed AutoMat ecosystem enables and stimulates parties from different sectors to focus on their core businesses and to excel collaboration with other partners. AutoMat therefore may serve as incubator for new business opportunities strengthening Europe's position as provider of innovative cross-sectorial and cross-border Big Data services. The latter aspect is actively stimulated during the project by an Open Service Contest based upon the AutoMat Big Data ecosystem.
24	EUBra-BIGSEA	EUrope-BRAzil Collaboration on BIG Data Scientific REsearch through Cloud-Centric Applications	http://cordis.europa.eu/project/rcn/198830_en.html	European Commission H2020 Call:EUB-1-2015	EUrope-BRAzil Collaboration on BIG Data Scientific REsearch through Cloud-Centric Applications aims at providing services in the cloud for the processing of massive data coming from highly connected societies, which impose multiple challenges on resource provision, performance, Quality of Service and privacy. Processing those data require rapidly provisioned infrastructures customised to Big Data requirements. The three main aims of the proposal will be: - The development of innovative Big Data services for capturing, federating and annotating on the order of PB of data on top of efficient programming models. Despite that MapReduce is a successful model in BigData (with a high impact on massive Geo-spatial and textual data), it has many limitations specially when dealing with real-time transactions or streamed data, the proposal would aim to introduce innovative evolutions on the capture, federation & annotation experience it can bring to the table with its partners. - The Development of advanced cloud services to support Big Data. This cloud services will address three main challenges: a) the advance on SLAs to support privacy (boundaries of protected data to be moved) and performance restrictions (convenience of moving data to computing resources or vice-versa); b) Quality of Service (vertical and horizontal elastic adjustment of resources allocates to meet deadlines and dynamic adjustment of workloads); and c) Business models (price-based dynamic re-scheduling of data searching for the best usage of resources invested).

					- The demonstration of such services on applications with high social and business impact, addressing main scenarios of high interest for both Europe and Brazil.
25	BigData4Cat	Big Data for Catalysis	http://cordis.europa.eu/project/rcn/198780_en.html	European Commission H2020 Call:ERC-PoC-2015	Catalysis is one of the scientific areas in which Europe has a leading position. The radical change in the use of raw materials from oil towards gas or biomass might compromise this position. Computational techniques have been identified as the third pillar in catalysis research and provide a great amount of data that can speed up the generation of new catalytic systems through rational design. Industries are now starting to focus on the large amount of data published in the open literature regarding mechanistic studies so that they can accelerate their discovering of new catalysts. However, the unstructured and unlinked nature of this information hinders a fast transference of published knowledge to the chemical industry. Our BigData4Cat proof of concept would generate a simple, unified platform: ioChem-BD, where all the data regarding atomistic theoretical simulations in catalysis could be stored and retrieved in a structured manner. The platform will highlight the links, establish the relationships between data from different sources, provide error bars, and allow inferring data from missing steps in complex reaction networks. Moreover, it will provide problem-targeted structured databases with data-mining options. The final goal to the project is to transfer the mature computational Chemistry methodology and data into growing research strategies through the ioChem-BD platform. The goal of the proof-of-concept will be to store, structure and search the Catalysis Big Data resources in a sustainable manner that can be adapted to different problems at academic, editorial and industrial levels.
26	GDC	A Genetic Data CUBE - An innovative business model applied to predictive and prescriptive analytics, exploring Big Data and empowering cloud-services and urgent computation	http://cordis.europa.eu/project/rcn/198412_en.html	European Commission H2020 Call:INSO-10-2015-1	Business Intelligence, Decision Making Processing, Predictive Analysis for economic and financial forecasting represent today a brick and mortar set of tools and methodologies to support important business functions aiming at investing purposes, storing customer information, marketing analysis and related tracking activities. Predictive analysis is therefore of paramount importance to diagnose and solve business problems and issues in our VUCA world (dominated by Volatility, Uncertainty, Complexity, Ambiguity). To meet this EU as well as global challenge, we intend to bring Genetic Data Cube (GDC) to market maturity. GDC is a data driven innovative business model powered by a huge trove of real-world, real-time preference/usage /feedback data coming from experienced business experts as well as end- users. Our goal is to architect, develop, and deploy world-class data systems to empower multiple services and facilitate SMEs as well large enterprises or industries in diagnosing and solving relevant business problems or challenges. We are introducing a new paradigm in predictive analysis, a step-change to “prescriptive” analysis, not only processing-based but also human-centred. Considering that Big Data are defined as the “oil of our times”, Big Data Analytic stands at the forefront to data management for economic and financial forecasting: to have a crystal-clear picture of the business opportunity, 90% of the data in the world has been created in the last 2 years alone; present estimates project that the Big Data market will top $84B in 2026. GDC will provide a unique, brand new business model supporting customers in understanding the past unstructured data to predict future activities in customers, investments and business

					development/growth perspective. Main benefits rely on improving quick-decision, profit, measure social media impact and prevents fraud by making application of advanced analytics and decision optimisation.
27	BIGCHEM	Big Data in Chemistry	http://cordis.europa.eu/project/rcn/198330_en.html	European Commission H2020 Call:MSCA-ITN-2015-EID	The advent of the big data era in chemistry and the life sciences requires the development of new computational analysis methods, which are not only of scientific, but also economic relevance. Currently, the international data market already grows six times faster than the entire IT sector, and growth rates further increase. Achieving and sustaining a leadership positions in the big data arena represent critically important challenges for the EU. The economic developments in the emerging big data field are science-driven. Due to complexity and heterogeneity of biochemical and biomedical data, large-scale data exploration and exploitation are intrinsically interdisciplinary tasks. BIGCHEM positions itself at interfaces between chemistry, computer science, and the life science to provide well-structured multidisciplinary training and educate high-in-demand computational specialists capable of operating in interdisciplinary and international research and business settings. Cornerstones of BIGCHEM's curriculum include on-line lectures and periodic schools taught by internationally leading experts in chemical and life science informatics, a balanced consortium of academia, SMEs, and large industry, and an unprecedented symbiosis of academic and industrial training and application components. Accordingly, BIGCHEM is well positioned to boost multilateral collaborations between academia and industry and train scientists who are highly competitive in the international big data market. In BIGCHEM's R&D and training activities, the development and evaluation of conceptually novel methods for large-scale data analysis, knowledge extraction, and information sharing with demonstrated practical application potential take center stage. The network has a clearly defined policy for exploitation of new IP through wide involvement of target users, SMEs, and large industry facilitated by the experienced technology transfer department of the coordinator's team.
28	BigDataFinance	Training for Big Data in Financial Research and Risk Management	http://cordis.europa.eu/project/rcn/198268_en.html	European Commission H2020 Call:MSCA-ITN-2015-ETN	BigDataFinance, a Marie Skłodowska-Curie Innovative Training Network "Training for Big Data in Financial Research and Risk Management", provides doctoral training in sophisticated data-driven risk management and research at the crossroads of Finance and Big Data for 13 researchers. The main objectives are i) to meet an increasing commercial demand for well-trained researchers experienced in both Big Data techniques and Finance and ii) to develop and implement new quantitative and econometric methods for empirical finance and risk management with large and complex datasets. To achieve the objectives, the emphasis is put on exploiting big data techniques to manage and use datasets that are too large and complex to process with conventional methods. Banks and other financial institutions must be able to manage, process, and use massive heterogeneous data sets in a fast and robust manner for successful risk management; nonetheless, financial research and training has been slow to address the data revolution. Compared to the USA, Europe is still at an early stage of adopting Big Data technologies and services. Immediate action is required to seize opportunities to exploit the huge potential of Big Data within the European financial world. This world-class network consists of eight academic participants and six companies, representing banks, asset management companies, and data and solution providers. The proposed research is relevant both academically and

					practically, because the program is built around real challenges faced both by the academic and private sector partners. To bridge research and practice, all researchers contribute to the private sector via secondments. BigDataFinance provides the European financial community with specialists with state-of-the-art skills in finance and data-analysis to facilitate the adoption of reliable and realistic methods in the industry. This increases the financial strength of banks and other financial institutions in Europe.
29	BYNSE	BUSINESS INTELLIGENCE SERVICE FOR THE MANAGEMENT OF CROPS BASED ON CLOUD & BIG DATA	http://cordis.europa.eu/project/rcn/197154_en.html	European Commission H2020 Call:SFS-08-2014-1	Agro-industrial companies face the necessity of acquiring more information and control over their crops and the operations they perform in order to produce them. The challenge lies in reducing the hydric and energy resources they use with the objective of increasing their competitiveness and reducing the impact on the environment. Under the brand name of bynse, our SME cubenube (www.cubenube.com) has developed the first big data solution for agriculture in the world. bynse allows managers and producers to analyse the current and future needs of their crops in real time, using an innovative and inexpensive microclimatic sensing system called bynsebox for collecting field data, and a cloud information service called bynsecloud for processing and displaying the information to the users. This data is intersected with data from weather forecasts and vegetative growth tasks (pruning, seeding, etc), and a number of recommendations are generated, leading to a 30% reduction in energy and water consumption, and up to 40% reduction in the use and subsequent cost of phytosanitary products. Other information, such as strength and vigour of the plants, cannot be extracted from sensors and require colour areal maps (providing data through chlorophyll levels). The combined information provided by these maps plus bynse sensor information has great market potential for our target market of owners and managers of high added-value crops with extensions of between 20 and 100 hectares of non-intensive, woody or extensive agriculture. The main goal of this innovation project is to evolve bynse offering a “universal” solution by linking big data to geographical data to provide a full Precision Agriculture solution by adding the capability of processing colour maps and geo-localization to the overall information analysis and recommendations.

30	EarthServer-2	Agile Analytics on Big Data Cubes	http://cordis.europa.eu/project/rcn/196704_en.html	European Commission H2020 Call:EINFRA-1-2014	EarthServer-2 makes Agile Analytics on Big Earth Data Cubes of sensor, image, simulation, and statistics data a commodity for non-experts and experts alike through • navigation, extraction, aggregation, and recombining of any-size space/time data cubes; • easy to install & maintain value-adding services extending the existing portfolio of data and compute centers; • based on open standards, in particular: the OGC Big Data standards and the forthcoming ISO SQL/MDA (“Multi-Dimensional Arrays”) standard. In the Joint Research Activity, the project will advance the existing, world-leading rasdaman Array Database technology wrt. query functionality, inter-federation data processing with automatic data and query distribution, tape archive integration, and 3D/4D visualization based on NASA’s virtual globe technology. In the Services Activity, large data centers (ECMWF, PML, MEEO/ESA, GeoScience Australia, JacobsUni) will set up water, air, weather, and planetary services on 3D & 4D data cubes up to Petabyte-size with user-tailored clients for both visual and textual ad-hoc mix&match. In the Networking Activity, the project will advance open Big Data standards in OGC, RDA, and ISO (in particular: write ISO SQL/MDA). Further, all adequate channels will be used for strong dissemination & exploitation, specifically: writing a monograph explaining OGC Big Geo Data standards; scientific publications & active conference organization; Earth science data user workshops for each domain addressed; actively contributing technology & experience to GEO / GEOSS and further bodies; establish standardized Big Geo Data benchmark and run it against EarthServer-2 and further relevant systems. Altogether, EarthServer-2 will maintain and extend the lead in Big Earth Data services established in the highly successful EarthServer-1 project. Being already supported by ESA, rasdaman will form an enabling building block for COPERNICUS / Sentinel.
31	BDoSM	Big Data of Small Molecules	http://cordis.europa.eu/project/rcn/196524_en.html	European Commission H2020 Call:BIOTEC-5a-2014-1	As Google mapped and photographed the streets and terrain of our planet, the main objective of the project is to map and spectrally fingerprint the small organic molecules of various organisms and introduce to the market a fundamentally new cloud driven compound identification platform. Our pioneering solution will allow continuous identity assignment of unknown molecules in biologically relevant samples by developing a technologically mature architecture and infrastructure for storage, processing, search, exchange, analysis, data mining and visualization of millions of spectral, structural, chromatographic and biochemical data of small molecules. Our project will become the world most comprehensive database platform of spectral fingerprints of endogenous metabolites, natural products and environmental contaminants that will allow identification of substances present in biological samples that could not be identified using conventional technology that is a crucial bottleneck in many biotechnological and pharmaceutical applications. Since the biomedical, pharmaceutical and biotechnology industry in general are constantly detecting numerous unknown compounds and searching for new biomarkers and disease specific targets, our technology platform is direct response to their needs with the prospect of new business opportunities and future growth. The feasibility study will be important for acquiring additional financing of the project, to review technological

					solution (most effective and sustainable, without infringing patents, copyrights, and trademarks of third parties), to precise marketing strategy for corporate, small group & individual users, to identify, clarify and protect own inventions, new approaches and new technical solutions. Finally, the Phase 1 outputs will give perfect review on project management plan.
32	BDuSSCR	BIG DATA using SECOND SCREEN AND CONTENT RECOGNITION (BDuSSCR)	https://ec.europa.eu/easme/en/sme/5086/big-data-using-second-screen-and-content-recognition-bdusscr	European Commission H2020 Call:H2020-SMEINST-1-2014	TV Viewers demand a personalized and richer experience while watching TV. Broadcast TV and producers try to develop interactive products to build a stronger engagement with their audiences. Advertisers are tired of ad strategies that are not measurable nor results?oriented. Most of the broadcasting agents (TV/radio broadcasters, content producers, advertisers and their agencies) have poor information about the current behavior of their audiences. To improve the quality of such information they need to incorporate, on a personalized way, Big Data tools that help develop new strategies into their business activities. Bridge Mediatech, S.L. (from now on BMT), a Spanish technology company, has developed a Cloud-based Platform embedding its patented Automatic Content Recognition (ACR) technology in order to provide a richer experience to the audience. It also provides the broadcasting players with an easy to “install and operate” Big Data Platform to improve the engagement with their audience. It helps getting to know them better and it allows to offer customized contents tailored to their needs product and services. After having proved the core of the technology in the world´s largest Spanish speaking broadcasting player like Televisa, it´s time to start thinking about large scale solution deployment and international expansion. For that reason BMT applies to this call to develop the Feasibility study needed to accomplish the technology consolidation and large scale deployment which will let BMT become the global leader in this emerging and innovative marketplace
33	TORUS	HIGH PERFORMANCE COMMUNICATIONS FOR FINANCIAL SERVICES AND BIG DATA	http://www.torusware.com/	European Commission H2020 Call:ICT-37-2014-1	The objective of this proposal is to receive funding and support for commercializing a set of disruptive high-speed communication protocols (IPR protected by patent and trade secret) which bypass the legacy and inefficient TCP/IP, the Internet protocol, designed 40 years ago. The technology is the result of 10 year R&D at University of A Corunna and the successful outcome of a Proof-of-Concept project (2011-2014) funded by Fundacion Barrie (0.5 M investment), which led to incorporate TORUS in April 2013 to further develop, commercialize and support the technology. TORUS has 3 products in the market (Java Fast Sockets, Universal Fast Sockets and FastMPJ) which are up to 40 times faster than TCP/IP. These products reduce significantly the time-to-results of critical software, providing significant competitive advantages in multiple sectors, particularly for financial services, real time Big Data analytics (with multiple applications, such as real time online advertising), biotech (supporting DNA processing in hours, instead of weeks), and defense/space (FastMPJ has reduced up to 94% the time to distribute satellite images of ESAs Gaia Mission). Additionally, the technology reduces costs (power and software licensing)

					significantly, as fewer servers with TORUS outperform a higher number of servers interconnected via TCP/IP. The specific objectives of the feasibility study are: (1) acquire knowledge on how clients perceive TORUS value; (2) develop an effective go-to-market strategy; and (3) partner with key industry players (TORUS is in talks with two multibillion companies listed in NASDAQ). TORUS is about to rise 1M for funding next commercialization steps during 2014-2015. A new round of 4M is expected in 18-24 months, to definitively enter in the growth stage. SME Instrument would definitively fuel this high-tech disruptive European project. Our mission is to make servers respond faster than the speed of thought, no more waiting for computers!
34	BIDAMARK	BIg DAta-driven MARKeting for better competitiveness	http://cordis.europa.eu/project/rcn/195669_en.html	European Commission H2020 Call:MSCA-IF-2014-EF	Big data analysis has become a source of competitive advantage for businesses and even whole economies. Big data is indeed a critical issue for the digital economy, innovation and services that are expected to drive growth and job creation in the EU. Compared to the USA, however, the EU has been slow in the adoption of big data analysis. Fortunately, big data opportunities still exist in many business sectors to improve the competitiveness of European firms. In particular, performance may be improved by making use of big customer data to offer goods and services that better meet customer needs, i.e., “big data-driven marketing (BDM)”. However, adopting BDM has a steep learning curve due to organizations’ lack of understanding regarding the diverse factors required to succeed. The BIDAMARK project aims to assess the impact of different success factors on BDM, and consequently on firms’ ability to achieve superior customer performance. To answer this question, a comparative analysis between France- and US-based firms, and between industry sectors, will be implemented. A quantitative survey methodology will be applied with structural equation modelling (PLS) as the primary analysis method. This project is a perfect springboard for the Principal Investigator (PI) through the opportunity to build durable research networks, and acquire new skills that are complementary with his current expertise. New skills include research skills (theoretical and methodological), technical skills (use of new software tools), project management skills (lead of interdisciplinary project), language skills (French), and training skills (course teaching). Moreover, the publication of conference papers (later developed into high-impact journal articles) and recommendations for decision-makers will contribute to the improvement of the visibility of the PI’s research, and positively influence his academic career. Finally, this project builds a foundation for future collaborations.

35	SSBD	Small Summaries for Big Data	http://cordis.europa.eu/project/rcn/194509_en.html	European Commission H2020 Call:ERC-CoG-2014	A fundamental challenge in processing the massive quantities of information generated by modern applications is in extracting suitable representations of the data that can be stored, manipulated and interrogated on a single machine. A promising approach is in the design and analysis of compact summaries: data structures which capture key features of the data, and which can be created effectively over distributed data sets. Popular summary structures include the Bloom filter, which compactly represents a set of items, and sketches which allow vector norms and products to be estimated. These are very attractive, since they can be computed in parallel and combined to yield a single, compact summary of the data. Yet the full potential of summaries is far from being fully realized. The Principal Investigator will lead a team, working on important problems around creating Small Summaries for Big Data. The goal is to substantially advance the state of the art in data summarization, to the point where accurate and effective summaries are available for a wide array of problems, and can be used seamlessly in applications that process big data. Several directions will be pursued, including: designing and evaluating new summaries for fundamental computations such as tracking the data distribution; summary techniques for complex structures, such as massive matrices, massive graphs, and beyond; and summaries that allow the verification of outsourced computation over big data. Success in any one of these areas could lead to substantial impact on practice, as evidenced by the influence of existing summary techniques. Support in the form of a five-year research grant will allow the PI to consolidate his research in this area, and build an expert team to focus on these challenging algorithmic questions.
36	BigStorage	Storage-based Convergence between HPC and Cloud to handle Big Data	http://bigstorage-project.eu/	European Commission H2020 Call:MSCA-ITN-2014-ETN	"The consortium of this European Training Network (ETN) "BigStorage: Storage-based Convergence between HPC and Cloud to handle Big Data" will train future data scientists in order to enable them and us to apply holistic and interdisciplinary approaches for taking advantage of a data-overwhelmed world, which requires HPC and Cloud infrastructures with a redefinition of storage architectures underpinning them - focusing on meeting highly ambitious performance and energy usage objectives. There has been an explosion of digital data, which is changing our knowledge about the world. This huge data collection, which cannot be managed by current data management systems, is known as Big Data. Techniques to address it are gradually combining with what has been traditionally known as High Performance Computing. Therefore, this ETN will focus on the convergence of Big Data, HPC, and Cloud data storage, ist management and analysis. To gain value from Big Data it must be addressed from many different angles: (i) applications, which can exploit this data, (ii) middleware, operating in the cloud and HPC environments, and (iii) infrastructure, which provides the Storage, and Computing capable of handling it. Big Data can only be effectively exploited if techniques and algorithms are available, which help to understand its content, so that it can be processed by decision-making models. This is the main goal of Data Science. We claim that this ETN project will be the ideal means to educate new researchers on the different facets of Data Science (across storage hardware and software architectures, large-scale distributed systems, data management

					services, data analysis, machine learning, decision making). Such a multifaceted expertise is mandatory to enable researchers to propose appropriate answers to applications requirements, while leveraging advanced data storage solutions unifying cloud and HPC storage facilities."
37	IOSTACK	Software Defined Storage for Big Data	http://www.iostack.eu/	European Commission H2020 Call:ICT-07-2014	The main objective is to create IOStack: a Software Defined Storage toolkit for Big Data on top of the OpenStack platform. IOStack will enable efficient execution of virtualized analytics applications over virtualized storage resources thanks to flexible, automated, and low cost data management models based on software defined storage (SDS). Major challenges are: 1) Storage and compute disaggregation and virtualization. Virtualizing data analytics to reduce costs implies disaggregation of existing hardware resources. This requires the creation a virtual model for compute, storage and networking that allows orchestration tools to manage resources in an efficient manner. We will provide policy-based provisioning tools so that the provisioning of virtual components for the analytics platform is made according to the set of QoS policies. 2) SDS Services for Analytics. The objective is to define, design, and build a stack of SDS data services enabling virtualized analytics services with improved performance and usability. Among these services we include native object store analytics that will allow running analytics close to the data without taxing initial migration, data reduction services, specialized persistent caching mechanisms, advanced prefetching, and data placement. 3) Orchestration and deployment of big data analytics services. The objective is to design and build efficient deployment strategies for virtualized analytic-as-a-service instances (both ephemeral and permanent). In particular, the focus of this work is on data-intensive systems such as Apache Hadoop and Apache Spark, which enable users to define both batch and latency-sensitive analytics. This objective includes the design of scalable algorithms that strive at optimizing a service-wide objective function (e.g., optimize performance, minimize cost) under different workloads. Finally, we will create a SDS toolkit for Big Data on top of the OpenStack projects Sahara, Cinder, Nova and Swift.
38	ExtraLytics	Big Data Analytics for Real Estate	http://cordis.europa.eu/project/rcn/193793_en.html	European Commission H2020 Call:ERC-PoC-2014	Data is becoming the key to effective decision making in more and more aspects of modern society. Big data analytics provide insight into complex processes considering a breadth of factors. This promises to increase the efficiency of decision making in areas such as policy making, research, product development, or financial and retail markets. Currently, big data analytics is employed only by technologically savvy organisations with control over the relevant data. This further exacerbates the information asymmetry that has been plaguing many of these areas: real estate where buyers have long suffered from limited information, often controlled by the seller; or retail markets where sellers are now employing sophisticated competitive price intelligence solutions unaffordable to buyers. Who can tell you if it is better to buy a house with a garden or with a garage; or where renovations will yield more

					rent over time? ExtraLytics answers these questions through a combination of big data extraction and analytics. ExtraLytics introduces analytics and prediction models into DIADEM's platform for accurate big data extraction from the web, which is able to extract required data at massive scale from the web. For the proof-of-concept, ExtraLytics initially focuses on residential real estate in the UK, a £4,135bn market (UBS), where buyers and renters are required to make fast decision with limited information—time on market is often below a week or even below a day as for desirable rentals in Oxford. "In an age when 'big data' is shaping our every move, how come an industry that takes so much profit can't keep pace with technology and use it to actually deliver a service worth the money they take in commission?" (The Telegraph). ExtraLytics addresses this lack of data-driven analytics by giving consumers and investors alike a tool for better understanding properties, their location, their neighborhood, and their investment potential compared to other properties.
39	OPTIMUM	Multi-source Big Data Fusion Driven Proactivity for Intelligent Mobility	http://cordis.europa.eu/project/rcn/193380_en.html	European Commission H2020 Call:MG-7.1-2014	Transportation sector undergoes a considerable transformation as it enters a new landscape where connectivity is seamless and mobility options and related business models are constantly increasing. Modern transportation systems and services have to mitigate problems emerging from complex mobility environments and intensive use of transport networks including excessive CO2 emissions, high congestion levels and reduced quality of life. Due to the saturation of most urban networks, innovative solutions to the above problems need to be underpinned by collecting, processing and broadcasting an abundance of data from various sensors, systems and service providers. Furthermore, such novel transport systems have to foresee situations in near real time and provide the means for proactive decisions, which in turn will deter problems before they even emerge. Our vision is to provide the required interoperability, adaptability and dynamicity in modern transport systems for a proactive and problem-free transportation system. OPTIMUM will establish a largely scalable, distributed architecture for the management and processing of multisource big-data, enabling continuous monitoring of transportation systems needs and proposing proactive decisions and actions in an (semi-) automatic way. OPTIMUM follows a cognitive approach based on the Observe, Orient, Decide, Act loop of the big data supply chain for continuous situational awareness. OPTIMUM's goals will be achieved by incorporating and advancing state of the art in transport and traffic modeling, travel behavior analysis, sentiment analysis, big data processing, predictive analysis and real-time event-based processing, persuasive technologies and proactive recommenders. The proposed solution will be deployed in real-life pilots in order to realise challenging use cases in the domains of proactive improvement of transport systems quality and efficiency, proactive charging for freight transport and Car2X communication integration.

40	-	CSR: Small: Cost Effective, High Performance Solutions Using Erasure Codes for Big Data Management in Large Data Centers	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Data and I/O availability is an increasing concern in todayÕs large data centers where both data volume and complexity are increasing dramatically. Most existing solutions are based on multi-replication techniques to provide data redundancy, where data chunks are replicated across storage server nodes. However, multi-replication techniques are insufficient to manage big data: itÕs a big challenge to efficiently replicate N copies of a data set of tens-to-hundreds of petabytes! As an alternative solution, erasure codes tolerating multiple failures can provide reliability and availability at much lower cost. However, the biggest challenge using erasure codes to manage big data is the performance problem due to the complex encoding/decoding operations, which limits the application of erasure codes in large-scale data centers. This project develops cost effective techniques to exploit erasure codes to achieve high availability and enhance performance in large data centers to efficiently manage big data via several research innovations. This project cohesively investigates how to utilize proper spatial cost and system/architecture techniques to improve the overall data access performance of server clusters built upon erasure codes. This research has fundamental contributions to pave the way to efficiently deploy data centers using erasure codes. It has potential to benefit numerous big data applications such as online searching, social network, e-business, health care, and so on which are typically data intensive.

41	-	BIGDATA: Small: DCM: DA: Building a Mergeable and Interactive Distributed Data Layer for Big Data Summarization Systems	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Big data today is stored in a distributed fashion across many different machines or data sources. This poses new algorithmic and system challenges to performing efficient analysis on the full data set. To address these difficulties, the PIs are building the MIDDLE (Mergeable and Interactive Distributed Data LayEr) Summarization System and deploying it on large real-world datasets. The MIDDLE system builds and maintains a special class of summaries that can be efficiently constructed and updated while still allowing fine-grained analysis on the heavy tail. Mergeable summaries can represent any data set with a guaranteed tradeoff between size and accuracy, and any two such summaries can be merged to create a new summary with the same size-accuracy tradeoff.Interactive summaries can be quickly adapted to a specified query range of data while maintaining the same size-accuracy tradeoffs relative to the data in that range. This allows accurate efficient analysis to zero-in on small subsets of big data.The MIDDLE system enables different big data users to develop a wide spectrum of efficient and scalable data analytic tasks through the use of data summaries. The MIDDLE system is being evaluated and refined with the aid of domain experts. Since the prospect of data-summary-based analytics becoming a part of standard techniques in processing big data is tantalizing, this research generates broader impacts on the nation's government agencies, research institutes, education system, and high-tech industries. Our broad impacts also extend to academia and community outreach, through the design and development big data curriculum and education, and the involvement of general public in understanding and using big data through concise summaries.

42	-	Student Travel Fellowships for the 2015 IEEE International Conference on Big Data (IEEE Big Data October 29 - November 1, 2015)	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	"Big Data" has emerged as a new approach to computing that is transforming science, engineering, medicine, health care, finance, business, and ultimately society itself. The IEEE International Conference on Big Data 2015 (IEEE BigData 2015) provides a leading venue for disseminating the latest research in big data research, development, and applications. The conference will take place in Santa Clara, California from October 29 - November 1, 2015. The conference co-chairs are Dr. Laura Haas (IBM), and Prof. Vipin Kumar (U. Minn.). The program committee co-chairs are: Dr. Howard Ho (IBM), Prof. Beng Chin Ooi (National Univ. of Singapore), and Prof. Mohammed J. Zaki (RPI). This award will help support travel of Ph.D. students at U.S. universities who are primary authors on full papers that have been accepted to the technical program or are participating in the doctoral student symposium. Special efforts are being made to recruit and fund women and minority students.This conference solicits high-quality original research papers in any aspect of big data, including infrastructure, management, search and mining, security and privacy, and applications. Contributions will advance the state of the art in techniques, algorithms, and systems.For further information see the conference homepage: http://cci.drexel.edu/bigdata/bigdata2015/

43	-	IGERT-CIF21: Big Data U: A Program for Integrated Multidisciplinary Education and Research for Big Data Science	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DGE	This Integrative Graduate Education and Research Traineeship (IGERT) award provides Ph.D. students at the University of Washington with multidisciplinary training in computer science, statistics, and domain sciences (oceanography, astronomy, chemistry, and genome sciences). Through this blended approach, trainees will learn how to manage, analyze, and visualize increasingly large amounts of data (known as ?Big Data?), thereby being prepared to address the challenges of cyberinfrastructure in the 21st century. Intellectual Merit: By developing a new Ph.D. program that involves partnerships with 11 leading companies and national labs in the field of Big Data, this program provides trainees with a collaborative approach to processing, scaling, and modeling massive and complex data sets for the scientific community. Trainees learn to create new statistical and machine learning techniques needed to manage large data sets. Additionally, this program builds an open-source system for scientists worldwide to access and analyze Big Data through a Cloud service.Broader Impacts: This IGERT traineeship program aims to create a new Big Data curriculum that will be delivered online and through University of Washington outreach initiatives. The program also prepares a new generation of scientists with the interdisciplinary tools to approach problems that will arise in the field of cyberinfrastructure. Moreover, this program promotes the development of a diverse STEM workforce by recruiting and training underrepresented groups, women, and students with disabilities, particularly through a partnership with the AccessComputing Alliance and the University of Washington DO-IT program.IGERT is an NSF-wide program intended to meet the challenges of educating U.S. Ph.D. scientists and engineers with the interdisciplinary background, deep knowledge in a chosen discipline, and the technical, professional, and personal skills needed for the career demands of the future. The program is intended to establish new models for graduate education and training in a fertile environment for collaborative research that transcends traditional disciplinary boundaries, and to engage students in understanding the processes by which research is translated to innovations for societal benefit.

44	-	BIGDATA: Collaborative Research: F: Making Big Data Accessible on Personal Devices: Big Network Algorithms, External Memory, and Data Streams	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Big networks are constantly growing in both size and relevance: from social networks such as Facebook and Twitter, to brain networks, gene regulatory networks, and health/disease networks. The traditional approach to analyzing such big datasets is to use powerful supercomputers (clusters), ideally large enough to store the data in main memory. The downsides to this approach are that many potential users of big data lack such powerful computational resources (e.g. point-of-sale Bitcoin blockchain analysis), and it can be difficult to solve unexpected problems within such a large infrastructure (e.g. image analysis after the Boston Marathon Bombing). The algorithms developed in this project will enable the processing of huge datasets on computational devices with a limited amount of fast memory, connected to a relatively slow external data source.This project will investigate the extent to which complex network analysis can be performed on a single computer, even a mobile device such as a smartphone. To this end, the project will develop external-memory, cache-oblivious, and streaming algorithms for analyzing and understanding big network data, even on relatively weak computational devices. These algorithms will make big data analysis accessible to a much broader audience, enabling new applications. The approach uniquely combines advanced algorithmic techniques, including approximation algorithms, parameterized algorithms, graph algorithms, graph structure theory, and computational geometry, to solve real-world problems on big networks.

45	-	BIGDATA: IA: Virtual Observatory of Innovation Communities and Ecosystems (VOICE): Advancing Big Data with Ecology Theory and Data Science	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Today's digital revolution is fostering remarkable innovations. The landscape of innovation is rapidly changing and thus difficult to navigate or study. Understanding broader participation in innovation requires large datasets on the innovation participants and their activities. As organizational and industry boundaries become fuzzy in the digitized world, small data analysis cannot fully explain how boundaries shift and evolve. Open innovation leads to overlapping roles of designer and user, making it inadequate to examine innovation development and adoption separately. Therefore, it is necessary to collect data on larger units of analysis encompassing multiple organizations and multiple industries engaging in both the production and use of innovations. Further, the fast pace and risky nature of digital innovations mean that it is no longer sufficient to retrospectively study only successful innovations , but it is important to analyze contemporaneous data on a full range of innovations, including emergent and potentially failing ones, as they unfold. Addressing these challenges in conducting and studying digital innovations, this research team will build the Virtual Observatory of Innovation Communities and Ecosystems (VOICE) that (1) collects, connects, and curates large datasets from multiple sources on the people, groups, and organizations engaging with multiple innovations; (2) employs data science techniques to process and analyze these datasets; and (3) advances an ecology theory of digital innovation.The project collects data on Big Data innovations from popular, professional, and trade press; academic publications; press releases; blogs and wikis; and social media. Further, a suite of theory-informed measures and models are developed for characterizing the evolving innovations and their communities in the Big Data ecosystem. Finally, via an open, layered modular information platform, the project's datasets and tools are shared widely and data and tools from other innovation ecosystems can be added as well. The VOICE platform would become a "Hubble Telescope" for everyone to observe and partake in digital innovations beyond those associated with Big Data. Policymakers are able to evaluate the mix of capabilities in innovation communities and recommend necessary interventions to broaden participation, reduce inequality, and cultivate partnerships. By monitoring emergent trends in different but related innovation communities, entrepreneurs and managers would be more likely to develop or adopt the right innovation at the right time. Overall, VOICE enhances the research and education infrastructure that supports individuals from underrepresented backgrounds, students, educators, and other citizens to engage mindfully with science and technology innovations.

46	-	BIGDATA: Collaborative Research: F: Making Big Data Accessible on Personal Devices: Big Network Algorithms, External Memory, and Data Streams	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Big networks are constantly growing in both size and relevance: from social networks such as Facebook and Twitter, to brain networks, gene regulatory networks, and health/disease networks. The traditional approach to analyzing such big datasets is to use powerful supercomputers (clusters), ideally large enough to store the data in main memory. The downsides to this approach are that many potential users of big data lack such powerful computational resources (e.g. point-of-sale Bitcoin blockchain analysis), and it can be difficult to solve unexpected problems within such a large infrastructure (e.g. image analysis after the Boston Marathon Bombing). The algorithms developed in this project will enable the processing of huge datasets on computational devices with a limited amount of fast memory, connected to a relatively slow external data source.This project will investigate the extent to which complex network analysis can be performed on a single computer, even a mobile device such as a smartphone. To this end, the project will develop external-memory, cache-oblivious, and streaming algorithms for analyzing and understanding big network data, even on relatively weak computational devices. These algorithms will make big data analysis accessible to a much broader audience, enabling new applications. The approach uniquely combines advanced algorithmic techniques, including approximation algorithms, parameterized algorithms, graph algorithms, graph structure theory, and computational geometry, to solve real-world problems on big networks.

47	-	IGERT: Big Data Social Science - An Integrative Education and Research Program in Social Data Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DGE	This Integrative Graduate Education and Research Traineeship (IGERT) award draws together a diverse interdisciplinary team of researchers to create a new training program in Social Data Analytics, aimed at producing a new type of scientist capable of meeting emerging big data challenges. Intellectual Merit: In response to massive new sources of data, "data science" and "analytics" are emerging as new fields of inquiry, merging statistics, computer science, and visualization. The greatest challenges and opportunities arise from socially-generated big data, observed as a result of human interactions that are increasingly recorded via web, mobile device, and distributed sensors, or revealed through digitization of historical records. Society faces a transformative data deluge, from which new scientific, economic, and social value can be extracted. The BDSS-IGERT project includes a new curriculum, training in advanced technologies of data science and analytics, a series of research rotations ? in both academic and nonacademic settings ? and a challenge mechanism, under which interdisciplinary teams compete to innovate solutions to real social data analytics problems. Broader Impacts: A recent report estimated that big data contain unexploited economic value over $1 trillion annually, but that 165,000 positions in "deep analytics" will need to be filled. Government statistics agencies need to produce useful statistical products, at lower cost, while protecting confidentiality. Big data technologies like crisis mapping present opportunities to address grand social challenges. Further, the BDSS-IGERT project expands the participation of underrepresented groups in data science, by combining an exciting new field with a focus on diversity as a research theme. IGERT is an NSF-wide program intended to meet the challenges of educating U.S. Ph.D. scientists and engineers with the interdisciplinary background, deep knowledge in a chosen discipline, and the technical, professional, and personal skills needed for the career demands of the future. The program is intended to establish new models for graduate education and training in a fertile environment for collaborative research that transcends traditional disciplinary boundaries, and to engage students in understanding the processes by which research is translated to innovations for societal benefit.

48	-	BIGDATA: Collaborative Research: IA: F: Fractured Subsurface Characterization Using High Performance Computing and Guided by Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Natural fractures act as major heterogeneities in the subsurface that controls flow and transport of subsurface fluids and chemical species. Their importance cannot be underestimated, because their transmissivity may result in undesired migration during geologic sequestration of CO2, they strongly control heat recovery from geothermal reservoirs, and they may lead to induced seismicity due to fluid injection into the subsurface. Advanced computational methods are critical to design subsurface processes in fractured media for successful environmental and energy applications. This project will address the following key BIG data and computer science challenges: (1) Computation of seismic wave propagation in fractured media; (2) BIG DATA analytics for inferring fracture characteristics; (3) High Performance Computation of flow and transport in fractured media; and (4) Integration of data from disparate sources for risk assessment and decision-making. This will enable design of technologies for addressing key societal issues such as safe energy extraction from the surface, long-term sequestration of large volumes of greenhouse gases, and safe storage of nuclear waste. The project will provide interdisciplinary training for a team of graduate students and postdoctoral fellows. Outreach to high schools teachers and minorities through a planned workshop will inspire interest in environmental green-engineering, mathematics, and computational science. Numerous applications will benefit from this research, including Computer and Information Science and Engineering (CISE), Geosciences (GEO), and Mathematical and Physical Sciences (MPS).The proposed research will emphasize high performance computation (HPC) approaches for characterizing fractures using large subsurface seismic data sets, BIG data analytics for extraction of fracture related information from seismic inversion results and long-duration dynamic data, and advanced computational approaches for modeling flow, transport, and geomechanics in fractured subsurface systems. The specific objectives are to: Develop an efficient forward modeling algorithm for seismic wave propagation in fractured media using efficient computational schemes. Compute flow and transport in fractured media using an efficient computational scheme implemented on GPUs such as mimetic finite differences. Perform efficient multiphysics simulation of flow and geomechanics in fractured media. Integrate information from time-lapse seismic inversion and flow/transport simulation using novel statistical schemes. Joint inversion of seismic and fluid flow data and uncertainty quantification using efficient computational schemes. Develop and deploy a scalable hybrid-staging based substrate that can support targeted workflows using staging-based in-situ/in-transit approaches. Computational simulation is critical to design subsurface processes for successful environmental and energy applications. Project URL: http://csm.ices.utexas.edu/current-projects/

49	-	EAGER: Council for Big Data, Ethics, and Society	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Big data are changing the ways in which individuals are tracked and the corresponding ways in which science must respond. Data uses run the gamut from public health modeling to personalization of online content to potentially discriminatory practices. Big data involve not just using new tools to get better answers but also new social, cultural and ethical issues. This award establishes a model for dealing with the ethical, social and legal impacts of big data projects from their outset, with an eye to developing next generation protocols to cover these kinds of impacts. A multidisciplinary council will be established to engage with and complement discussions already underway in mathematics, the sciences, engineering, and computer science. The goal of the Council will be to address issues such as security, privacy, equality, and access in order to help guard against the repetition of earlier mistakes and inadequate preparation. Through public commentary, events, white papers, and direct engagement with data analytics projects, the Council will develop frameworks to help researchers, practitioners, and the public understand the social, ethical, legal, and policy issues that underpin big data phenomena.Functionally, the goal will be to help guard against the repetition of known mistakes and inadequate preparation by working across domains and disciplines involved in big data projects. In the process, those working on this project will develop new and powerful paradigms for identifying and understanding leading edge social, political, ethical, and legal issues. Both the research outputs and the coordinated network will help inform the design of scientific projects. Furthermore, the public-oriented nature and accessible outputs of this project will provide input for public discussions surrounding the big data phenomena by engaging with journalists, educators, and public policy makers. This project will create an influential community of thought leaders that can help shape the understanding of the complexities of the big data and also provide engagement for young scientists.

50	-	CSR: Small: Collaborative Research: Comprehensive Algorithmic Resilience (CAR) for Big Data Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Big Data analytics is the process of mining useful knowledge in very large data sets, critical to the advancement of many research and application fields. With manufacturing technology downscaling of processor size coupled with increasing power densities, modern computer systems suffer from potential hardware and software failures, which can manifest themselves as errors. The errors happen when long-running Big Data analytics are executed on the systems, causing system crashes or worse, returning undetected incorrect results. While numerous hardware-based resilience methods exist, they often come at the cost of excessive power efficiency reduction and substantial design complexity enlargement among others. The project aims to reinforce popular Big Data analytics by embracing a host of comprehensive algorithmic resilience (CAR) software techniques that include concurrent error detection, coordinated checkpointing, and execution recovery, for high execution resilience. Upon detecting potential hardware and software errors concurrently during analytics, CAR enables execution recovery from detected errors without lofty overhead common to hardware-based resilience methods. Research activities of the project aim to achieve six main objectives that focus on addressing several technical challenges to realize CAR, based on investigators' encouraging preliminary results and prior work. The success of this project can benefit wide scientific and industrial applications due to its better support of Big Data analytics and processing. Research advances from this research are to be incorporated into undergraduate and graduate education, to be disseminated and shared broadly through technical presentations and by a website, and to inspire high school students for their STEM interest.

51	-	CC-NIE Integration: Bridging, Transferring and Analyzing Big Data over 10Gbps Campus-Wide Software Defined Networks	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	This CC-NIE Integration project performed at the Louisiana State University is developing a cyber-infrastructure integrating six different large scale scientific research groups with high performance computing (HPC) clusters at LSU using software defined network technology and Hadoop and MPI (Message Passing Interface) based parallel frameworks.The project consists of three objectives: (i) Building 10Gbps software-defined network (SDN) with OpenFlow switches and controllers to provide multiple virtual network slices to each group; (ii) Transferring Big Data with automatically tuned operation through multiple optical paths over the SDN network achieving at most 20Gbps of aggregated disk-to-disk transfer rate; and (iii) Analyzing Big Data by developing data-intensive distributed computing frameworks with Hadoop and MPI technologies parallelizing large number of jobs over HPC clusters.Those three components are integrated with a web portal service and a GENI-enabled network management system. To achieve high disk-to-disk transfer rate at 20Gbps, the project has an industrial partnership with Samsung Electronics that contributes 70TB Solid State Drive (SSD) storage and optimizes the I/O bandwidth.The cyber-infrastructure accelerates the current Big Data research projects spanning a wide range of research areas including gene sequencing research at Biology and Vet School, computational chemistry, big data mining at Computer Science, coastal hazard simulation research at civil and environmental engineering. In the end, the project will establish a methodology to build integrated cyber-infrastructures consisting of high speed networks, high performance computing, and high speed storage for the Big Data Science and Engineering.

52	-	III: Small: Linking and Resolving Entities in Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	This project will explore the challenge of cleaning data in the context of analysis pipelines over big data. Data cleaning has traditionally been designed to improve data quality in ETL systems where enterprise data is collected, prepared, staged, transformed, and loaded into a data warehouse to support offline data analysis. In the era of big data, such back-end processes are quickly giving way to interactive exploratory data analysis where analysts immerse themselves in data (possibly collected from heterogeneous data sources) in order to drive online (near-) real-time decision making. Existing systems do not scale to the volume, velocity, or the variability of the dynamically generated data (e.g., social media streams) and the offline architecture is unsuited for the online real-time nature of analysis. The market is abuzz with innovations in data transformation technologies, e.g., TriFacta allows analysts to visually manipulate data to generate complex analytical transformations and Data Tamer is exploring scalable data curation from diverse sources. Data quality (and hence data cleaning technologies) remain at the core of big-data analytics. Many popular media (as well as academic) articles have highlighted challenges such as entity linking and resolution as among the most important and immediate roadblocks for big data analytics. The key insight on which this project is based is that data cleaning to support analytics over big data is not simply a matter of scaling up known approaches to larger data sets by exploiting more hardware. While scale up is important, big data analytics in streaming, real-time, and interactive settings requires a paradigm shift in how data cleaning is performed. This project will significantly impact and change the modern practices of data cleaning and the way cleaning is integrated in the Big Data analysis pipeline and will explore broader impact through: (a) technology transfer opportunities with a relevant industrial partner whose existing products could benefit from the proposed research; and (b) open source effort in the context of the ongoing social media analytics system (SoDAS), currently under development, in which the proposed research algorithms will be integrated.This research will explore two new innovations that will help advance data cleaning to enable Big Data analysis. The first innovation explores a progressive approach to entity resolution to support progressive analysis. The research will explore an approach where progressiveness is pervasive spanning all the phases of the cleaning process especially in scenarios when cleaning is based on complex logic possibly requiring dynamic acquisition of additional contextual information. The second innovation is the analysis-aware data cleaning that is developed for structured queries (e.g., Hive and SQL) for both one-time and continuous query scenarios that are issued on top of static and streaming data. The project will address these methodologies at the higher conceptual level as well as implement them on modern highly-parallel computing platforms and frameworks that run on a cluster of machines. The project will exploit two concrete contexts to guide the research exploration: (a) supporting analytical queries over structured web data sources such as fusion tables; and (b) online analysis of social media data. These application contexts will serve as vehicles for testing and demonstrating the research. The planned research, system development, and educational activities (e.g., curriculum changes to incorporate projects related to big data and data quality in the CS curriculum at UCI) will significantly enhance the educational experience of students, preparing them for a brighter future in the today?s knowledge driven society. More information about the project can be found at http://sherlock.ics.uci.edu.

53	-	BIGDATA: F: DKM: Collaborative Research: Making Big Data Active: From Petabytes to Megafolks in Milliseconds	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	A wealth of digital information is being generated daily through social networks, blogs, online communities, news sources, and mobile applications in an increasingly sensed world. Organizations and researchers recognize that tremendous value and insight can be gained by capturing this emerging data and making it available for querying and analysis. First-generation Big Data management efforts have been passive in nature -- queries, updates, and/or analysis tasks were mainly scaled to handle very large volumes of data. In contrast, this project will develop new techniques for continuously and reliably capturing Big Data collections (arising from social, mobile, Web, and sensed data sources) and will enable timely delivery of the right information to the relevant end users. In short, this project will provide a scalable foundation for moving from Big Passive Data to Big Active Data. Techniques should be developed to enable the accumulation and monitoring of petabytes of data of potential interest to millions of end users; when "interesting" new data appears, it should be delivered to end users in a time frame measured in (100's of) milliseconds. This project will build such an Active Big Data Management system and make it available as open source to the community. Students will be trained in technologies related to Big Active Data management and applications; such training is critical to addressing the information explosion that social media and the mobile Web are driving today. The general-purpose foundation for active information dissemination from Big Data will have broader impacts in areas such as public safety and public health. There are many challenges involved in building a foundation for Big Active Data. On the "data in" side, these include resource management in very large scale, LSM-based storage systems and the provision of a highly available, elastic facility for fast data ingestion. On the "data processing" side, challenges include the parallel evaluation of a large number of declarative data subscriptions over multiple) highly partitioned data sets. Amplifying this challenge is a need to efficiently support spatial, temporal, and similarity predicates in data subscriptions. Big Data also makes result ranking and diversification techniques critical in order for large result sets to be manageable. On the "data out" side, challenges include the reliable and timely dissemination of data of interest to a sometimes-connected subscriber base of unprecedented scale. As a software base, this project will be jump-started by using AsterixDB(http://asterixdb.ics.uci.edu/), an open-source Big Data Management System that supports the scalable storage, searching, and analysis of mass quantities of semi-structured data. For further information see the project web sites at https://www.ics.uci.edu/BigActiveData and http://www.cs.ucr.edu/~tsotras/BigActiveData

54	-	CI-ADDO-NEW: ASTERIX: A Community Software Platform for Big Data Research, Analysis, and Management	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Vast quantities of digital information are being generated today on a daily basis through social networks, blogs, online communities, news sources, and mobile applications as well as our increasingly sensed surroundings. Tremendous insight can be gained by storing and making such ?Big Data? available for exploration in a wide variety of domains. Likely beneficiaries include business, social sciences, public health, national security, political science, public safety, medicine, and government policy. Researchers exploring these benefits need software to manage and analyze Big Data, and researchers investigating algorithms and programming models for Big Data can benefit tremendously by being provided with shared building blocks to use as a foundation for their efforts.Over the past 3.5 years we have developed an initial version of AsterixDB, a powerful new Big Data Management System (BDMS) for scalably storing, managing, searching, and analyzing collections of Big Data using clusters of commodity computers. AsterixDB has a layered code base that consists of a scalable runtime platform (Hyracks), a model-neutral framework for parallel query compilation for Big Data (Algebricks), and the end-user-targeted AsterixDB BDMS itself. This NSF project is turning AsterixDB and its internal software stack into robust, supported, open-source resources for use by the Big Data applications and technology research communities. AsterixDB and its components will be helpful in training students in Computer Science and other data-related sciences, at universities everywhere, about Big Data technologies. This is critical for addressing the information explosion being brought to us courtesy of social media and the mobile Web.

55	-	REU Site: BigDataX: From Theory to Practice in Big Data Computing at Extreme Scales	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CCF	This award establishes a new Research Experiences for Undergraduates (REU) site at Illinois Institute of Technology and the University of Chicago. The site, which is named BigDataX, focuses on undergraduate research in both the theory and practice of big data computing at extreme scales. BigDataX includes a diverse group of 8 undergraduate students and 4 mentors spread out over the two institutions. Students will conduct research in the area of big data and how it will address issues related to the design, analysis, and implementation of run-time systems and storage systems to support big data applications. This work includes making extreme scale computing more tractable, touching every branch of computing in high-end computing and datacenters. These advancements will impact scientific discovery and economic development at the national level, and they will strengthen a wide range of research activities enabling efficient access, processing, storage, and sharing of valuable scientific data from many disciplines. The primary focus of this award is to promote a data-centric view of scientific and technical computing, at the intersection of distributed systems theory and practice. The project team has identified various data-intensive applications from many disciplines such as astronomy, bioinformatics, medical imaging, that demonstrate characteristics of big-data applications. This work focuses on the design, implementation, and optimization of runtime systems to support parallel programming systems for both Many-Task Computing and High-Performance Computing. The work centers on distributed scheduling, dynamic provisioning, improved fault tolerance, and support for heterogeneous computing. To better support big data applications, the team is exploring distributed file systems and improvements to a variety of critical components such as metadata management, I/O access pattern coalescing, distributed provenance as well as exploring novel interfaces into distributed storage systems. This work involves real applications, real data, and real testbeds ranging from small clusters at IIT and UChicago, supercomputers at Argonne National Laboratory, to the Amazon AWS cloud.

56	-	Collaborative Research: Leveraging "Big Data" to Explore Big Ideas: Utilizing the Paleobiology Database to Provide Hands-on Research Opportunities for Undergraduates	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DUE	With the rapid growth of "big data" in the sciences, the need to understand and manage their use in undergraduate education has become increasingly important. This project will investigate how the Paleobiology Database (PBDB), a large database of fossil information, can be leveraged to provide research experiences to undergraduates at two and four-year colleges that do not necessarily have extensive fossil collections of their own. It will also investigate how students' attitudes towards scientific research change after engaging in research experiences using a large database such as the PBDB. Two year colleges and distance-based learning initiatives seek research-based alternatives to traditional lab activities, and "big data" provides opportunities for the hands-on science they pursue. "Big data" science initiatives, such as the Paleobiology Database (PBDB), provide affordable and easily accessible research opportunities, but the extent of their value in the undergraduate classroom has yet to be tested and documented. The PBDB contains a wealth of data on fossils entered by experts from around the world. Research skills, such as critical thinking, statistics, and computing, are essential to scientifically literate citizens. This project will guide how other large scientific databases can be used to craft research experiences for undergraduate students, and provide the means for other earth science programs to engage their undergraduates in scientific research.Integration of research experiences into the undergraduate classroom can result in increased recruitment, retention, persistence, and motivation of science students. "Big data" science initiatives, such as the Paleobiology Database (PBDB), can provide inexpensive and accessible research opportunities, but the value of these programs in the undergraduate classroom has yet to be tested and documented. This project will: (1) evaluate how research experiences using the PBDB compare to field or lab-based research experiences; (2) establish how a large online database can be leveraged as a tool to provide effective research experiences for students at two and four-year colleges; (3) and determine how research databases can be used to teach STEM principles to non-majors, majors, and pre- service teachers. To accomplish these goals, guided research activities and a new user interface will be developed for the PBDB. The project will implement the PBDB-based research activities across two pairs of two-year and four-year institutions in Virginia that attract students with broadly diverse backgrounds. This project will assess how applicable "big data" science is to undergraduate education, specifically to what extent online databases can engage students in authentic research experiences. Research activities will focus on content crucial to scientific literacy, and will be scaffolded to build transferable skills including critical thinking, quantitative skills, and database savvy. The new PBDB interface and research activities resulting from this work will be implemented across a diversity of two- and four-year colleges, allowing the evaluation of its applicability across institution types and student populations, including distance learning versus traditional educational settings.

57	-	NeTS: CSR: Medium: Collaborative Research: Enabling Flexible and High Performance Big Data Analytics Over Geo-Distributed Clouds	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Large organizations and small enterprises alike leverage datacenters across the globe to offer Internet services to their users. These sites routinely gather data pertaining to end user activities to provide better services, and they collect server monitoring logs and performance counters to ensure uninterrupted service. Although fast, efficient, and cost-effective analyses of these large datasets can significantly improve users' quality of experience and enable novel applications, the wide area network (WAN) that connects the datacenters poses a considerable challenge: because WAN bandwidth is limited and expensive, and WAN latency is high and variable, both the performance and timeliness of analytics are affected by the WAN.This project aims to build a new WAN-aware big data stack customized for flexible geo-distributed data analytics. The project will not impose any constraints on the set of queries that can be issued, and it will support a variety of performance objectives including obtaining timely responses, minimizing batch completion times, or using minimal bandwidth. To account for unpredictable and fine-timescale changes to WAN conditions and to enable coordination among the actions taken by different layers of the analytics stack, this project will enable holistic, cross-layer visibility and optimizations. It will incorporate awareness of the geo-distributed setting in the stack's upper layers (e.g., query optimization) and of application-level objectives in the lower layers (e.g., networking). This will result in a radical re-factoring of the API and interfaces between query optimization, query execution, resource negotiation, wide-area storage, and network routing/scheduling.Software artifacts from this project will be incorporated into existing open source big data stacks, making the research outcomes broadly available for public reuse. The experimental harnesses will be made available to ensure repeatability and to foster follow up research. The research outcomes will guide industry evolution as the industry slowly shifts from single-datacenter to geo-distributed settings. The project has a substantial educational component involving the introduction of new courses on big data systems at both graduate and undergraduate levels that will involve hands-on exercises with state-of-the-art big data software, and it will reach out to high-school students, women, and underrepresented minorities through big data boot camps.

58	-	NeTS: CSR: Medium: Collaborative Research: Enabling Flexible and High Performance Big Data Analytics Over Geo-Distributed Clouds	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Large organizations and small enterprises alike leverage datacenters across the globe to offer Internet services to their users. These sites routinely gather data pertaining to end user activities to provide better services, and they collect server monitoring logs and performance counters to ensure uninterrupted service. Although fast, efficient, and cost-effective analyses of these large datasets can significantly improve users' quality of experience and enable novel applications, the wide area network (WAN) that connects the datacenters poses a considerable challenge: because WAN bandwidth is limited and expensive, and WAN latency is high and variable, both the performance and timeliness of analytics are affected by the WAN.This project aims to build a new WAN-aware big data stack customized for flexible geo-distributed data analytics. The project will not impose any constraints on the set of queries that can be issued, and it will support a variety of performance objectives including obtaining timely responses, minimizing batch completion times, or using minimal bandwidth. To account for unpredictable and fine-timescale changes to WAN conditions and to enable coordination among the actions taken by different layers of the analytics stack, this project will enable holistic, cross-layer visibility and optimizations. It will incorporate awareness of the geo-distributed setting in the stack's upper layers (e.g., query optimization) and of application-level objectives in the lower layers (e.g., networking). This will result in a radical re-factoring of the API and interfaces between query optimization, query execution, resource negotiation, wide-area storage, and network routing/scheduling.Software artifacts from this project will be incorporated into existing open source big data stacks, making the research outcomes broadly available for public reuse. The experimental harnesses will be made available to ensure repeatability and to foster follow up research. The research outcomes will guide industry evolution as the industry slowly shifts from single-datacenter to geo-distributed settings. The project has a substantial educational component involving the introduction of new courses on big data systems at both graduate and undergraduate levels that will involve hands-on exercises with state-of-the-art big data software, and it will reach out to high-school students, women, and underrepresented minorities through big data boot camps.

59	-	Collaborative Research: Leveraging "Big Data" to Explore Big Ideas: Utilizing the Paleobiology Database to Provide Hands-on Research Opportunities for Undergraduates	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DUE	With the rapid growth of "big data" in the sciences, the need to understand and manage their use in undergraduate education has become increasingly important. This project will investigate how the Paleobiology Database (PBDB), a large database of fossil information, can be leveraged to provide research experiences to undergraduates at two and four-year colleges that do not necessarily have extensive fossil collections of their own. It will also investigate how students' attitudes towards scientific research change after engaging in research experiences using a large database such as the PBDB. Two year colleges and distance-based learning initiatives seek research-based alternatives to traditional lab activities, and "big data" provides opportunities for the hands-on science they pursue. "Big data" science initiatives, such as the Paleobiology Database (PBDB), provide affordable and easily accessible research opportunities, but the extent of their value in the undergraduate classroom has yet to be tested and documented. The PBDB contains a wealth of data on fossils entered by experts from around the world. Research skills, such as critical thinking, statistics, and computing, are essential to scientifically literate citizens. This project will guide how other large scientific databases can be used to craft research experiences for undergraduate students, and provide the means for other earth science programs to engage their undergraduates in scientific research.Integration of research experiences into the undergraduate classroom can result in increased recruitment, retention, persistence, and motivation of science students. "Big data" science initiatives, such as the Paleobiology Database (PBDB), can provide inexpensive and accessible research opportunities, but the value of these programs in the undergraduate classroom has yet to be tested and documented. This project will: (1) evaluate how research experiences using the PBDB compare to field or lab-based research experiences; (2) establish how a large online database can be leveraged as a tool to provide effective research experiences for students at two and four-year colleges; (3) and determine how research databases can be used to teach STEM principles to non-majors, majors, and pre- service teachers. To accomplish these goals, guided research activities and a new user interface will be developed for the PBDB. The project will implement the PBDB-based research activities across two pairs of two-year and four-year institutions in Virginia that attract students with broadly diverse backgrounds. This project will assess how applicable "big data" science is to undergraduate education, specifically to what extent online databases can engage students in authentic research experiences. Research activities will focus on content crucial to scientific literacy, and will be scaffolded to build transferable skills including critical thinking, quantitative skills, and database savvy. The new PBDB interface and research activities resulting from this work will be implemented across a diversity of two- and four-year colleges, allowing the evaluation of its applicability across institution types and student populations, including distance learning versus traditional educational settings.
60	-	CAREER: Modeling Personalized Brain Development	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	IIS	Big data offer an opportunity to study specific control populations (age / sex / environmental factors / demographics / genetics) and identify substantive homogeneous sub-cohorts so that one may understand the roles that potential factors play in brain development, differentiating abnormal trajectories from normal development. The image processing, statistical, and informatics tools to effectively and efficiently use big data imaging archives for quantitative population-level research and personalized medicine do not yet exist. This research will enable discovery science on a scale considerably larger than routinely possible with traditional study designs by creating

		with Big Data	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		novel informatics resources that tie archives of 3-D images into accessible research databases. This research will discover genetic and environmental factors that influence an individual's brain development and characterize the developing human brain through personal developmental trajectories. To accomplish this goal, new informatics technologies will be created to enable (1) image processing and segmentation based on image content in the context of heterogeneous, low quality, and error prone data with minimal human oversight and (2) routine archival, query, and image processing of large medical imaging datasets. This research will impact the areas of (1) informatics via novel computation models, (2) neuroscience via a new structural model of brain development, and (3) public health via newly accessible data sets for research. The science and technology innovations enabled by using big data to understand personalized brain development will be communicated in a tiered method. Outreach to the K-12 audience will target conceptualizing design criteria, inspiring students with interactive demonstrations, and providing capabilities for students to apply key concepts in hands-on engineering projects. For advanced students and researchers, new accessible course materials and online modules will be developed so that others may build upon the foundations established by this research.Novel software, data wrangling tools, and resources will be created through two research thrusts organized around a novel test bed infrastructure and synthesized in a third education/outreach thrust. Thrust 1 (Personal Brain Trajectories) will focus on extracting meaningful information from medical images when performed at scale through (1) creating automated methods robust to variations in image quality, acquisition, and transfer errors, and (2) enabling efficient human-in-loop control at scale. The research will extend novel statistical models for image content labeling while adapting quality control techniques from industrial engineering. Thrust 2 (Novel Storage & Processing) will create novel medical imaging data models to describe data acquisition / retrieval, storage, cleaning, access / security, query and processing by integrating of medical imaging standards with big data architecture derived from social network and e-commerce communities. This infrastructure will provide practical access to petabyte imaging archives, integrate with existing data workflows, and effectively function with commodity hardware. The PI will develop and release a reference test bed to evaluate new technologies in the context of computer-aided detection (CADe) of brain abnormalities while considering age, sex, and demographics. Using the test bed, researchers and students will be able to efficiently evaluate existing and emerging image processing software to screen for potential prognostic markers. In Thrust 3 (Education and Outreach), the research results will be integrated into two classes targeting undergraduate students and interactive online modules created and released through an established graduate student/faculty training program. Each summer, an undergraduate and high school student will participate in research by implementing and extending research contributions within an interactive demonstration platform. In the second through fifth summers, a high school teacher will assist in the development of curricula targeting high school students using the demonstration platform. High school students and teachers will be recruited from Nashville Metro schools with a high underrepresented minority / reduced cost lunch populations. These efforts will create an open-source, open-hardware system for public demonstration and K-12 classroom exercises.

61	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in Universities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and

					understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.
62	-	I-Corps: Feasibility Study for Commercializing a Domain-Specific Big Data Analytics Cloud Software Stack	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	IIP	In the so-called big data era, the data in many companies are becoming too big to be processed and stored by traditional computing platforms. There are urgent needs to address the challenge from ever-increasing big data facing these industries, especially for small companies and contractors who do not have large scale computing infrastructure. Those big companies that have the large scale computing infrastructure, are currently using the traditional High Performance Computing (HPC) techniques to handle the big data requirements. However, the traditional HPC techniques were designed to process computation intensive works, instead of data-intensive jobs. Moreover, the traditional HPC requires deep knowledge and expertise in HPC architectures and uses low-level parallel programming models, which typically create an obstacle for the regular users and domain experts to utilize HPC infrastructure directly. This I-Corps team proposes the cloud computing technique that provides a cost-effective solution to address the computation and storage requirements, which is especially useful to small companies and startups.This project is to explore the commercialization opportunities for a software stack that stores, processes and analyzes big volume of data in a variety of domains, including petroleum seismic data, medical images, retails, energy and more. Besides the scalable performance and fault-tolerance features provided by the software stack, it also provides high-level programming templates, workflow and a domain specific language to simplify the usage of large scale computing platforms. The proposed cloud-based software stack will deliver a customized computing platform service via a user-friendly web interface that is designed to fill the gap between the big data analytics platforms and the customer?s requirements in a variety of domains. The targeted users are data scientists, analysts, researchers and developers in the petroleum, medical and other industrial domains dealing with big volume of data.

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
63	-	Student Travel Support: ACM International Workshop on Big Data in Life Sciences, Seattle, WA, October 2, 2016	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperato	CCF	The ACM International Workshop on Big Data in Life Sciences (BigLS) is a workshop series focusing on computational and data challenges in broadly defined life sciences. The workshop was initiated in 2013, and it is hosted together with the ACM Conference on Bioinformatics, Computational Biology and Health Informatics - the flagship conference of the Association for Computing Machinery Special Interest Group on Bioinformatics, Computational Biology, and Biomedical Informatics (SIGBio). The workshop brings together leading researchers and practitioners working on a diverse range of big data problems relating to biology and medicine, and engages them in a discussion about current big data questions, the state of computational tools and analytics, the challenges and the future trends within life sciences.This project will support up to ten students and postdoctoral researchers from US academic institutions to attend BigLS 2016 to be held on October 2, 2016 in Seattle, WA. Awardees will be selected via a widely advertised competitive process involving the submission of a travel grant application, and review by the program committee. Preference will be given to women and underrepresented minority groups, first-generation college students, and undergraduate researchers.Big data problems and challenges have become a reality in modern day computational biology, bioinformatics and biomedical informatics. The workshop will feature peer-reviewed papers and invited talks on five key research themes that underline big data research in life sciences: 1) scalable algorithms and techniques for big data analytics in molecular biology; 2) statistical and integrative approaches to big data biology; 3) emerging machine learning and AI techniques for big data biology; 4) high performance computing methods and software for big data biology; 5) software and hardware foundations for managing big data in biomedical informatics.

			r=		
64	-	EAGER: Big Data Analytics for Advanced Manufacturing Improvement	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CMMI	Advanced manufacturing, is a group of manufacturing activities that use advanced technologies, cutting-edge materials, automation, and computation to manufacture new products and to improve manufacturing methods of existing products. Advanced manufacturing provides a capability to produce customized and complex products, while reducing both production time and cost. To ensure the quality and reliability of products, most advanced manufacturing systems are equipped with hundreds of sensors that are used to collect quality-related data such as the part dimensions, strength, etc. This results in a tremendous amount of data (namely, Big Data) that provide unique opportunities for improving advanced manufacturing systems. This EArly-Grant for Exploratory Research (EAGER) award supports fundamental research that enhances existing knowledge for the development of new analytical methods for Big Data that help improve the quality of manufactured products. The results of this research will have a significant impact on a variety of high profile application domains including automotive, aerospace, defense, biomedical, and medical by, preventing catastrophic failures, reducing scrap and rework costs, and improving product quality, all of which increase customer satisfaction and aid in economic growth. The use of statistical methods for improving manufacturing processes has been extensively studied and various techniques have been developed in this area. However, most existing methods fail to effectively analyze the big manufacturing process data due to their complex structure, high-volume, and high sampling rates. To address these issues, this research will propose and validate a new framework that enables modeling and analysis of Big Data in real-time by exploiting and integrating statistical learning, optimization, advanced computing, and manufacturing principles. The research team will develop a set of nonparametric methods for real-time sensing and measurement of Big Data, effective data compression, and process monitoring, change detection, and diagnosis.

			DateOperator=&ExpDateOperator=		
65	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in Universities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer,

			DateOperator=&ExpDateOperator=		store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.
66	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in Universities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmou	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of

			nt=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.
67	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in Universities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific

			=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.
68	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&Boole	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work,

		Prediction Research and Education in Universities	anElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.

69	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in Universities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and

					understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.
70	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in Universities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.
71	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in Universities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&Award	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of

			NumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.
72	-	Collaborative Research: SI2-SSI: Big Weather Web: A Common and Sustainable Big Data Infrastructure in Support of Weather Prediction Research and Education in	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&P	ACI	Earth science communities need to rely on access to large and growing amounts of curated data to make progress in research and provide adequate education. Existing infrastructures pose significant barriers to this access, especially for small to mid-size research groups and primarily undergraduate institutions: cloud services disappear when funding runs out to pay for them and therefore do not provide the long-term availability required for curated data. Similarly, in-house IT infrastructure is maintenance-intensive and requires dedicated resources for which long-term funding is often unavailable. The goal of the Big Weather Web is to make in-house IT infrastructure affordable by combining the application of three recent technologies: virtualization, federated smart storage, and big data management. Virtualization allows push-button deployment and maintenance of complex systems, smart storage provides automatic, community-wide data availability guarantees, and big data management allows for easy curation of data and its products. The combination of these three technologies allows communities to create a standard community-specific computational environment and efficiently refine it with minimal repetition of work, introducing a high degree of reproducibility in research and education. This reproducibility accelerates learning and amplifies everyone?s contribution. Due to virtualization, it can easily take advantage of cloud services whenever they become available, and it can run on in-house IT infrastructure using significantly reduced maintenance

		Universities	rogram=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		resources. The Big Weather Web will be developed in the context of numerical weather prediction with the expectation that the resulting infrastructure can be easily adapted to other data-intensive scientific communities.The volume, variety, and velocity of scientific data generated is growing exponentially. Small to mid-size research groups and especially primarily undergraduate institutions (PUIs) do not have the resources to manage large amounts of data locally and share their data products globally at high availability. This lack of resources has a number of consequences in education and research that have been well-documented in recent EarthCube workshops: (1) data-intensive scientific results are not easily reproducible, whether in the context of research or education, (2) limited or non-existent availability of intermediate results causes a lot of unnecessary duplication of work and makes learning curves unnecessarily steep, and consequently (3) scientific communities of practice are falling behind technological innovations. This Big Weather Web project focuses on the numerical weather prediction community. Numerical weather models produce terabytes of output per day, comprising a wealth of information that can be used for research and education, but this amount of data is difficult to transfer, store, or analyze for most universities. The Big Weather Web addresses this situation with the design, implementation, and deployment of "nuclei," which are shared artifacts that enable reliable and efficient access and sharing of data, encode best practices, and are sustainably maintained and improved by the community. These nuclei use existing and well-established technologies, but the integration of these technologies will significantly reduce the resource burden mentioned above. Nucleus 1 is a large ensemble distributed over seven universities. Nucleus 2 is a common storage, linking, and cataloging methodology implemented as an appliance-like Data Investigation and Sharing Environment (DISE) that is extremely easy to maintain and that automatically ensures data availability and safety. Nucleus 3 is a versioned virtualization and container technology for easy deployment and reproducibility of computational environments. Together, these nuclei will advance discovery and understanding through sharing of data products and methods to replicate scientific results while promoting teaching, training, and learning by creating a shared environment for scientific communities of practice. These shared environments are particularly important for underrepresented groups who otherwise have limited access to knowledge that is primarily propagated by social means. Our approach is a significant step towards improving reproducibility in the complex computational environments found in many scientific communities.

73	-	CRII: CI: Exploring Advanced Cyber-Infrastructure Co-Design for Big Data Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	The proliferation of digital data provides new opportunities in all areas of science, engineering and industry. However, the increasing data volumes and rates, and the associated costs in terms of latencies and energy, quickly dominate and limit data analytics applications' ability to leverage this data in an effective and timely manner. The co-design process enables scientists to reason about the rich design spaces available in software and hardware, and is expected to help in constructing a new class of cyber-infrastructure. Nevertheless, current solutions for big data analytics have important limitations that make these solutions infeasible for the next generation of cyber-infrastructure. As a result, exploring key co-design issues in the scope of big data analytics has become critical.The overarching goals of this research are to understand the performance and power behaviors and tradeoffs related to data placement, movement, and processing of big data analytics on systems with emerging architectures, and to develop models that can fundamentally enable big data analytics on ongoing cyber-infrastructure. In contrast to existing work, this research effort focuses on the co-design process that has been exploited in the context of large-scale scientific applications. Key research activities include building a framework to evaluate different classes of big data analytics, characterizing data analytics applications in terms of performance and power, and developing a methodology to construct models to understand and explore the design space.As data analytics applications become increasingly important in a wide range of domains, the ability to develop large-scale and sustainable platforms and software infrastructure to support these applications has significant potential to drive research and innovation in science and business domains. This project provides an understanding of hardware/software characteristics and requirements for big data analytics and future cyber-infrastructure design. These research activities are integrated with graduate and undergraduate student research, and leverage minority student outreach networks at Rutgers.

74	-	Collaborative Research: Efficient Parallel Iterative Monte Carlo Methods for Statistical Analysis of Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DMS	The integration of computer technology into science and daily life has enabled the collection of massive volumes of data. To analyze these data, one may have to resort to parallel and distributed architectures. While the parallel and distributed architectures present new capabilities for storage and manipulation of big data, it is unclear, from the inferential point of view, how the current statistical methodology can be transported to the paradigm of big data. Also, growing data size typically comes together with a growing complexity of data structures and of the models needed to account for the structures. Although iterative Monte Carlo algorithms, such as the Markov chain Monte Carlo (MCMC), stochastic approximation, and expectation-maximization (EM) algorithms, have proven to be very powerful and typically unique computational tools for analyzing data of complex structures, they are infeasible for big data as for which a large number of iterations and a complete scan of the full dataset for each iteration are typically required. Big data have put a great challenge on the current statistical methodology. The investigators propose a general principle for developing Monte Carlo algorithms that are feasible for big data and workable on parallel and distributed architectures; that is, using Monte Carlo averages calculated in parallel from subsamples to approximate the quantities that originally need to calculate from the full dataset. This principle avoids the requirement for repeated scans of full data in algorithm iterations, while enabling the algorithm to produce statistically sensible solutions to the problem under consideration. Under this principle, a general algorithm, the so-called subsampling approximation-based parallel stochastic approximation algorithm, is proposed for parameter estimation for big data problems. Unlike the existing algorithms, such as the bag of little bootstraps, aggregated estimation equation, and split-and-conquer algorithms, the proposed algorithm works for the problems for which the observations are generally dependent. Under the same principle, a subsampling approximation-based parallel Metropolis-Hastings algorithm is proposed for Bayesian analysis of big data, and a subsampling approximation-based parallel Monte Carlo EM algorithm is proposed for parameter estimation for the big data problems with missing observations. In addition to the subsampling approximation-based parallel iterative Monte Carlo algorithms, an embarrassingly parallel MCMC algorithm is proposed for Bayesian analysis of big data based on the popular idea of divide-and-conquer. Various schemes of dataset partition and results aggregation are proposed. The validity of the proposed parallel iterative Monte Carlo algorithms, including both the subsampling approximation-based and embarrassingly parallel ones, will be rigorously studied. The proposed algorithms will be applied to spatio-temporal modeling of satellite climate data, genome-wide association study, and stream data analysis.The intellectual merit of this project is to propose a general principle for statistical analysis of big data: Using Monte Carlo averages of subsamples to approximate the quantities that originally need to calculate from the full dataset. This principle provides a general strategy for transporting the current statistical methodology to the paradigm of big data. Under this principle, a few subsampling approximation-based parallel iterative Monte Carlo algorithms are proposed. The proposed algorithms address the core problem of big data analysis?how to make a statistically sensible analysis for big data while avoiding repeated scans of the full dataset. This project will have broader impacts because big data are ubiquitous throughout almost all fields of science and technology. A successful research program in theory and methods of parallel iterative Monte Carlo computations can have immense benefit

					widely throughout science and technology. The research results will be disseminated to the communities of interest, such as atmospheric science, biomedical science, engineering, and social science, via direct collaboration with researchers in these disciplines, conference presentations, books, and papers to be published in academic journals. The project will have also significant impacts on education through direct involvement of graduate students in the project and incorporation of results into undergraduate and graduate courses. In addition, the package Distributed Iterative Statistical Computing (DISC) that will be developed under this project is designed to provide a platform for Ph.D. students and researchers like the investigators with network-connected computers to experiment new ideas of developing efficient iterative Monte Carlo algorithms in parallel or, more exactly, grid computing environments.
75	-	BIGDATA: Mid-Scale: DCM: A Formal Foundation for Big Data Management	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	The ability to analyze massive-scale datasets has become an important tool both in industry and in the sciences and many systems have recently emerged to support it. However, effective methods for deep data analytics are currently high-touch processes: they require a highly specialized expert who thoroughly understands the application domain and pertinent disparate data sources and who needs to perform repeatedly a series of data exploration, manipulation and transformation steps to prepare the data for querying, machine learning or data mining algorithms. This project explores the foundations of big data management with the ultimate goal of significantly improving the productivity in big data analytics by accelerating the bottleneck step of data exploration. The project integrates two thrusts: a theoretical study, which leads to new fundamental results regarding the complexity of various new (ad hoc) data transformations in modern massive-scale systems, and a systems study, which leads to a multi-platform software middleware for expressing and optimizing ad hoc data analytics techniques. The middleware is designed to augment and integrate existing analytics solutions in order to facilitate and improve methods of interest to the community and compatible with many existing platforms.The results of this project will make it easier for domain experts to conduct complex data analysis on big data and on large computer clusters. All research results will be released in a middleware package layered on top of existing big-data systems. The middleware includes all the new algorithms, optimization techniques, fault-tolerance and skew mitigation mechanisms, and generalized aggregates developed during the project. In addition, the project develops and deploys a Web-based query-as-a-service interface to the new middleware. The project Web site (http://myriadb.cs.washington.edu) provides access to the software, additional results and information. Project results will be included in educational and outreach activities in big data analytics, including new curricula at the undergraduate, graduate, and professional levels.

			DateOperator=&ExpDateOperator=		
76	-	CAREER: Uncertainty Quantification and Big Data Analysis in Interconnected Systems: Algorithms, Computations, and Applications	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DMS	Uncertainty quantification (UQ) and big data analysis have received increasing attention in recent years. Extensive research effort has been devoted to these topics, and novel numerical methods have been developed to efficiently deal with large-scale data sets and complex problems with uncertainty. Both UQ and big data analysis enable us to better understand the impacts of various uncertain inputs (boundary and initial data, parameter values, geometry, network etc.) to numerical predictions. UQ and big data analysis are thus critical to many important practical problems such as climate modeling, weather prediction, ocean dynamics, and smart grids. As the data size and dimensions of parameter space increase, one of the biggest challenges in UQ computations and big data analysis is the computational cost for analyzing the data and running the simulations. For large-scale complex interconnected systems, deterministic simulations can be very time-consuming, and conducting UQ simulations further increases the simulation cost and can be prohibitively expensive. This project aims to address these critical challenges. A novel set of highly efficient UQ and big data analysis algorithms will be developed to make big data analysis and UQ simulations amenable for large-scale complex interconnected systems. The new algorithms will significantly advance the current state of the art of UQ and big data analysis methods. The project also integrates educational opportunities, including exposing a range of undergraduate students to UQ and big data, giving graduate students the advanced skills needed to apply them, and mentoring Ph.D. students to be leaders in UQ and big data education and research. The approach under development in this research project is based on scalable algorithms for multivariate Bayesian-treed Gaussian process and power network reduction; high-dimensional UQ algorithms; dynamic state estimation and model calibration for non-Gaussian noisy data; and advanced stochastic contingency analysis. The new algorithms will be based on building multi-fidelity models in both network models and probability space. Such algorithms can accommodate big data in linear time. In addition, while current contingency analysis allows only assessment of a static power grid status without considering uncertainty, the new approaches will allow analysis of contingency dynamically and probabilistically for cascade failures. The new algorithms will allow investigators to establish an efficient framework to rigorously quantify the uncertainty, analyze big data, and endow smart grid simulations with a composite error bar.

77	-	Conference: Drawing Causal Inference from Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	A conference titled "Drawing Causal Inference from Big Data" will be held March 26 and 27, 2015, in the National Academy of Sciences auditorium in Washington DC. The purpose of this conference is to present state-of-the-art approaches to the problem, and to bring together leading experts, both the featured speakers and other experts, who will generate progress through their interactions. In many respects the subject of this conference is in its infancy because the many methods that have been developed and used for causal inference in small data do not scale up, because Big Data is often collected in the field in uncontrolled fashion, and because of the sheer size of the data that, contrary to popular belief, make it more rather than less difficult to identify causal effects. The problems in dealing with Big Data are in good part rooted in the limitations of human cognition, so ongoing efforts are aimed at the development of computational algorithms. However it is likely that computational techniques are best viewed as augmenting rather than replacing human insight: Current algorithms can find complex patterns and associations but most are not aimed to discover causal explanations. The conference also addresses the appropriate way to define causality in large data collected from chaotic and noisy systems, and the way to find causes that lie outside the measured variables. For example a correlation observed in a health survey based on genetic mapping might be due to an unmeasured environmental factor such as poverty. The subject of the conference is of vital and current interest to every field of study, business, and government agencies. Our society has developed methods of collecting and storing enormous amounts of data, and is increasingly doing so. The data can arrive from controlled experiments, but most often comes from relatively uncontrolled field observations, such as those from social networks, human medical and genetic measurements, and patterns of purchases. The amount of data has far outstripped our ability to discern what important patterns are in the data, and most important, what explains those patterns. In a typical large database there are huge number of variables that can be measured, and virtually uncountable numbers of correlations between different subgroups of those variables. There are enormous potential benefits to science, business, government, and society if the critical patterns in Big Data can not only be ascertained but explained. Explanation is the goal of this conference, represented by the phrase, "drawing causal inference." The most pressing questions we face are causal in nature. In health we might observe that a particular treatment is associated with a decrease of cancer deaths, but need to know if the treatment is the cause of the decrease. In education we might observe that students held back in early grades tend to drop out of high school, but need to know if the treatment causes that result.

78	-	REU Site: BIGDatA - Big Data Analytics for Cyber-Physical Systems	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	NON-TECHNICAL SUMMARYThis Research Experience for Undergraduate (REU) site will support NSF's mission to promote progress of science by introducing big data analytics in Cyber-physical Systems (CPS) to undergraduate students, helping advance the state of art, and preparing them for the future scientific workforce. The site's objectives include (i) motivating and solidifying students' interest in Computer Science (CS) in general, and big data analytics in particular; (ii) providing students with problem solving skills for conducting research in big data analytics and for presenting scientific findings verbally and in writing; (iii) equipping students with working knowledge of applying data management and data analytics techniques, by involving students in research projects related to different aspects of CPS; (iv) enhancing self-confidence and self-efficacy of participants, by creating a sense of belonging to a diverse community; and (v) broadening participation in computing, and contributing to diversify the computing workforce. The research outcome will advance the state of the art of big data analytics and push the research agenda of utilizing big data techniques for CPS. The REU site will reach undergraduate students beyond the New Mexico State University (NMSU) campus and train workforce for big data analytics in CPS. Special emphasis will be placed on recruiting underrepresented students nationwide. The participating students will be mentored by the site researchers to disseminate their research projects' findings via professional conferences and through the REU site website. TECHNICAL SUMMARY: The goal of this REU site initiative is to inspire and prepare undergraduate students to pursue careers in STEM with a focus on big data analytics for Cyber-physical Systems (CPS). PI's propose research projects to explore big data analytics for CPS in three intertwined layers: (L1) systems and architecture, (L2) models and algorithms, and (L3) visualization, spanning across four CPS application areas including smart grids, wireless sensor networks, smart homes for elderly and disabled, and disaster response. The planned student activities include: (i) team-based research activities on focused research projects, (ii) creation of cohorts, which will engage in training workshops, to develop research skills and to prepare for graduate schools, (iii) field trips to companies, and local and national labs to broaden students' research horizon, (iv) workshop and conference participation to present research findings, and (v) mentoring by faculty members and interaction with other student researchers. The proposed research outcome has the potential to advance the operation, protection, and utility of CPS. The research activities in the projects will equip student participants with skills and knowledge related to both big data analytics and CPS. The training workshops will help students develop research skills and their career paths. The field trips will broaden participants' understanding about data analytics research for CPS.

79	-	Workshop on Big Data Benchmarking: THE BIGDATA TOP100 LIST	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	OCI 1338373Chaitanya Baru - U. of California at San DiegoWorkshop on Big Data Benchmarking: The BigData Top 100 ListThe Workshop series on Big Data Benchmarking (WBDB) was created in 2012 by the Center for Large-scale Data Systems Research (CLDS) at the San Diego Supercomputer Center, UC San Diego to foster a community activity in the area of benchmarking of big data systems. The first workshop was held in May 2012 in San Jose, California and was sponsored by the National Science Foundation along with a few industry sponsors. Subsequent meetings followed, and a second workshop held in December 2012 in Pune, India. The meetings and the second workshop substantiated the initial ideas for a big data benchmark leading to the concept of the BigData Top100 List for ranking big data systems according to performance on a specified big data workload, while also recording and reporting system efficiency, for example, in terms of price/performance.This current action is for the Third Workshop on Big Data Benchmarking workshop to be held on July 16-17, 2013 in Xi'an, China. The local hosts for this meeting will be the Xi'an University of Posts and Telecommunications and the Shanxi HPC Center. Intellectual Merit: While a mature transaction processing and data management industry has been established over the last two decades based primarily on relational database (RDBMS) technologies, the emergence of the Big Data phenomenon characterized by the so-called "3Vs" - volume, velocity, and variety) of data along with the need for agile development of data-driven applications, has introduced a new set of challenges. A variety of new techniques and technologies have emerged to address these challenges. A big data benchmark could, indeed, be specified with the following components: (i) a definition for a synthetic benchmark dataset with a well-defined and well-specified data generation procedure; (ii) a representative workload for big data applications; and (iii) a set of metrics, run rules and full disclosure reports for fair comparisons among technologies and platforms. These results would then form the basis of the BigData Top100 List.Broader Impacts: The first two WBDB meetings have had a total of about 100 (invited) attendees from about 60 different organizations from industry and academia. The BigData Top100 List initiative was officially introduced at a dedicated session at the Strata Conference in Santa Clara, on February 28th, 2013. This is a community effort and the benchmark will be released via the bigdatatop100.org website.

80	-	Full Scale Development: Global Soundscapes! The Big Data, Big Screens, Open Ears Project	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DRL	The Global Soundscapes! Big Data, Big Screens, Open Ears Project uses the new science of soundscape ecology to design a variety of informal science learning experiences that engage participants through acoustic discovery Soundscape ecology is an interdisciplinary science that studies how humans relate to place through sound and how humans influence the environment through the alteration of natural sound composition. The project includes: (1) an interface to the NSF-funded Global Sustainable Soundscapes Network, which includes 12 universities around the world; (2) sound-based learning experiences targeting middle-school students (grades 5-8), visually impaired and urban students, and the general public; and(3) professional development for informal science educators. Project educational components include: - the first interactive, sound-based digital theater experience; - hands-on Your Ecosystem Listening Labs (YELLS), a 1-2 day program for school classes and out-of school groups; - a soundscape database that will assist researchers in developing a soundscape Big Database; and - iListen, a virtual online portal for learning and discovery about soundscape. The project team includes Purdue-based researchers involved in soundscape and other ecological research; Foxfire Interactive, an award-winning educational media company; science museum partners with digital theaters; the National Audubon Society and its national network of field stations; the Perkins School for the Blind; and Multimedia Research (as the external evaluator).

81	-	Big Data PI Workshop	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	This workshop will bring together, for the first time, all of the PIs of the NSF BIGDATA program, at a venue in the Washington DC/Northern Virginia area. The goals of the NSF BIGDATA program are to "develop new methods for deriving knowledge from data; construct new infrastructure to manage, curate and serve data to communities; and forge new approaches for associated education and training." Given the range of possible activities, there is a need and opportunity to introduce the funded project PI's to the breadth of work funded under the Big Data Research Initiative and to new potential research partners. The workshop is by invitation, and open to researchers who are currently funded through the NSF BIGDATA program. The goal of this two-day workshop is to bring together PIs/co-PIs with government and industry invitees to discuss current research, as well as identify and elaborate on new, emerging challenges, and future directions.This workshop will bring together experts in different disciplines comprising the Big Data community, funded by NSF. The meeting will identify gaps, needs, resources, and barriers related to Big Data research, and also develop a forward-looking research agenda that will identify future directions in this area. The workshop will provide current PIs with opportunities to showcase and discuss their current work; establish new collaborations; develop knowledge of available research and infrastructure resources; and gain insight into new areas of research. The broader impact of this meeting will be the ability to more broadly share research results in the foundational as well as application aspects of big data in society.
82	-	RI: Medium: Collaborative Research: Next-Generation Statistical	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	IIS	This project develops a new generation of optimization methods to address data mining and knowledge discovery challenges in large-scale scientific data analysis. The project is constructed in the context that modern computing architectures are enabling us to fit complex statistical models (Big Models) on large and complex datasets (Big Data). However, despite significant progress in each subfield of Big Data, Big Model, and modern computing architecture, we are still lacking powerful optimization techniques to effectively integrate these key components.One important bottleneck is that many general-purpose optimization methods are not specifically

		Optimization Methods for Big Data Computing	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		designed for statistical learning problems. Even some of them are tailored to utilize specific problem structures, they have not actually incorporated sophisticated statistical thinking into algorithm design and analysis. To tackle this bottleneck, the project extends traditional theory to open new possibilities for nontraditional optimization problems, such as nonconvex and infinite-dimensional examples. The project develops deeper theoretical understanding of several challenging issues in optimization (such as nonconvexity), develops new algorithms that will lead to better practical methods in the big data era, and demonstrates the new methods on challenging bio-informatics problems.The project is closely related to NSF's mission to promote Big Data research, and will have broad impacts. In the Big Data era, we see an urgent need for powerful optimization methods to handle the increasing complexity of modern datasets. However, we still lack adequate methods, theory, and computational techniques. By simultaneously addressing these aspects, this project will deliver novel and useful statistical optimization methods that benefit all relevant scientific areas. The project will deliver easy-to-use software packages which directly help scientists to explore and analyze complex datasets. Both PIs will also design and develop new classes to teach modern techniques in handling big data optimization problems. All the course materials - including lecture notes, problem sets, source code, solutions and working examples - will be freely accessed online. Moreover, both PIs will write tutorial papers and disseminate the results of this research through the internet, academic conferences, workshops, and journals. Through senior theses and potentially the REU (Research Experiences for Undergraduates) program, the proposed project will also actively include undergraduates and engage under-represented minority groups.To achieve these goals, this project develops (i) a new research area named statistical optimization, which incorporates sophisticated statistical thinking into modern optimization, and will effectively bridge machine learning, statistics, optimization, and stochastic analysis; (ii) new theoretical frameworks and computational methods for nonconvex and infinite-dimensional optimization, which will motivate effective optimization methods with theoretical guarantees that are applicable to a wide variety of prominent statistical models; (iii) new scalable optimization methods, which aim at fully harnessing the horsepower of modern large-scale distributed computing infrastructure. The project will shed new theoretical light on large-scale optimization, advance practice through novel algorithms and software, and demonstrate the methods on challenging bio-informatics problems.

83	-	RI: Medium: Collaborative Research: Next-Generation Statistical Optimization Methods for Big Data Computing	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	This project develops a new generation of optimization methods to address data mining and knowledge discovery challenges in large-scale scientific data analysis. The project is constructed in the context that modern computing architectures are enabling us to fit complex statistical models (Big Models) on large and complex datasets (Big Data). However, despite significant progress in each subfield of Big Data, Big Model, and modern computing architecture, we are still lacking powerful optimization techniques to effectively integrate these key components.One important bottleneck is that many general-purpose optimization methods are not specifically designed for statistical learning problems. Even some of them are tailored to utilize specific problem structures, they have not actually incorporated sophisticated statistical thinking into algorithm design and analysis. To tackle this bottleneck, the project extends traditional theory to open new possibilities for nontraditional optimization problems, such as nonconvex and infinite-dimensional examples. The project develops deeper theoretical understanding of several challenging issues in optimization (such as nonconvexity), develops new algorithms that will lead to better practical methods in the big data era, and demonstrates the new methods on challenging bio-informatics problems.The project is closely related to NSF's mission to promote Big Data research, and will have broad impacts. In the Big Data era, we see an urgent need for powerful optimization methods to handle the increasing complexity of modern datasets. However, we still lack adequate methods, theory, and computational techniques. By simultaneously addressing these aspects, this project will deliver novel and useful statistical optimization methods that benefit all relevant scientific areas. The project will deliver easy-to-use software packages which directly help scientists to explore and analyze complex datasets. Both PIs will also design and develop new classes to teach modern techniques in handling big data optimization problems. All the course materials - including lecture notes, problem sets, source code, solutions and working examples - will be freely accessed online. Moreover, both PIs will write tutorial papers and disseminate the results of this research through the internet, academic conferences, workshops, and journals. Through senior theses and potentially the REU (Research Experiences for Undergraduates) program, the proposed project will also actively include undergraduates and engage under-represented minority groups.To achieve these goals, this project develops (i) a new research area named statistical optimization, which incorporates sophisticated statistical thinking into modern optimization, and will effectively bridge machine learning, statistics, optimization, and stochastic analysis; (ii) new theoretical frameworks and computational methods for nonconvex and infinite-dimensional optimization, which will motivate effective optimization methods with theoretical guarantees that are applicable to a wide variety of prominent statistical models; (iii) new scalable optimization methods, which aim at fully harnessing the horsepower of modern large-scale distributed computing infrastructure. The project will shed new theoretical light on large-scale optimization, advance practice through novel algorithms and software, and demonstrate the methods on challenging bio-informatics problems.

84	-	CAREER: High Performance Spatial Queries and Analytics for Spatial Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	The rise of big data is changing our way of thinking about the world by providing new insights and creating new forms of value. The challenges for big data come not only from the volume but also the complexity, such as the multi-dimensional nature of spatial data. In this CAREER project, we will deliver a scalable and efficient spatial big data processing system that can take advantage of fast increasing processing power of computers and their latest supporting programming environments. This software can be used for a variety of applications in medical image processing and in GIS (Geographical Information Systems), e.g., for city planning, transportation planning, disaster response, military planning.The fundamental goal of this CAREER project is to address the research challenges for delivering a high performance software system for spatial queries and analytics of spatial big data on MapReduce and CPU-GPU hybrid platforms, promote the use of the created open source software to support problem solving in multiple disciplines, and educate the next generation workforce in big data. Specifically, the following research aims will be pursued in this project: 1) Create new spatial data processing methods and pipelines with spatial partition level parallelism through MapReduce and propose multi-level indexing methods to accelerate spatial data processing; 2) Research two critical components to enable data parallelism: effective and scalable spatial partitioning in MapReduce, and query normalization methods for partition effect; 3) Research efficient GPU-based spatial operations to support object level and intra-object level parallelism, and integrate them into MapReduce pipelines; 4) Investigate optimization methods for data processing pipelines, data skew mitigation, and CPU/GPU resource coordination in MapReduce; and 5) Provide declarative spatial queries and create a query translator to automatically translate the queries into MapReduce applications.The project will provide a high performance scalable spatial computing infrastructure to be deployed by researchers and application users world-wide from various disciplines, and the source codes will be made open source and fully available. The project will provide a strong foundation to solve spatial big data problems such as location based services, remote sensing based applications, and map based applications. It will also enable the fast solving of scientific problems such as pathology imaging at large scale. The education activities include a revised undergraduate course with a new spatial big data theme, a revised graduate course with a focus on big data management, involvement of undergraduate, graduate and underrepresented students in research, symposia and science projects for K-12 students, and a software infrastructure to support the education. For further information see the project web site: http://fushengwang.net/hadoop-gisKeywords: spatial big data, MapReduce, CPU-GPU, spatial queries, spatial analytics
85	-	RAPID: Socio-technical systems and Big Data Analytics in the Ebola	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	IIS	In response to the 2014 Ebola virus outbreak in western Africa, dozens of humanitarian relief organizations as well as the CDC and U.S. military are providing medical assistance or logistics support to the relief effort. At the same time, a diverse range of volunteer technical communities (VTCs) and academics, as well as the humanitarian relief organizations themselves, are attempting to make use of "big data" to improve the response. These big data analyses are based on diverse data from diverse sources, including call records from mobile phone companies, health worker inventories from ministries, and daily case-data reports aggregated from multiple organizations. The

		Response	anization=&PISta te=&PIZip=&PICo untry=&ProgOrg anization=&Prog EleCode=&Boole anElement=All& ProgRefCode=&B ooleanRef=All&P rogram=&ProgOf ficer=&Keyword =%22Big+Data% 22&AwardTitleO nly=true&Award NumberOperato r=&AwardAmou nt=&AwardInstr ument=&ActiveA wards=true&Ori ginalAwardDate Operator=&Start DateOperator=& ExpDateOperato r=		analyses have generated outputs such as visually appealing maps and predictions of outbreak trajectories. However, precisely how, when and where these analyses can be used effectively by response organizations are still open questions. The project will develop knowledge to guide response organizations interested in leveraging existing and emerging big data from a variety of sources (response organizations, firms, government, individuals), which in turn may improve the speed, quality, and efficiency of crisis response. The research team of computer and social scientists will partner with a consultant with expertise in crisis information management deployed in the Ebola response. They will examine both organizational and technical dimensions of the use of big data analytics in the Ebola response organizations, carrying out a series of interviews to investigate how and where data is used (field, headquarters, or both) and the work involved to make big data analyses usable in the decision making of response organizations. The results will inform the development of a socio-technical systems framework to explain what makes big data analyses useable. The social dimensions of the framework will include the response context as well as decision making processes. The technical dimensions will include data availability, data analyses and output formats. In the process of developing this socio-technical framework, this research will identify mechanisms for matching organizational needs with big data analyses. More importantly, however, by identifying these mechanisms, the research will shed light on the fundamental roles of multi-level governance and articulation work in making effective use of big data analyses. The multi-level approach helps explain and predict the location in an organization?s hierarchy of technical and decision making expertise. Within these levels, a focus on articulation work helps specify the necessary tasks for using data in a highly dynamic environment. The project will extend the scholarship in the crisis informatics sub-field of big data beyond its current focus on social media data and provide clarification of "last mile" issues in big data aimed at ensuring the usefulness of output.
86	-	BD Hubs: Collaborative Proposal: SOUTH: A Big Data Innovation Hub for the South Region	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&	IIS	Big Data (BD) is a prominent, rapidly emerging discipline with far-reaching scientific and economic potential. While significant fundamental research has already taken place, there is a gap in the translation of BD research findings into economic growth and end-user impact. In addition, big data often arises in the practice of industry and government, and in broad geographic settings, making multi-stakeholder partnerships vital to exploiting its full potential. To successfully address these challenges, academic institutions from 16 states in the South region and the District of Columbia, in partnership with collaborators from industry, government, and non-profit organizations, will establish the South Big Data Regional Innovation Hub (South BD Hub) with lead institutions, Georgia Institute of Technology (GT) and the University of North Carolina at Chapel Hill (UNC-CH). The South BD Hub will serve as the primary vehicle for interdisciplinary, multi-stakeholder partnerships designed to pursue BD projects of interest to the South region and the nation as a whole. It will collectively nurture a BD innovation ecosystem, promote collaboration, provide essential workforce training, and serve as a national asset.The proposed Hub will establish a community-driven governance plan for long-term sustainability, guided by a Steering Council composed of

			ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		stakeholders that reflect the breadth, involvement, organizational nature, geography, and diversity of BD Hub interests. An Executive Committee and several Standing Committees also will be established. Birds of a Feather groups and more formal Working Groups may be established in targeted areas over the award period as needed. The Hub will have dual locations in Atlanta and Chapel Hill, with co-Executive Directors broadly accountable to the South BD Hub partners. GT and UNC-CH will act as host institutions providing logistics, operational, and other support for creation and long-term sustenance of the hub. Many partner organizations in the South region are committed to make diverse contributions to the Hub, including BD science "spoke" projects chosen by South region stakeholders. Initial spoke areas include: 1) Healthcare; 2) Coastal Hazards; 3) Industrial Big Data; 4) Materials and Manufacturing; and 5) Habitat Planning. To support these and other activities, the Hub also will launch cross-cutting spoke projects in BD Sharing and Infrastructure and BD Economics, Privacy, and Policy Issues. The national need for data scientists will also be addressed, through training programs, workshops, online resources, and other workforce development activities.For further information see the project web site: southbdhub.org
87	-	MRI: Acquisition of Big Data Training and Research Laboratory	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleO	CNS	This project, procuring and acquiring a large computing cluster appropriate for Big Data research, aims to enable research in a number of fields of national concern such as bioinformatics, ocean energy, social media mining, environmental and climate modeling, image processing and analysis emergency response, health and medical informatics, national security, infrastructure maintenance and reliability, law enforcement, commerce, and manufacturing. Of particular note are seven projects spanning a wide range of application domains. These include the- Extension of LexisNexis's High Performance Computing Cluster (HPCC) platform to incorporate a Wider Range of Algorithms;- Analysis of Big Data for Bioinformatics;- Use of Machine Condition Monitoring and Prognostic Health Monitoring to Improve Ocean Turbine Reliability;- General Challenges found when Mining Streaming data;- Analysis of Underwater Acoustic Signals for Tasks such as Unexploded Ordinance Detection; - Use of Machine Learning to Improve Video Compression Schemes; and- Challenges of Climate Modeling.All these demand the advanced computation resources under this acquisition. The study of Big Data encompasses the analysis of extremely large datasets, building models which are able to incorporate vast numbers of instances and features in order to make reliable predictions and connections.Maintaining and promoting the growth of Big Data has become an essential activity to ensure that the problems that now seem insurmountable may be solved tomorrow. Two aspects of this growth are developing and providing courses for students focused on the tools and techniques necessary for Big Data research, and more focused training in these tools for existing researchers whose area of expertise lies in other aspects of research. Best practices will be followed for cultural diversity, involving students and researchers

			nly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		from underrepresented groups.
88	-	BIGDATA: F: New Algorithms of Online Machine Learning for Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	IIS	This project is developing innovative, theoretically rigorous algorithms to learn from continuously arriving (streaming) data. Specific challenges addressed are class imbalance (one of the concepts to be learned is very rare, as in disease detection), cost constraints on both obtaining features (e.g., computationally expensive image processing), and cost constraints on obtaining class labels (e.g., human annotation.) The algorithms developed in this project make it possible to effectively address big data challenges in streaming data due to increased complexities in various aspects such as heavily imbalanced data distributions, ultrahigh dimensional features, a large number of labels, highly complex constraints, etc. The project will also contribute to training future professionals in big data analytics, including participation in the University of Iowa's undergraduate summer research program and high school student training program.Most work devoted to online learning algorithms and their analysis were developed with the goal of minimizing a symmetric measure (e.g., the classification error) and without considering practical constraints arising in big data. This project addresses imbalanced data by developing online learning algorithms for minimizing asymmetric measures including F-score, area under the ROC curve, and area under precision and recall curve. Convex or non-convex surrogate loss functions that well-approximate these asymmetric measures are constructed and minimized in an online fashion. The project also develops online algorithms under three types of constraints arising in big data context namely constraints on computing costs, on query costs, and complex inequality constraints, by exploring techniques in randomized algorithms, active learning and convex optimization. The developed algorithms are being evaluated in real applications including biomedical semantic indexing, social media mining, and image annotation.

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
89	-	BIGDATA: F: DKM: Collaborative Research: Making Big Data Active: From Petabytes to Megafolks in Milliseconds	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	A wealth of digital information is being generated daily through social networks, blogs, online communities, news sources, and mobile applications in an increasingly sensed world. Organizations and researchers recognize that tremendous value and insight can be gained by capturing this emerging data and making it available for querying and analysis. First-generation Big Data management efforts have been passive in nature -- queries, updates, and/or analysis tasks were mainly scaled to handle very large volumes of data. In contrast, this project will develop new techniques for continuously and reliably capturing Big Data collections (arising from social, mobile, Web, and sensed data sources) and will enable timely delivery of the right information to the relevant end users. In short, this project will provide a scalable foundation for moving from Big Passive Data to Big Active Data. Techniques should be developed to enable the accumulation and monitoring of petabytes of data of potential interest to millions of end users; when "interesting" new data appears, it should be delivered to end users in a time frame measured in (100's of) milliseconds. This project will build such an Active Big Data Management system and make it available as open source to the community. Students will be trained in technologies related to Big Active Data management and applications; such training is critical to addressing the information explosion that social media and the mobile Web are driving today. The general-purpose foundation for active information dissemination from Big Data will have broader impacts in areas such as public safety and public health. There are many challenges involved in building a foundation for Big Active Data. On the "data in" side, these include resource management in very large scale, LSM-based storage systems and the provision of a highly available, elastic facility for fast data ingestion. On the "data processing" side, challenges include the parallel evaluation of a large number of declarative data subscriptions over multiple) highly partitioned data sets. Amplifying this challenge is a need to efficiently support spatial, temporal, and similarity predicates in data subscriptions. Big Data also makes result ranking and diversification techniques critical in order for large result sets to be manageable. On the "data out" side, challenges include the reliable and timely dissemination of data of interest to a sometimes-connected subscriber base of unprecedented scale. As a software base, this project will be jump-started by using AsterixDB(http://asterixdb.ics.uci.edu/), an open-source Big Data Management System that supports the scalable storage, searching, and analysis of mass quantities of semi-structured data. For further information see the project web sites at https://www.ics.uci.edu/BigActiveData

			DateOperator=&ExpDateOperator=		and http://www.cs.ucr.edu/~tsotras/BigActiveData
90	-	BIGDATA: Collaborative Research: IA: F: Fractured Subsurface Characterization using High Performance Computing and Guided by Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperato	IIS	Natural fractures act as major heterogeneities in the subsurface that controls flow and transport of subsurface fluids and chemical species. Their importance cannot be underestimated, because their transmissivity may result in undesired migration during geologic sequestration of CO2, they strongly control heat recovery from geothermal reservoirs, and they may lead to induced seismicity due to fluid injection into the subsurface. Advanced computational methods are critical to design subsurface processes in fractured media for successful environmental and energy applications. This project will address the following key BIG data and computer science challenges: (1) Computation of seismic wave propagation in fractured media; (2) BIG DATA analytics for inferring fracture characteristics; (3) High Performance Computation of flow and transport in fractured media; and (4) Integration of data from disparate sources for risk assessment and decision-making. This will enable design of technologies for addressing key societal issues such as safe energy extraction from the surface, long-term sequestration of large volumes of greenhouse gases, and safe storage of nuclear waste. The project will provide interdisciplinary training for a team of graduate students and postdoctoral fellows. Outreach to high schools teachers and minorities through a planned workshop will inspire interest in environmental green-engineering, mathematics, and computational science. Numerous applications will benefit from this research, including Computer and Information Science and Engineering (CISE), Geosciences (GEO), and Mathematical and Physical Sciences (MPS).The proposed research will emphasize high performance computation (HPC) approaches for characterizing fractures using large subsurface seismic data sets, BIG data analytics for extraction of fracture related information from seismic inversion results and long-duration dynamic data, and advanced computational approaches for modeling flow, transport, and geomechanics in fractured subsurface systems. The specific objectives are to: Develop an efficient forward modeling algorithm for seismic wave propagation in fractured media using efficient computational schemes. Compute flow and transport in fractured media using an efficient computational scheme implemented on GPUs such as mimetic finite differences. Perform efficient multiphysics simulation of flow and geomechanics in fractured media. Integrate information from time-lapse seismic inversion and flow/transport simulation using novel statistical schemes. Joint inversion of seismic and fluid flow data and uncertainty quantification using efficient computational schemes. Develop and deploy a scalable hybrid-staging based substrate that can support targeted workflows using staging-based in-situ/in-transit approaches. Computational simulation is critical to design subsurface processes for successful environmental and energy applications. Project URL: http://csm.ices.utexas.edu/current-projects/

			r=		
91	-	BIGDATA: F: DKM: DKA: Big Data Modeling and Analysis with Depth and Scale	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	An important step in understanding large volumes of data is the construction of a model: a succinct but abstract representation of the phenomenon that produced the data. In order to understand a phenomenon, a data analyst needs to be able to propose a model, evaluate how the proposed model explains the data, and refine the model as new data becomes available. Statistical models, which specify relationships among random variables, have traditionally been used to understand large volumes of noisy data. Logical models have been used widely in databases and knowledge bases for organizing and reasoning with large and complex data sets. This project is aimed at developing a programming language and system for the creation, evaluation and refinement of combined statistical and logical models for the express purpose of understanding very large and complex data sets. Apart from its direct effect on model development for Big Data problems, the semantic foundations and scalable computing infrastructure resulting from this project is expected to directly impact the areas of system development and verification, planning, and optimization, with broad application in Science and Engineering. The tools developed in this project will facilitate the training of a new generation of scientists capable of transforming data into knowledge for use across disciplines. The project's education and outreach component is designed to train select undergraduate students on Big Data modeling and analysis via annual workshops and research mentorship; and graduate students via curriculum modifications including a specialization in Data Science.The project will develop Px, a language with well-defined declarative semantics, to support high-level model construction and analysis. Px will be capable of expressing generative and discriminative probabilistic and relational models, and the Px system will support complex queries over such models. The project will encompass three significant and complementary research directions, aimed at developing: (1) semantic foundations, including language constructs needed for succinct specification of complex models with rich logical and statistical structure; (2) scalable inference techniques combining exact and approximate methods, and query optimizations over combined logic/statistical models; and (3) programming extensions as well as static and dynamic analyses to support the creation and refinement of complex models. The Px language and system will be evaluated using two important and diverse application problems: (1) analysis and verification of infinite-state probabilistic systems, including parameterized systems, and (2) construction of phylogenetic trees from phenomic data, used in the Tree of Life project, for mapping the evolutionary history of organisms. The project is expected to make significant contributions towards creating a unifying framework combining probabilistic inference, logical inference, and constraint processing, with an emphasis on semantic clarity, efficiency, and scalability. The project will also demonstrate the practical utility of the proposed integrated framework by developing complex models from big data that take advantage of this technology in fundamental ways.

92	-	BD Hubs: Collaborative Proposal: SOUTH: A Big Data Innovation Hub for the South Region	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Big Data (BD) is a prominent, rapidly emerging discipline with far-reaching scientific and economic potential. While significant fundamental research has already taken place, there is a gap in the translation of BD research findings into economic growth and end-user impact. In addition, big data often arises in the practice of industry and government, and in broad geographic settings, making multi-stakeholder partnerships vital to exploiting its full potential. To successfully address these challenges, academic institutions from 16 states in the South region and the District of Columbia, in partnership with collaborators from industry, government, and non-profit organizations, will establish the South Big Data Regional Innovation Hub (South BD Hub) with lead institutions, Georgia Institute of Technology (GT) and the University of North Carolina at Chapel Hill (UNC-CH). The South BD Hub will serve as the primary vehicle for interdisciplinary, multi-stakeholder partnerships designed to pursue BD projects of interest to the South region and the nation as a whole. It will collectively nurture a BD innovation ecosystem, promote collaboration, provide essential workforce training, and serve as a national asset.The proposed Hub will establish a community-driven governance plan for long-term sustainability, guided by a Steering Council composed of stakeholders that reflect the breadth, involvement, organizational nature, geography, and diversity of BD Hub interests. An Executive Committee and several Standing Committees also will be established. Birds of a Feather groups and more formal Working Groups may be established in targeted areas over the award period as needed. The Hub will have dual locations in Atlanta and Chapel Hill, with co-Executive Directors broadly accountable to the South BD Hub partners. GT and UNC-CH will act as host institutions providing logistics, operational, and other support for creation and long-term sustenance of the hub. Many partner organizations in the South region are committed to make diverse contributions to the Hub, including BD science "spoke" projects chosen by South region stakeholders. Initial spoke areas include: 1) Healthcare; 2) Coastal Hazards; 3) Industrial Big Data; 4) Materials and Manufacturing; and 5) Habitat Planning. To support these and other activities, the Hub also will launch cross-cutting spoke projects in BD Sharing and Infrastructure and BD Economics, Privacy, and Policy Issues. The national need for data scientists will also be addressed, through training programs, workshops, online resources, and other workforce development activities.For further information see the project web site: southbdhub.org

93	-	BIGDATA: F: DKA: Collaborative Research: Theory and Algorithms for Parallel Probabilistic Inference with Big Data, via Big Model, in Realistic Distributed Computing Environments	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	This project develops a new framework that enables machine learning (ML) systems to automatically comprehend and mine massive and complex data via parallel Bayesian inference on large computer clusters. The research has a profound impact on the practice and direction of Big Learning. The developed technologies have a catalytic effect on both ML research and applications: ML scientists are able to rapidly experiment on novel, cutting-edge ML models with minimal programming effort, unhindered by the limitations of single machines. Researchers from other fields, like biology and social sciences, are able to run contemporary advanced ML methods that transcend the capabilities of simple models, yielding new scientific insights on data whose size would otherwise be daunting. Data scientists at small start-ups are able to conduct ML analytics with complex models, putting their capabilities on par with huge companies possessing dedicated engineering and infrastructure teams. Students and beginners are able to witness distributed ML in action with just a few lines of code, driving ML education to new heights. Technically, this research focuses on scaling up and parallelizing Bayesian machine learning, which provides a powerful, elegant and theoretically justified framework for modeling a wide variety of datasets. The research team develops a suite of complementary distributed inference algorithms for hierarchical Bayesian models, which cover most commonly used Bayesian ML methods. The project focuses on combining speed and scalability with theoretical guarantees that allow us to assess the accuracy of the resulting methods, and allow practitioners to make trade-offs between speed and accuracy. Rather than focus on a few disconnected models, the project develops techniques applicable to a broad spectrum of hierarchical Bayesian models, resulting in a toolkit of building blocks that can be combined as needed for arbitrary probabilistic models - be they parametric or nonparametric, discriminative or generative. This is in contrast to much existing work on parallel inference, which tends to focus on parallelization in a specific model and cannot be easily extended. The project provides a solid algorithmic foundation for learning on Big Data with powerful models. The research contributes to democratizing advanced and large-scale ML methods for broad applications, by offering the user and developer community a library of general-purpose parallelizable algorithms for working on diverse problems using computer clusters and the cloud, bridging the gap between practical needs from data and basic research in ML.

94	-	BIGDATA: Small: DA: Choosing a Needle in a Big Data Haystack	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	This big data project develops tools and algorithms to support users in the task of choosing one (or a few) object(s) from a very large set, particularly when there is a great deal of complex data on which to base this choice.Consider a traveler looking at hotel options on a travel site, a scientist trying to identify proteins to investigate further based upon the results of a high throughput experiment, or an intelligence analyst trying to identify suspected terrorists. In all of these cases we have a big data challenge in that there are likely to be hundreds, perhaps thousands or even millions, of options to choose from. While there are some criteria that can be expressed as simple functions of attribute values, e.g. price for a hotel room, these criteria capture only a part of the objective function. Other considerations, such as stylishness of a hotel, can be much harder to determine as a function of known attributes. The user may be compelled to examine candidate options individually. The computer's task is to help minimize the number of candidates examined, and to optimize the order of examination. This project examines how best to accomplish this task.echniques explored include supporting human specification of information need against a variety of big data sources and machine presentation of relevant results with the volume of big data. The broader impact of this project is in effectively harnessing the power of big data in a variety of applications, including business, science, and national defense.

95	-	EAGER: Learning Big Data Analytic Skills through Scientific Workflows	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	Big data analytics has emerged as a widely desirable skill in many areas. Although courses are now available on a variety of aspects of big data, there is a lack of a broad and accessible course that covers the variety of topics that concern big data analytics. As a result, acquiring practical data analytics skills is out of reach for many students and professionals, posing severe limitations to our ability as a society to take advantage of our vast digital data resources. The goal of this work is to develop curriculum materials for big data analytics to provide broad and practical training in data analytics in the context of real-world and science-grade datasets and data analytics methods. A key technical basis of the approach is the use of workflows that capture expert analytic methods that will be presented to users for practice with real-world datasets within pre-defined lesson units.The results of this work include lesson units for learning expert-level skills in big data analytics, a framework for non-programmers to understand basic concepts in big data analytics, and a hands-on workflow framework to learn by direct experimentation and exploration with scientific data.The work focuses on big data problems relevant to geosciences, such as water quality analysis, hydrology, lake ecosystem sustainability, climate science, and earth modeling.This project will supplement existing academic training materials in big data. The PIs will use real-world geosciences data and domain tasks. All the materials will be available under open source licenses. The proposed work will have great impact in the ability of students to pursue careers in big data analytics. The framework will be accessible to students who lack the programming skills required to assemble themselves end-to-end data analysis systems for experimentation and practical learning. The wide adoption of the proposed approach could ultimately lead to broad societal impact by changing the way people interact with data, learn from using scientific data, and their ability to participate in big data analysis.

96	-	MRI: Acquisition of a Big Data Compute Cluster for Interdisciplinary Research	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DBI	An award is made to the University of California, Riverside (UCR) to acquire a highly scalable Big Data compute cluster dedicated to long-term support for data-intensive NSF research. By freeing and expanding currently overbooked research compute resources, the proposed compute cluster will have a significant impact on training and educating graduate and undergraduate students, including a high percentage from populations traditionally underrepresented in STEM disciplines. Based on URM enrollment, UCR has been designated by the Dept. of Education as an Hispanic Serving Institution (HSI). The availability of adequate computing and its beneficial impact on research programs will serve to attract outstanding students to a large number of undergraduate and graduate programs, including successful NSF-Funded REU and IGERT programs. New courses planned in conjunction with system acquisition will train students in parallel computing concepts and offer expanded access to integrate Big Data computing into their research projects. The Big Data compute cluster of this project also will support several smaller biotechnology companies in California that already use IIGB's computer facility. Future users from external research institutions, minority-serving colleges and industrial partners, particularly start-ups, will be actively recruited to gain access to IIGB's computing resources. Combined with UCR's diverse ethnicity and research mission, this investment will benefit a wide array of research directions and technology-based economic development initiatives at UCR, an institution that serves as an important driver of economic development in California's Inland Empire.The goal of this project is to enable Big Data driven research at UCR to address grand challenge questions in a highly interdisciplinary environment. Questions include: How do different organism groups adapt to, and defend themselves against extreme environmental conditions or pathogens? How can more efficient and selective small molecules be developed to accelerate discovery-oriented chemical biology and genomics research? How can this knowledge translate into improved stress and pathogen tolerance, for example in crops to respond to global climate change and feed a growing world population? Recent advances in high-throughput and monitoring technologies offer, for the first time, novel methods to address these challenges systematically, comprehensively, and with unprecedented resolution. The new Big Data compute cluster substantially strengthens UCR's high-performance compute infrastructure and provides a critical enabling resource for UCR researchers from a broad spectrum of research specializations, including environmental science, chemical genomics, evolution, statistics, computational biology, and genome biology of multiple organism groups. Since this research relies on high-throughput and computational modeling approaches, its success and future growth is critically dependent on high-performance computer resources to manage and process large and rapidly increasing data sets. The system will be managed by experienced personnel in the Research Compute Facility of the Institute for Integrative Genome Biology (IIGB). The IIGB facility serves a broad user population distributed across departments from more than 50 research groups with 160 active users. The requested computing system therefore will reach a maximum number of NSF-funded investigators at UCR and constitutes a cost-effective investment of NSF and UCR funds.

97	-	An Undergraduate Workshop on "Big Data, Human Health and Statistics"	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DMS	This workshop is a concluding event to an Undergraduate Summer Institute on "Big Data, Human Health and Statistics" held at the Department of Biostatistics, University of Michigan, Ann Arbor from June 1-26 and is intended to be an event focusing on the undergraduates. The training of the next generation of quantitative scientists needs to change to meet the demands of the data. We define "Big Data" as datasets of enormous size and complexity (either in number of observations, and/or in the number/nature of predictors/outcomes). Classical theory, computation and intuition often fail for such irregular, sparse data sets of vast size. More training in data management, data storage, visualization, high dimensional statistics, optimization, causal methods, modeling sparse data, machine learning are needed to equip students to tackle these big data challenges. It is expected that the knowledge obtained from these massive heterogeneous data sources will inform prevention, screening, prognosis, and treatment of human diseases and play a major role in biology, medicine, and public health in the coming decade. This workshop lies in the intersection of Big Data, Human Health and Statistics. In this two-day symposium, the first day will feature talks by distinguished researchers in areas of relevance to Big Data, including mobile health, precision medicine, genomics, data visualization. There will be a poster session by undergraduate attendees and an oral presentation session by the undergraduate participants of the summer institute. The second day will be a professional development workshop for undergraduate attendees. Registration will be open to anyone interested (with a maximum limit of 120 participants). The grant will support participation cost of selected undergraduate attendees, faculty mentors and outside speakers.
98	-	CC*DNI Engineer: Big Data Enabler for the UW-DMZ	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	ACI	The University of Wyoming is developing an environment providing researchers and students with access to local, regional, national, and international computational, networking, and data storage resources needed to transform scientific discovery and education. The university is developing a perimeter network named UW-DMZ (University of Wyoming "science DMZ") that contains and exposes the university's external facing services to Internet2. UW-DMZ enhances Big Computing and Big Data based research by providing infrastructure needed to leverage high performance, dynamic network services. Research is informed by a wide variety of data collected from field work

			anization=&PISta te=&PIZip=&PICo untry=&ProgOrg anization=&Prog EleCode=&Boole anElement=All& ProgRefCode=&B ooleanRef=All&P rogram=&ProgOf ficer=&Keyword =%22Big+Data% 22&AwardTitleO nly=true&Award NumberOperato r=&AwardAmou nt=&AwardInstr ument=&ActiveA wards=true&Ori ginalAwardDate Operator=&Start DateOperator=& ExpDateOperato r=		and sensor networks, mined from databases, observed by lab experiments, or generated by simulations. The movement of data from their source to computational and visualization resources and then to the scientific world is critical.The CI Engineer provides essential knowledge for how to best use the UW-DMZ facilities to researchers and students who have not experienced very fast computer networking or Big Data environments. The CI Engineer provides short courses for students and researchers that provide deeper, enriching experiences. Through the university's significant research collaborations with Jackson State University on environmental hydrology, their students and faculty are trained in accessing and using computational and visualization resources and gain firsthand experience in how to facilitate scientific discovery through management and integration of Big Computing and Big Data.
99	-	CC*IIE Campus Design: Building The 10Gbps Network for Big Data in the Sciences	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&	ACI	This project expands access at St. Olaf College to massive, complex datasets and delivers the analytic power they require. The 2-year effort upgrades the current campus network so researchers in physics, biology, economics, computer science and mathematics can accelerate their projects and expand the range of datasets they investigate. In particular, greater network capacity advances the NSF-funded Center for Interdisciplinary Research and the CSinParallel initiative. "The Network for Big Data" also enhances formal instruction in data science across the STEM disciplines. Looking ahead, the Network for Big Data will enable St. Olaf researchers to access massive data sets generated via remote sensing methods, including Light Detection and Ranging (LIDAR)."The Network for Big Data" continues St. Olaf's tradition of taking innovative, cost-effective approaches to advanced computing. All data transfer nodes and Big Data enclave devices reside in a single virtual local area network to serve St. Olaf's science and mathematics data management zone. Once core and data center switches are consolidated in one enhanced switch, external connectivity, which includes Internet2, will increase to 10 Gbps. The fiber infrastructure connecting St. Olaf's network border to its core and to key STEM buildings is upgraded to support 10Gbps connections, with

			ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		IPv6 campus-wide deployment, and continuous performance monitoring performance and outcomes of these upgrades using a perfSONAR node.
100	-	CI-ADDO-NEW: ASTERIX: A Community Software Platform for Big Data Research, Analysis, and Management	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleO	CNS	Vast quantities of digital information are being generated today on a daily basis through social networks, blogs, online communities, news sources, and mobile applications as well as our increasingly sensed surroundings. Tremendous insight can be gained by storing and making such ?Big Data? available for exploration in a wide variety of domains. Likely beneficiaries include business, social sciences, public health, national security, political science, public safety, medicine, and government policy. Researchers exploring these benefits need software to manage and analyze Big Data, and researchers investigating algorithms and programming models for Big Data can benefit tremendously by being provided with shared building blocks to use as a foundation for their efforts.Over the past 3.5 years we have developed an initial version of AsterixDB, a powerful new Big Data Management System (BDMS) for scalably storing, managing, searching, and analyzing collections of Big Data using clusters of commodity computers. AsterixDB has a layered code base that consists of a scalable runtime platform (Hyracks), a model-neutral framework for parallel query compilation for Big Data (Algebricks), and the end-user-targeted AsterixDB BDMS itself. This NSF project is turning AsterixDB and its internal software stack into robust, supported, open-source resources for use by the Big Data applications and technology research communities. AsterixDB and its components will be helpful in training students in Computer Science and other data-related sciences, at universities everywhere, about Big Data technologies. This is critical for addressing the information explosion being brought to us courtesy of social media and the mobile Web.

			nly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
101	-	II-NEW: BOLD: Big Data and Optical Lightpaths-Driven Networked Systems Research Infrastructure	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	CNS	This project will deploy an optically networked systems research infrastructure named BOLD that integrates high-performance, low-power, optical networking devices and programmable packet switches to enable transformative and inter-disciplinary research. It enables experimental exploration of different architectural design choices and it will potentially suggest new approaches to integrated hardware-software designs. At the hardware level, it enables research on novel optical network devices that can provide powerful communication capabilities, possibly tailored to the specific needs of big data applications, as well as experimentation with the device prototypes under real application traffic. BOLD enables a broad range of transformative big data-driven research. At the system software level, it enables research on storage, network, and application control software that are designed from the ground up to coordinate for optimal performance. At the application level, BOLD motivates research on how fundamental algorithms for big data applications should be designed to leverage the new network capabilities. BOLD has the unique potential to bridge the gap between nano-photonics researchers, networked systems researchers, and big data application researchers, creating inter-disciplinary research opportunities. Furthermore, because BOLD is designed to operate alongside Rice?s existing NSF-funded computing infrastructures BOLD can support big data experiments at a substantial scale.Results from the inter-disciplinary research enabled by BOLD will lead to future big data processing system architectures that dramatically speed up a wide range of computational scientific discoveries. Because optical networking devices are unique in that they consume very little power, yet can support enormous data rates, BOLD will inspire a new class of high performance, high energy efficiency system architectures. Research enabled by BOLD could inform the design of future nation-wide networking infrastructures by showing how optical networks can be harnessed as shared ?cloud? resources. BOLD will serve as a platform for the training and education of numerous undergraduate and graduate students, including under-represented groups, in cutting edge big data-driven research.

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
102	-	BD Hubs: Collaborative Proposal: WEST: A Big Data Innovation Hub for the Western United States	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperato	IIS	The Big Data Innovation Hub for the Western United States will join stakeholders from academia, industry, non-profit institutions and the community who share common challenges and innovative approaches related to the acquisition, storage, analysis and integration of large or "messy" data, commonly referred to as Big Data. The West's Innovation Hub (Hub) will serve 13 states with Montana, Colorado and New Mexico marking the eastern boundary. This project will develop the organizational and governance structures for the Hub, and initiate efforts toward defining spoke activities for subsequent phases of the data innovation hubs program. The initial themes include Big Data technology, data-enabled scientific discovery and learning, managing natural resources and hazards, metro data science, and precision medicine. Partnerships fostered through the Hub will enable the use of Big Data to assess risks related to regional and long-term decisions. The Hub's structure will enable impact in later phases of the Hubs program that may include data-driven models for managing natural resources to tools for integrating self-collected patient data for more precise care options. Through coordination activities that inspire the action of its members, the Hub has the potential to facilitate the improved flow of commercial technologies in ways that maximize competitiveness for member organizations, such as universities, and vice versa: the Hub has the potential to expand the impact of its members' technologies through greater adoption or via start-ups. The Hub will be impactful by facilitating cross-discipline approaches to Big Data innovation and problem solving, influencing the next generation of thought leaders and data scientists. The partnerships enabled by the Hub will lead to professional certificate programs and student internships, creating a pipeline of graduates from partner institutions to impact corporations, public and governmental agencies, national labs, resource-planning agencies, and regulatory commissions.Project URL: BDHub.SDSC.edu

			r=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
103	-	BIGDATA: Collaborative Research: IA: F: Fractured Subsurface Characterization using High Performance Computing and Guided by Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Natural fractures act as major heterogeneities in the subsurface that controls flow and transport of subsurface fluids and chemical species. Their importance cannot be underestimated, because their transmissivity may result in undesired migration during geologic sequestration of CO2, they strongly control heat recovery from geothermal reservoirs, and they may lead to induced seismicity due to fluid injection into the subsurface. Advanced computational methods are critical to design subsurface processes in fractured media for successful environmental and energy applications. This project will address the following key BIG data and computer science challenges: (1) Computation of seismic wave propagation in fractured media; (2) BIG DATA analytics for inferring fracture characteristics; (3) High Performance Computation of flow and transport in fractured media; and (4) Integration of data from disparate sources for risk assessment and decision-making. This will enable design of technologies for addressing key societal issues such as safe energy extraction from the surface, long-term sequestration of large volumes of greenhouse gases, and safe storage of nuclear waste. The project will provide interdisciplinary training for a team of graduate students and postdoctoral fellows. Outreach to high schools teachers and minorities through a planned workshop will inspire interest in environmental green-engineering, mathematics, and computational science. Numerous applications will benefit from this research, including Computer and Information Science and Engineering (CISE), Geosciences (GEO), and Mathematical and Physical Sciences (MPS).The proposed research will emphasize high performance computation (HPC) approaches for characterizing fractures using large subsurface seismic data sets, BIG data analytics for extraction of fracture related information from seismic inversion results and long-duration dynamic data, and advanced computational approaches for modeling flow, transport, and geomechanics in fractured subsurface systems. The specific objectives are to: Develop an efficient forward modeling algorithm for seismic wave propagation in fractured media using efficient computational schemes. Compute flow and transport in fractured media using an efficient computational scheme implemented on GPUs such as mimetic finite differences. Perform efficient multiphysics simulation of flow and geomechanics in fractured media. Integrate information from time-lapse seismic inversion and flow/transport simulation using novel statistical schemes. Joint inversion of seismic and fluid flow data and uncertainty quantification using efficient computational schemes. Develop and deploy a scalable hybrid-staging based substrate that can support targeted workflows using staging-based in-situ/in-transit approaches. Computational simulation is critical to design subsurface processes for

			DateOperator=&ExpDateOperator=		successful environmental and energy applications. Project URL: http://csm.ices.utexas.edu/current-projects/
104	-	BIGDATA: F: DKA: Collaborative Research: Theory and Algorithms for Parallel Probabilistic Inference with Big Data, via Big Model, in Realistic Distributed Computing Environments	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	IIS	This project develops a new framework that enables machine learning (ML) systems to automatically comprehend and mine massive and complex data via parallel Bayesian inference on large computer clusters. The research has a profound impact on the practice and direction of Big Learning. The developed technologies have a catalytic effect on both ML research and applications: ML scientists are able to rapidly experiment on novel, cutting-edge ML models with minimal programming effort, unhindered by the limitations of single machines. Researchers from other fields, like biology and social sciences, are able to run contemporary advanced ML methods that transcend the capabilities of simple models, yielding new scientific insights on data whose size would otherwise be daunting. Data scientists at small start-ups are able to conduct ML analytics with complex models, putting their capabilities on par with huge companies possessing dedicated engineering and infrastructure teams. Students and beginners are able to witness distributed ML in action with just a few lines of code, driving ML education to new heights. Technically, this research focuses on scaling up and parallelizing Bayesian machine learning, which provides a powerful, elegant and theoretically justified framework for modeling a wide variety of datasets. The research team develops a suite of complementary distributed inference algorithms for hierarchical Bayesian models, which cover most commonly used Bayesian ML methods. The project focuses on combining speed and scalability with theoretical guarantees that allow us to assess the accuracy of the resulting methods, and allow practitioners to make trade-offs between speed and accuracy. Rather than focus on a few disconnected models, the project develops techniques applicable to a broad spectrum of hierarchical Bayesian models, resulting in a toolkit of building blocks that can be combined as needed for arbitrary probabilistic models - be they parametric or nonparametric, discriminative or generative. This is in contrast to much existing work on parallel inference, which tends to focus on parallelization in a specific model and cannot be easily extended. The project provides a solid algorithmic foundation for learning on Big Data with powerful models. The research contributes to democratizing advanced and large-scale ML methods for broad applications, by offering the user and developer community a library of general-purpose parallelizable algorithms for working on diverse problems using computer clusters and the cloud, bridging the gap between practical needs

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		from data and basic research in ML.
105	-	Creating Pathways for Big Data Careers	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DUE	The quantity of digital data created globally is growing exponentially, and workers in many professions need skills to organize and analyze the deluge of data that confronts them. To address the urgent need for workers who have "big data" skills, the Oceans of Data Institute (ODI) of the Education Development Center will work with four community colleges that are leaders in the area of big data, data science, and data analytics to create a career pathway model for big data careers, with special attention to the "middle skill" jobs that can be filled by graduates of community college programs. Subject matter experts representing a variety of professional fields will identify the entry-level, middle-level, and high-level big data skills, knowledge, and behaviors that are needed in the workplace, and these skills, knowledge, and behaviors will then be organized and aligned with big data jobs. ODI and partners Bunker Hill Community College (Massachusetts), Normandale Community College (Minnesota), Johnson County Community College (Kansas), and Sinclair Community College (Ohio) will work together to design the big data career pathway model. They will document pathways to these careers via a system of stackable credentials that identify associate degrees, certificates, and four-year programs leading to employment. Each college will adapt the pathway model to fit the big data labor demand in its geographic region. To build the capacity of faculty to teach big data skills, Bunker Hill Community College will leverage its relationship with the Broadening Advanced Technological Education Connections (BATEC) center to deliver appropriate professional development for faculty. ODI and the partner colleges will disseminate the project's products to a national audience of community colleges (including minority serving institutions and tribal colleges) and business/industry representatives hiring for big data positions.The methodology that the project team will use to define the "Middle-Skill Big Data-Enabled Specialist" (MSBDES) has been used successfully to define other emerging occupations, has been widely validated, and has been used internationally to identify the skills needed by front-line workers. The project's products will include an occupational profile of the MSBDES, which will articulate the big data skills, knowledge, and behaviors expected by industry. The profile will be paired with rubrics that describe proficient performance of the big data middle skills. The profile and performance rubrics will provide an assessment framework that will assist community college faculty in evaluating student proficiency. These materials, along with the stackable credentials, will guide other community colleges seeking to develop big data courses and programs.

106	-	BD Hubs: Collaborative Proposal: WEST: A Big Data Innovation Hub for the Western United States	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	The Big Data Innovation Hub for the Western United States will join stakeholders from academia, industry, non-profit institutions and the community who share common challenges and innovative approaches related to the acquisition, storage, analysis and integration of large or "messy" data, commonly referred to as Big Data. The West's Innovation Hub (Hub) will serve 13 states with Montana, Colorado and New Mexico marking the eastern boundary. This project will develop the organizational and governance structures for the Hub, and initiate efforts toward defining spoke activities for subsequent phases of the data innovation hubs program. The initial themes include Big Data technology, data-enabled scientific discovery and learning, managing natural resources and hazards, metro data science, and precision medicine. Partnerships fostered through the Hub will enable the use of Big Data to assess risks related to regional and long-term decisions. The Hub's structure will enable impact in later phases of the Hubs program that may include data-driven models for managing natural resources to tools for integrating self-collected patient data for more precise care options. Through coordination activities that inspire the action of its members, the Hub has the potential to facilitate the improved flow of commercial technologies in ways that maximize competitiveness for member organizations, such as universities, and vice versa: the Hub has the potential to expand the impact of its members' technologies through greater adoption or via start-ups. The Hub will be impactful by facilitating cross-discipline approaches to Big Data innovation and problem solving, influencing the next generation of thought leaders and data scientists. The partnerships enabled by the Hub will lead to professional certificate programs and student internships, creating a pipeline of graduates from partner institutions to impact corporations, public and governmental agencies, national labs, resource-planning agencies, and regulatory commissions.Project URL: BDHub.SDSC.edu

107	-	BD Hubs: Collaborative Proposal: WEST: A Big Data Innovation Hub for the Western United States	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	The Big Data Innovation Hub for the Western United States will join stakeholders from academia, industry, non-profit institutions and the community who share common challenges and innovative approaches related to the acquisition, storage, analysis and integration of large or "messy" data, commonly referred to as Big Data. The West's Innovation Hub (Hub) will serve 13 states with Montana, Colorado and New Mexico marking the eastern boundary. This project will develop the organizational and governance structures for the Hub, and initiate efforts toward defining spoke activities for subsequent phases of the data innovation hubs program. The initial themes include Big Data technology, data-enabled scientific discovery and learning, managing natural resources and hazards, metro data science, and precision medicine. Partnerships fostered through the Hub will enable the use of Big Data to assess risks related to regional and long-term decisions. The Hub's structure will enable impact in later phases of the Hubs program that may include data-driven models for managing natural resources to tools for integrating self-collected patient data for more precise care options. Through coordination activities that inspire the action of its members, the Hub has the potential to facilitate the improved flow of commercial technologies in ways that maximize competitiveness for member organizations, such as universities, and vice versa: the Hub has the potential to expand the impact of its members' technologies through greater adoption or via start-ups. The Hub will be impactful by facilitating cross-discipline approaches to Big Data innovation and problem solving, influencing the next generation of thought leaders and data scientists. The partnerships enabled by the Hub will lead to professional certificate programs and student internships, creating a pipeline of graduates from partner institutions to impact corporations, public and governmental agencies, national labs, resource-planning agencies, and regulatory commissions.Project URL: BDHub.SDSC.edu

108	-	RCN: SAVI: Adaptive Management and Use of Resilient Infrastructures in Smart Cities: Support for Global Collaborative Research on Real-Time Analytics of Heterogeneous Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Cities provide ready and efficient access to facilities and amenities through shared civil infrastructures such as transportation and healthcare. Making such critical infrastructures resilient to sudden changes, e.g., caused by large-scale disasters, requires careful management of limited and varying resources. The rapidly growing big data from both physical sensors and social media in real-time suggest an unprecedented opportunity for information technology to enable increasing efficiency and effectiveness of adaptive resource management techniques in response to sharp changes in supply and/or demand on critical infrastructures. Within the general areas of resilient infrastructures and big data, this project will focus on the integration of heterogeneous Big Data and real-time analytics that will improve the adaptive management of resources when critical infrastructures are under stress. The integration of heterogeneous data sources is essential because many kinds of physical sensors and social media provide useful information on various critical infrastructures, particularly when they are under stress. This Research Coordination Network (RCN) will promote meetings and activities that stimulate and enable new research on integration of heterogeneous physical sensor data and social media for real-time big data analytics in support of resilient critical infrastructures such as transportation and healthcare in smart cities. As first example, the RCN will support participation from young faculty attending the Early Career Investigators' Workshop on Cyber-Physical Systems in Smart Cities (ECI-CPS) at CPSweek (April of each year) and young faculty attending the Workshop on Big Data Analytics for Cyber-physical Systems (BDACPS). As a second example, the RCN will support contributions to a Special Track on Big Data Analytics for Resilient Infrastructures at the IEEE Big Data Congress. As a third example, the RCN will support participation in International meetings organized by other countries, e.g., Japan's Big Data program by Japan Science and Technology Agency (JST). The project will also maintain a repository of research resources. Concretely, the RCN will actively collect and make readily available public data sets (e.g., physical and social sensor data) and software tools (e.g., to support real-time big data analytics). The technologies and tools that arise from RCN-enabled research will be applied to socially and economically impactful areas such as reducing congestion and personalized healthcare in smart cities.

109	-	NeTS: Small: Big Data and Optical Lightpaths-Driven Software Defined Networking	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Science and engineering are increasingly relying on data and the ability to process a massive amount of data to solve hard problems and drive fundamental discoveries and innovations. Challenges arising from this trend are often referred to as "Big Data" problems. Examples of big data processing applications include seismic data analysis, data-intensive text processing, assembly of large genomes, machine learning, data mining, and social-network analysis. This project will investigate new directions in software defined networking (SDN) that are motivated by the networking challenges stemming from big data processing applications and by the potential benefits of using optical lightpaths for big data transport. The project will develop effective solutions for jointly configuring a rich set of optical devices and SDN switches to realize network services that meet the needs of big data applications. Specifically, the project will develop optical device resource allocation algorithms, topology design and routing algorithms, comparisons between greedy and guaranteed resource allocation policies, co-scheduling systems for traffic and network, techniques for data shuffle transmissions, and co-designed application and network controllers. The approaches, algorithms, and software developed by this project will be evaluated in a realistic experimental infrastructure called BOLD. The project may have far reaching societal impacts beyond the computing discipline. Results from the project can dramatically speed up a wide range of computational scientific discoveries. Optical networking devices consume very little power, yet can support enormous data rates; the project results could lead to a more environmentally sustainable future for the IT industry. The project activities will provide exciting opportunities for training and education of undergraduate and graduate students, and particularly under-represented minority students, in cutting edge big data-driven networking. Finally, software, data, and curriculum materials produced by this project will be disseminated as free open-source resources for the wider community's use.
110	-	BIGDATA: F: DKA: Usable Multiple Scale Big Data Analytics through	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	IIS	Gaining big insight from big data requires big analytics, which poses big usability problems. Analyses of big data often rely on several computational and statistical models that operate on multiple levels of data scale to discover and characterize noteworthy patterns. The models work jointly or in sequence to filter, group, summarize, and visualize big data so that analysts may assess the data. As a simple example in big text analytics, massive text is first sampled for relevant or representative words, then further reduced by a complex form of modeling (e.g., topic modeling), then visualized by applying a dimension reduction algorithm. As the size of data increases, so does the

		Interactive Visualization	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		number of models and, likewise, the need for human interaction in the analytical process. By interacting, humans include expert judgment into the analytical process, and efficiently explore and make sense of big data from varying perspectives. However, for a variety of reasons, interacting with any individual model is difficult, let alone a growing number of models. Thus, current human-computer-interaction research is merged with complex statistical methods and fast computation to develop a usable, multi-model analytic framework for big data. Wrapped in software, the framework will be accessible to both professional and student users alike; i.e., available to make new discoveries in current government and industrial big datasets, as well as, educate future analysts at the undergraduate and graduate levels given new teaching modules. The new analytic framework extends Visual-to-Parametric Interaction (V2PI) to Multi-scale V2PI (MV2PI). V2PI currently supports usable small-data analytics, and enables users to adjust model parameters by interacting directly with data in visualizations. That is, V2PI interprets visual interactions quantitatively to update underlying model parameters and produce new visualizations. MV2PI now links together several models that operate at multiple levels of data-scale in a unified interactive space. In MV2PI, small-scale data interactions in visualizations propagate to larger scale models (by inverting them and updating their parameters) and new visualizations are generated. In the text analytics example, if users drag several data points together to hypothesize a cluster, the inverted dimension reduction model computes updated dimension weights, queries relevant new hits at the large scale, identifies changed topics, and updates the layout to show big-data support for the new cluster. With MV2PI, users may interactively explore large-scale data and complex inter-relationships between models in real time, and in a usable fashion that directly supports their natural cognitive sensemaking process. Development of MV2PI involves: (1) formulation of an explicitly stated framework ; (2) creation of new interactive models (e.g., Interactive K-means and Interactive Latent Dirichlet Allocation) that cover different levels of scale and support MV2PI model inversion; (3) implementation of computational methods to support high-performance, real-time model updates; and (4) evaluation of MV2PI software framework for usability and effectiveness. The project web site (http://www.apps.stat.vt.edu/bava/mv2pi.html) will include information on MV2PI development, access to software, datasets, educational materials, and publications.
111	-	BDD: Disaster Preparation and Response via Big Data Analysis and Robust Networking	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&Boole	CNS	Disasters are events with dire consequences, requiring multiple-agency responses and resources beyond the capability of a single community. Natural disasters, such as the 2011 Great East Japan Earthquake, can threaten the lives of many people and cause inordinate economic losses. Communication is critical to disaster preparation, response, and recovery, but may be damaged during the disaster. In this project, researchers from the US and Japan study novel approaches to disaster preparation, response and recovery using survivable communication networks and big data analysis of social media data. This collaborative effort involves expertise in disaster research, social media mining and big data analysis, network science, wireless communications, and machine learning, to examine resilient network architecture and algorithms, data collection and analysis before the disaster, and decision making and information dissemination during the disaster. The resilient network incorporates both wired and wireless communications to deal with multiple disaster-induced failures, aiming for efficient algorithms serving emergency applications. State-of-the-art data collection and analysis techniques will help build an important knowledge base in

			anElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		proactive preparation for disasters. Real time decision making and information dissemination during a disaster can assist disaster response and recovery effectively. The proposed research aims to provide valuable guidance for disaster preparation, response, and recovery for both the US and Japan, and spearhead a new research direction in survivable communication network design and big data analysis. This project provides a conducive environment to further research collaboration of big data analysis and disaster relief between the US and Japan. Graduate students will be jointly trained in this international research project to actively collaborate in carrying out the proposed research tasks. Special efforts will be made to engage minority students and underrepresented groups.
112	-	BIGDATA: F: DKM: Collaborative Research: Scalable Middleware for Managing and Processing Big Data on Next Generation HPC Systems	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%	IIS	Managing and processing large volumes of data and gaining meaningful insights is a significant challenge facing the Big Data community. Thus, it is critical that data-intensive computing middleware (such as Hadoop, HBase and Spark) to process such data are diligently designed, with high performance and scalability, in order to meet the growing demands of such Big Data applications. While Hadoop, Spark and HBase are gaining popularity for processing Big Data applications, these middleware and the associated Big Data applications are not able to take advantage of the advanced features on modern High Performance Computing (HPC) systems widely deployed all over the world, including many of of the multi-Petaflop systems in the XSEDE environment. Modern HPC systems and the associated middleware (such as MPI and Parallel File systems) have been exploiting the advances in HPC technologies (multi/many-core architectures, RDMA-enabled networking, NVRAMs and SSDs) during the last decade. However, Big Data middleware (such as Hadoop, HBase and Spark) have not embraced such technologies. These disparities are taking HPC and Big Data processing into "divergent trajectories." The proposed research, undertaken by a team of computer and application scientists from OSU and SDSC, aim to bring HPC and Big Data processing into a "convergent trajectory." The investigators will specifically address the following challenges: 1) designing novel communication and I/O runtime for Big Data processing while exploiting the features of modern multi-/many-core, networking and storage technologies; 2) redesigning Big Data middleware (such as Hadoop, HBase and Spark) to deliver performance and scalability on modern and next-generation HPC systems; and 3) demonstrating the benefits of the proposed approach for a set of driving Big Data applications on HPC system. The

			22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		proposed work targets four major workloads and applications in the Big Data community (namely data analytics, query, interactive, and iterative) using the popular Big Data middleware (Hadoop, HBase and Spark). The proposed framework will be validated on a variety of Big Data benchmarks and applications. The proposed middleware and runtimes will be made publicly available to the community. The research enables curricular advancements via research in pedagogy for key courses in the new data analytics program at Ohio State and SDSC -- among the first of its kind nationwide.
113	-	MRI: Development of Advanced Visualization Instrumentation for the Collaborative Exploration of Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveA	CNS	Proposal #: 13-38192PI(s): Kuester, Falko; DeFanti, Thomas A.; Rosing, Tajana S.; Schulze, Jurgen P.Institution: University of California - San DiegoTitle: MRI/Dev.: Advanced Visualization Instrumentation for the Collaborative Exploration of Big DataProject Proposed:This project, developing a Scalable Omni-Presence Environment (ScOPE), a next generation visualization system for collaborative exploration of large volumes of data, provides an environment for analyzing, processing, and visualizing Big Data resulting from many different areas of science and engineering. The instrument serves as an integrative, virtual metaphor for a combined microscope and telescope, enabling users to explore data from the nano to macro to mega scale. ScOPE provides researchers the ability to study simulated and acquired data at a level of precision previously unmatched. It is expected to become the platform for training a new generation of users to be fluent in data analytics in collaborative environments. Initially, three universities will have direct access to the ScOPE instrument and all its features: U. California-San Diego (UCSD), Jackson State U. (JSU), and U of Texas Medical Branch (UTMB). Nonetheless, following the tradition of the project team (effectively done with earlier generations of visualization technologies (e.g., OptIPortal tile display walls now installed at more than 100 institutions worldwide), the critical components of the infrastructure will be broken such that they may be replicated for use at remote locations by other research or educational institutions. The developers anticipate that private-sector collaborators, such as Qualcomm and Intel, will help popularize use of specific components for the nation?s big-data analytics infrastructure. Notwithstanding, the broadest impact of the instrument should be evident in the discoveries and advances made by engineers and scientist that use ScOPE to enhance collaboration and analysis in the disciplines that have been singled out as ?Domain Drivers? for the project. These include projects led by researchers in ocean sciences (and ocean observatories); cyber-archaeology and cultural heritage diagnostics; real-time brain imaging; digital cinema and very-high quality digital media; integrative computational biology; underwater microscopy; molecular dynamics; structural biology and computational chemistry; and large-scale numerical simulation. In turn, these domain specialists will work alongside computer scientists who will address grand challenges in system architecture, data transport, security,

			wards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		representation, arching, processing multi-modal analytics, and human-computer interaction. ScOPE?s long-distance collaboration will be supported by telepresence at bandwidths ranging up to 40 Gigabits per second. Thus, the project creates a highly interactive collaboration space equipped with a natural human-computer interface and advanced 3D modeling and rendering at a sufficient scale to tackle complex experiments and analyze large amount of visual and numerical data pertaining to phenomena of wide dimensions and extreme time scales. Domain drivers have been identified to ensure that the resulting environment and tools are applicable to a broad array of scientific disciplines. These include earth system sciences, civil and structural engineering, mechanical and aerospace engineering, biomedical and electrical (and ocean observatories engineering, social sciences, and anthropology. This project takes a great leap forward into a new generation of collaborative environment that until recently was unthinkable. The display capabilities will no longer be passive; envisioned is a continuous spatial workspace imaging, including eye, skin response, and even mobile electroencephalography sensing, allowing ScOPE to respond to and infer user intent. The environment will be designed specifically to handle ?big data,? using a failure-tolerant and cloud-centric approach while also downsizing the supercomputer flash memory architecture. ?Big Data.? The instrument will enable scientific discoveries as well as research on how best to process, analyze, and visualize Scope will serve as a prototype for other similar instruments. The research enabled by ScOPE will have impacts in many areas of science.Broader Impacts: As previously mentioned, the ScOPE instrument provides researchers the ability to study simulated and acquired data at a level of precision previously unmatched. ScOPE is expected to become the platform for training a new generation of users to be fluent in data analytics in collaborative environments. The developers anticipate that private-sector collaborators, such as Qualcomm and Intel, will help popularize the use of specific components for the nation?s big-data analytics infrastructure. Notwithstanding, the broadest impact of the instrument should be evident in the discoveries and advances made by engineers and scientist that use ScOPE to enhance collaboration and analysis in the disciplines that have been singled out as ?Domain Drivers? for the project. These include projects led by researchers in ocean sciences (and ocean observatories); cyber-archaeology and cultural heritage diagnostics; real-time brain imaging; digital cinema and very-high quality digital media; integrative computational biology; underwater microscopy; molecular dynamics; structural biology and computational chemistry; and large-scale numerical simulation. In turn, these domain specialists will work alongside computer scientists who will address grand challenges in system architecture, data transport, security, representation, arching, processing multi-modal analytics, and human-computer interaction. The instrument will have direct impact on three universities, while the technology developed in building the instrument will inform the construction of similar instruments around the nation. To promote greater public appreciation of scientific research, the public will be invited to tour the visualization facilities, hopefully encouraging young people to enter career in science and engineering. The area of ScOPE accessible to the public will have significant impact on the public?s impression of academic research. ScOPE?s capabilities are likely to transform our ability to collaborate with distributed research teams and be directly applied to day-to-day research.

114	-	TUES: EAGER: Scaffolding Big Data for Authentic Learning of Computing	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DUE	The continued rapid evolution of computation and its impact on the world creates new challenges for the design of learning experiences in computation at the university level. Deeply informed skills and knowledge about computation are increasingly needed in all fields of study, including emerging fields like the "digital humanities." Growing awareness of "computational thinking" as a 21st century competency requires that learning the basics of computation becomes a part of every university student's education. However, weaving together the curriculum, pedagogy, and tools that engage learners with different dispositions and expectations about their learning of computation is a critical challenge. This exploratory project will investigate an approach to meeting this challenge by crafting authentic and engaging learning experiences using "big data" that are about real phenomenon and are realistic in scale and complexity. This approach stimulates motivation to learn and places real world problems at the core of the learning experience. The real world focus encourages additional study of computation and is believed to be especially useful to recruit and retain women in computing-intensive fields.The project will create curriculum and technologies for use in a computational thinking course and two introductory computer science courses. The work will (1) develop novel curriculum resources leveraging "big data", including interactive learning materials integrating graphical programming, visualization, and in-line execution; (2) expand and enhance a framework used to scaffold access to big data streams by adding many new streams and including interfaces to visualization and other critical services; and (3) develop, apply, and analyze an extensive set of assessment measures including assessments related to achievement of learning objectives, student motivation, and the dynamics of student cohorts. This work can serve as a national test-bed that: (1) offers a model for other universities grappling with the challenge of providing education in computation for all students, (2) provides an on-ramp for developing minor courses of study in computer science, (3) creates resources that infuse existing computer science courses with engaging big data projects, and (4) contributes to evolving discussions of theories on how students are best introduced to the computer science.
115	-	CSR: Small: Collaborative Research: Comprehensive Algorithmic Resilience	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	CNS	Big Data analytics is the process of mining useful knowledge in very large data sets, critical to the advancement of many research and application fields. With manufacturing technology downscaling coupled with increasing power densities, modern computer systems suffer from potential hardware and software failures, which can manifest themselves as errors. The errors are to happen when long-running Big Data analytics are executed on the systems, causing crash or worse, returning incorrect results silently. While numerous hardware-based resilience methods exist, they often come at the cost of excessive power efficiency reduction and substantial design complexity

		(CAR) for Big Data Analytics	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		enlargement among others. The project aims to reinforce popular Big Data analytics by embracing a host of comprehensive algorithmic resilience (CAR) software techniques that include concurrent error detection, coordinated checkpointing, and execution recovery, for high execution resilience. Upon detecting potential hardware and software errors concurrently during analytics, CAR enables execution recovery from detected errors without lofty overhead common to hardware-based resilience methods. Research activities of the project aim to achieve six main objectives that focus on addressing several technical challenges to realize CAR, based on investigators? encouraging preliminary results and prior work. The success of this project can benefit wide scientific and industrial applications due to its better support of Big Data analytics and processing. Research advances from this research are to be incorporated into undergraduate and graduate education, to be disseminated and shared broadly through technical presentations and by a website, and to inspire high school students for their STEM interest.
116	-	Student Travel Support: International Workshop on Big Data in Life Sciences, Atlanta, GA, September 9, 2015	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&	CCF	Big Data problems and challenges have become a reality in modern day computational biology and bioinformatics. The objective of this proposal is to request funds that will be used to support students and postdoctoral researchers from U.S. institutions to travel to the Association for Computing Machinery (ACM) International Workshop on Big Data in Life Sciences (BigLS), to be held on September 9, 2015 in Atlanta, GA, in conjunction with the 2015 ACM Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB). This travel award will help orient students and postdoctoral researchers in the emerging area of Big Data analytics in life sciences. It will create a unique opportunity for supported students to present their on-going work, network with experts and peers in the field, receive mentoring, while contributing to the workshop's technical program. It will help to prepare the future generation scientific workforce to meet the challenges of this critical area of life sciences.Funds will be used to support up to 10 students and postdoctoral researchers, working in areas at the intersection of Big Data and life sciences. Recipients of the support will be selected via a widely advertised competitive process involving the submission of a travel grant application, and review by the workshop's organizational committee. Preference will be

			ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		given to women and underrepresented minority groups, first-generation college students, and undergraduate researchers. The workshop will feature peer-reviewed papers and invited talks on five key research themes that underline Big Data research in life sciences: i) scalable algorithms and techniques for Big Data analytics in molecular biology, ii) statistical and integrative approaches to Big Data biology and informatics, iii) high performance computing methods and software for Big Data biology, iv) scientific workflows and platforms for Big Data in life sciences, v) software and hardware solutions for managing Big Data in biomedical informatics. This year's special emphasis theme is on scientific workflows for Big Data analytics, including workflows design, optimization, management and execution in HPC and cloud environments.
117	-	EAGER: Scalable Big Data Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleO	IIS	Big Data analytics requires bridging the gap between data-intensive computing and data-driven computing to obtain actionable insights. The former has primarily focused on optimizing data movement, reuse, organization and storage, while the latter has focused on hypothesis-driven, bottom-up data-to-discovery and the two fields have evolved somewhat independently. This exploratory project aims to investigate a holistic Ecosystem that optimizes data generation from simulations, sensors, or business processes (Transaction Step); organizes this data (possibly combining with other data) to enable reduction, pre-processing for downstream data analysis (Organization Step); performs knowledge discovery, learning and mining models from this data (Prediction Step); and leads to actions (e.g., refining models, new experiments, recommendation) (Feedback Step). Intellectual Merit: As opposed to the current practice of considering optimizations in each step in isolation, the project considers scalability and optimizations of the entire Ecosystem for big data analytics as part of the design strategy. The project aims to consider big data challenges in designing algorithms, software, analytics, and data management. This strategy contrasts with traditional approaches that first design algorithms for small data sizes and then scale them up. The project aims to treat data complexity, computational requirement, and data access patterns as a whole when designing and implementing algorithms, software and applications. Broader Impacts: The project could advance the state of the art in big data analytics across a number of key applications such as Climate Informatics and Social Media Analytics. The software resulting from the project is being made available to the broder scientific community under open source license. The project offers enhanced opportunities for education and training of graduate

			nly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		students and postdoctoral researchers at Northwestern University.
118	-	Collaborative Research: Efficient Parallel Iterative Monte Carlo Methods for Statistical Analysis of Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	DMS	The integration of computer technology into science and daily life has enabled the collection of massive volumes of data. To analyze these data, one may have to resort to parallel and distributed architectures. While the parallel and distributed architectures present new capabilities for storage and manipulation of big data, it is unclear, from the inferential point of view, how the current statistical methodology can be transported to the paradigm of big data. Also, growing data size typically comes together with a growing complexity of data structures and of the models needed to account for the structures. Although iterative Monte Carlo algorithms, such as the Markov chain Monte Carlo (MCMC), stochastic approximation, and expectation-maximization (EM) algorithms, have proven to be very powerful and typically unique computational tools for analyzing data of complex structures, they are infeasible for big data as for which a large number of iterations and a complete scan of the full dataset for each iteration are typically required. Big data have put a great challenge on the current statistical methodology. The investigators propose a general principle for developing Monte Carlo algorithms that are feasible for big data and workable on parallel and distributed architectures; that is, using Monte Carlo averages calculated in parallel from subsamples to approximate the quantities that originally need to calculate from the full dataset. This principle avoids the requirement for repeated scans of full data in algorithm iterations, while enabling the algorithm to produce statistically sensible solutions to the problem under consideration. Under this principle, a general algorithm, the so-called subsampling approximation-based parallel stochastic approximation algorithm, is proposed for parameter estimation for big data problems. Unlike the existing algorithms, such as the bag of little bootstraps, aggregated estimation equation, and split-and-conquer algorithms, the proposed algorithm works for the problems for which the observations are generally dependent. Under the same principle, a subsampling approximation-based parallel Metropolis-Hastings algorithm is proposed for Bayesian analysis of big data, and a subsampling approximation-based parallel Monte Carlo EM algorithm is proposed for parameter estimation for the big data problems with missing observations. In addition to the subsampling approximation-based parallel iterative Monte Carlo algorithms, an embarrassingly parallel MCMC algorithm is proposed for Bayesian analysis of big data based on the popular idea of divide-and-conquer. Various schemes of dataset partition and results aggregation are proposed. The

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		validity of the proposed parallel iterative Monte Carlo algorithms, including both the subsampling approximation-based and embarrassingly parallel ones, will be rigorously studied. The proposed algorithms will be applied to spatio-temporal modeling of satellite climate data, genome-wide association study, and stream data analysis.The intellectual merit of this project is to propose a general principle for statistical analysis of big data: Using Monte Carlo averages of subsamples to approximate the quantities that originally need to calculate from the full dataset. This principle provides a general strategy for transporting the current statistical methodology to the paradigm of big data. Under this principle, a few subsampling approximation-based parallel iterative Monte Carlo algorithms are proposed. The proposed algorithms address the core problem of big data analysis:how to make a statistically sensible analysis for big data while avoiding repeated scans of the full dataset? This project will have broader impacts because big data are ubiquitous throughout almost all fields of science and technology. A successful research program in theory and methods of parallel iterative Monte Carlo computations can have immense benefit widely throughout science and technology. The research results will be disseminated to the communities of interest, such as atmospheric science, biomedical science, engineering, and social science, via direct collaboration with researchers in these disciplines, conference presentations, books, and papers to be published in academic journals. The project will have also significant impacts on education through direct involvement of graduate students in the project and incorporation of results into undergraduate and graduate courses. In addition, the package Distributed Iterative Statistical Computing (DISC) that will be developed under this project is designed to provide a platform for Ph.D. students and researchers like the investigators with network-connected computers to experiment new ideas of developing efficient iterative Monte Carlo algorithms in parallel or, more exactly, grid computing environments.
119	-	Building a Big Data Analytics Workforce in iSchools	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOf	DUE	The significance and importance of this project resides in the introduction of big data analytics into the education landscape. There is increasing demand for skilled personnel in big data industries, but existing big data curricula at the university level focus primarily on students with a strong computational background, ignoring a large segment of students who might otherwise pursue education and training in this vital area, but who will be faced with big data issues in the workplace. This project aims at addressing the national demand for professionals with knowledge in big data and broadening the pool for a big data analytics workforce. Part of this effort will involve research as to whether the newly developed learning modules are more effective at increasing students' big data competencies, e.g., knowledge, skills, and analysis.The goal of this project is to develop three innovative learning modules. These modules will be designed to: (i) utilize both group-based and contextualized learning methods and (ii) be applicable and accessible to students majoring in disciplines outside, but related to main-stream computer science (e.g., iSchools). The first module will involve digital exercises where students will be asked to develop their own narratives about the relevance and significance of big data in solving real-life problems and will be expected to become knowledgeable of, and proficient with, big data concepts and applications. The second module will be more technical in nature and will allow students to discover the efficacy of big data concepts in solving practical problems in information security. Finally, the third module will introduce more advanced topics in big data mining, such as

			ficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		examining a large amount of complex data to unearth important patterns and knowledge, and introducing how to interpret the results to arrive at appropriate decisions in a specific context. Analysis of the research question surrounding the learning effectiveness will employ quasi-experimental designs that use pretests and posttests with control groups. Students will choose a course section without knowledge of which section will include the new learning modules. In analyzing the data, a hierarchical linear modeling approach will be used to evaluate the effectiveness of the intervention.
120	-	BIGDATA: F: DKM: Collaborative Research: Scalable Middleware for Managing and Processing Big Data on Next Generation HPC Systems	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=	IIS	Managing and processing large volumes of data and gaining meaningful insights is a significant challenge facing the Big Data community. Thus, it is critical that data-intensive computing middleware (such as Hadoop, HBase and Spark) to process such data are diligently designed, with high performance and scalability, in order to meet the growing demands of such Big Data applications. While Hadoop, Spark and HBase are gaining popularity for processing Big Data applications, these middleware and the associated Big Data applications are not able to take advantage of the advanced features on modern High Performance Computing (HPC) systems widely deployed all over the world, including many of of the multi-Petaflop systems in the XSEDE environment. Modern HPC systems and the associated middleware (such as MPI and Parallel File systems) have been exploiting the advances in HPC technologies (multi/many-core architectures, RDMA-enabled networking, NVRAMs and SSDs) during the last decade. However, Big Data middleware (such as Hadoop, HBase and Spark) have not embraced such technologies. These disparities are taking HPC and Big Data processing into "divergent trajectories." The proposed research, undertaken by a team of computer and application scientists from OSU and SDSC, aim to bring HPC and Big Data processing into a "convergent trajectory." The investigators will specifically address the following challenges: 1) designing novel communication and I/O runtime for Big Data processing while exploiting the features of modern multi-/many-core, networking and storage technologies; 2) redesigning Big Data middleware (such as Hadoop, HBase and Spark) to deliver performance and scalability on modern and next-generation HPC systems; and 3) demonstrating the benefits of the proposed approach for a set of driving Big Data applications on HPC system. The proposed work targets four major workloads and applications in the Big Data community (namely data analytics, query, interactive, and iterative) using the popular Big Data middleware (Hadoop, HBase and Spark). The proposed framework will be validated on a variety of Big Data benchmarks and applications. The proposed middleware and runtimes will be made publicly available to the community. The research enables curricular advancements via

			nt=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		research in pedagogy for key courses in the new data analytics program at Ohio State and SDSC -- among the first of its kind nationwide.
121	-	CAREER: An Integrated Inferential Framework for Big Data Research and Education	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDate	DMS	This project addresses several fundamental challenges in modern data analysis and aims to create a new research area named Big Data Inference. Currently available literature regarding Big Data research mainly focuses on developing new estimators for complex data. However, most of these estimators are still in lack of systematic inferential methods for uncertainty assessment. This project hopes to bridge this gap by developing new inferential theory for modern estimators unique to Big Data analysis. The deliverables of this project include easy-to-use software packages, which directly help scientists to explore and analyze complex datasets. The principal investigator is also actively collaborating with many scientists to ensure the more direct impact of this project to the targeted scientific communities.This project aims to develop novel inferential methods for assessing uncertainty (e.g., constructing confidence intervals or testing hypotheses) of modern statistical procedures unique to Big Data analysis. In particular, it develops innovative statistical inferential tools for a variety of machine learning methods which have not yet been equipped with inferential power. It also provides necessary inferential tools for the next generation of scientists to be competitive in modern data analysis.

			Operator=&StartDateOperator=&ExpDateOperator=		
122	-	AF: SMALL: Learning to Parsimoniously Model and Compute with Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CCF	This project develops mathematical and computational approaches for big data exploitation. Fast and onlinealgorithms that learn and adapt as data arrives and changes are developed. How to automaticallyunderstand and reduce redundancy in the data, for a given task, is also addressed in this project. Big datacomes in multiple forms, e.g., audio and video, audio and text, video and weather, video from multiplesources, brain imaging from multiple modalities, friendship networks and individual preferences. This isalso addressed in this project. The broad impact of the research is born in the large and diverseapplicability of big data and in the techniques here developed. In the education arena, the developedInternet classes have an audience of tens of thousands, and the project provides unique integration ofresearch and undergraduate education via different Duke initiatives.The framework follows the parsimony theory of sparse modeling. Challenges are addressed with a gamechangingparadigm: learning to optimize; on-line learning what the task-dependent optimizer is expectedto do, developing computationally efficient algorithms to approximate the ideal behavior of sometimesunknown optimizers. The work derives novel multi-modal formulations for network inference, and realtimeon-line robust PCA and robust NMF, fundamental tools in big data modeling and exploitation; aswell as robust 3D shape, networks, and multi-modal matching. The formulation elegantly solves bileveloptimization problems rendering it efficient for classification and signal separation tasks. Sparsemodeling is extended to new venues and algorithms, making such techniques usable for big data. Theformulations and theoretical foundations are complemented with numerous applications.

			DateOperator=&ExpDateOperator=		
123	-	BD Hubs: NORTHEAST: The Northeast Big Data Innovation Hub	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	The Northeast Big Data Innovation Hub will speed innovation and research by creating an ecosystem for a diverse group of researchers and practitioners to learn from each other. The Hub will encourage collaboration through workshops, visiting research positions, student ambassadors, virtual meetings, research projects and an online web presence, and facilitate sharing of data, tools, infrastructure, techniques, and insights to address challenges using big data. It will also facilitate the development, collection and organization of educational materials and learning opportunities for teachers in pre-K through high school. To communicate the importance of big data in everyday life, the Hub will work with museum educators and curators to develop materials and exhibits on this topic and share work already being done across the region. Finally, the Hub will develop programs for government and the general public on how to take advantage of data analytics. The scale of this project will allow the hub to make substantial intellectual gains. By identifying shared challenges and facilitating the deployment of human and financial resources, solutions will be developed that can be applied to a range of research questions, government and economic problems and that can be useful to society generally. The Northeast Hub will be organized in a "hub and spoke model" with common theme activities (connectors) that are centrally organized and spokes that reflect regional interests aligned with national priorities. Initial areas of focus, or spokes, will be: Health, Energy, Finance, Cities/Regions, Discovery Science and Data Science in Education. Cross-cutting connectors will be: Data Sharing, Privacy & Security, Ethics & Policy and Education. Topics of importance to the northeast will be added to the Hub during the first three years.For further information see the project web site: http://northeastbdhub.dsi.columbia.edu/

			DateOperator=&ExpDateOperator=		
124	-	BIGDATA: Collaborative Research: F: Big Data, It's Not So Big: Exploiting Low-Dimensional Geometry for Learning and Inference	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	This research will leverage ideas from algebraic and differential geometry to address core problems in modern high-dimensional and massive data science. The project will develop statistical methods and numerical tools, grounded in solid mathematical, statistical, and computational foundations, to extract low dimensional geometry from massive data with applications in clustering, data summarization, prediction, dimension reduction, and visualization. The solutions developed as part of this project can result in fundamental advances in practical applications across fields as diverse as biology, medicine, social sciences, communication networks, and engineering. In addition to internal validation via statistical and mathematical theory and simulation studies, the methods developed in the project will involve external validation via interdisciplinary applications. These applications include: (1) inference of population structure from genomic data; (2) document analysis via topic models; and (3) inference of subsets of putative gene networks relevant to drug resistance in melanoma.The research is motivated by the central premise that, even though the amount of data may be massive, a compact model can represent these data. Specifically, high-dimensional and/or massive data can be reasonably approximated by a mixture of subspaces, for which sparse representations exist. A mixture of subspaces of potentially different dimensions is a flexible, rich representation of data with nice mathematical properties that can scale to large data. There are several fundamental challenges in modeling mixtures of subspaces that will be addressed in this research: 1) the subspaces will be of different dimensions, 2) both the subspace parameters and the mixing parameters need to be inferred, 3) efficient algorithms for inference are required for both high-dimensional and massive data. The central foundational impediment in all of these challenges is that the model is a stratified space (a union of manifolds), and therefore has singularities. The key insight in this research is that there exist embeddings and representations of the model space that mitigate these singularities. These ideas are implemented as concrete Bayesian, frequentist, and numerical algorithms and models to address the real world examples listed above.

			DateOperator=&ExpDateOperator=		
125	-	AF: Small: Collaborative Research: Algorithmic and Computational Frontiers of MapReduce for Big Data Analysis	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CCF	Modern science and engineering heavily relies on processing massive data sets and the size of the data requires applications to run using distributed computing frameworks. However, many existing methods essential to the applications are not easily adapted to work in distributed settings. This project aims to develop new efficient ways of processing large data sets in widely used distributed computing platforms. The project will reveal new methods for processing diverse and complex data sets of massive size and allow for various applications scale to large inputs. The work has the potential to fundamentally change algorithmic techniques used in distributed computing, helping to shape big data research, the computing industry, and the growing economy reliant on big data analysis. Research outcomes will be integrated with education by writing an extensive survey/tutorial on the core algorithmic ideas used in the new discoveries to make the ideas transparent to the algorithmic developers and practitioners. The PIs will make some of the discovered algorithmic ideas accessible even to undergraduate students, helping them get prepared to cope with algorithmic challenges in distributed computing for large data sets. Special efforts will be made to include women and minorities in advising and mentoring plans.The main goal of the project is to find new ways of unlocking the underlying power of MapReduce, a popular distributed platform, through the development of new algorithmics. The developed algorithms should have provably strong guarantees and demonstrate the effectiveness via empirical experiments. Considering the increasing demand for large data analysis, establishing a solid theoretical MapReduce model and developing new algorithmic ideas will have the potential to establish faster and memory efficient algorithms for distributed computing. The PIs will consider a collection of carefully chosen problems to understand in the MapReduce setting that not only have strong connections to theoretical work but also have the potential for high impact in real world Big Data applications: Clustering, Distributed Dynamic Programming, and Limitations of MapReduce. This will be done in parallel with the attempt to better understand the currently accepted MapReduce models that have been developed and to perhaps further refine them to better connect models with practice.

			DateOperator=&ExpDateOperator=		
126	-	AF: Small: Collaborative Research: Algorithmic and Computational Frontiers of MapReduce for Big Data Analysis	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CCF	Modern science and engineering heavily relies on processing massive data sets and the size of the data requires applications to run using distributed computing frameworks. However, many existing methods essential to the applications are not easily adapted to work in distributed settings. This project aims to develop new efficient ways of processing large data sets in widely used distributed computing platforms. The project will reveal new methods for processing diverse and complex data sets of massive size and allow for various applications scale to large inputs. The work has the potential to fundamentally change algorithmic techniques used in distributed computing, helping to shape big data research, the computing industry, and the growing economy reliant on big data analysis. Research outcomes will be integrated with education by writing an extensive survey/tutorial on the core algorithmic ideas used in the new discoveries to make the ideas transparent to the algorithmic developers and practitioners. The PIs will make some of the discovered algorithmic ideas accessible even to undergraduate students, helping them get prepared to cope with algorithmic challenges in distributed computing for large data sets. Special efforts will be made to include women and minorities in advising and mentoring plans.The main goal of the project is to find new ways of unlocking the underlying power of MapReduce, a popular distributed platform, through the development of new algorithmics. The developed algorithms should have provably strong guarantees and demonstrate the effectiveness via empirical experiments. Considering the increasing demand for large data analysis, establishing a solid theoretical MapReduce model and developing new algorithmic ideas will have the potential to establish faster and memory efficient algorithms for distributed computing. The PIs will consider a collection of carefully chosen problems to understand in the MapReduce setting that not only have strong connections to theoretical work but also have the potential for high impact in real world Big Data applications: Clustering, Distributed Dynamic Programming, and Limitations of MapReduce. This will be done in parallel with the attempt to better understand the currently accepted MapReduce models that have been developed and to perhaps further refine them to better connect models with practice.

			DateOperator=&ExpDateOperator=		
127	-	MRI: Acquisition of an Instrument for Research in Irregularly Parallel Big Data Computation	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CNS	Proposal #: 13-37884PI(s): Cook, Jonathan E. Cao, Huiping; Cook, Jeanine M.; Pontelli, Enrico; Song, MingzhouInstitution: New Mexico State University Title: MRI/Acq.: Instrument for Research in Irregularly Parallel Big Data ComputationProject Proposed:This project, acquiring a computational instrument configured to support data-driven graph computations (DDGC) that might enable-to-scale and improve parallel computation that is generally irregular and hard to scale and improve. (Very regular computations--for example, fluid dynamics--already have a long history of development and are highly optimized to run on modern high performance (HPC) instruments.) The instrument, with 320 cores and a node and system architecture, is designed specifically with potentially transformative DDGC research in mind. The local node memory hierarchy includes both fast solid state secondary storage and traditional mechanical disk storage. Research projects explore ways to exploit the new layer of fast solid state storage, not just as a standalone data container, but within the memory hierarchy of the local node. The system has in its configuration both GPU and FPGA computing support. Some of the research projects exploit non-traditional, yet potentially powerful computation mechanisms. In particular, the instrument supports the following projects:- Genome Assembly and Annotation Computations,- Data Mining over Large Graphs,- Reasoning with Big Knowledge,- Hardware Acceleration for Scalable Graph-Based Computations, and- Data Driven Monitoring and Analysis of Scientific Computations.Each of these projects will make use of the general architecture of the instrument. GPU and FPGA capabilities will be used to further explore and enhance the possibilities to improve performance of data-driven graph computations. The instrument architecture is intended to enable cross-fertilization between the research projects that will, in turn, contribute to the development of new approaches that can perform DDG computations efficiently.Broader Impacts: Due to the growth of data analytics in so many areas of society, the techniques developed utilizing the instrument should be valuable across society (from economic capacity to homeland security needs). Techniques that can improve the DDG computations within many of the big data computations (from improved algorithms to newly demonstrated hardware approaches) are immediately useful. Moreover, the instrumentation enriches the research and educational activities at a minority-serving institution within an EPSCoR jurisdiction. The project offers opportunities for students and professors to participate in novel computer research.

			DateOperator=&ExpDateOperator=		
128	-	EAGER: Science in the Time of Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	DEB	The scientific community is awash in 'big data' but few practicing ecologists use these data to answer important ecological questions. They rely instead on the traditional approach of collecting new, experimental data focused on particular species, habitats, or problems. In addition, the data-intensive computational methods commonly needed to analyze big datasets are not easily accessible to most researchers. This high-risk, high-reward project could dramatically alter both the ways in which ecologists address questions and the types of questions that they tackle. It therefore represents a major contribution to NSF's efforts to extend ecological research in new directions to provide answers to more complex questions. A knowledge-driven, open access system that 'learns' and becomes more efficient and easier to use as data streams increase in variety and size is needed for timely scientific progress in an era of big data. This approach is centered on establishing linkages between databases and hypothesis-based inquiry that result in the derivation of new or refined hypotheses as a result of improved access to dynamic databases. The investigators recently implemented a hypothesis-driven, process-based analytical methodology that was conceptually integrated with a data-intensive machine learning approach. This integrated approach allowed them to use multiple long-term datasets to narrow a diverse suite of mechanistic explanations to a single, most likely process. This process was then tested by a short-term experiment that saved time and money and yielded a more definitive answer than the more traditional approach described above. To further this approach, this project will test, refine, and automate this new integrative effort to develop a prototype cyber-infrastructure capable of significantly advancing the environmental sciences. Open access data, programming scripts, and derived data products will reduce the time lag for knowledge transfer from an individual to the research community, likely increase the speed of scientific progress, and provide a filter and memory for how to deal with large amounts of data of mixed quality. A postdoctoral researcher will work collaboratively with computer scientists, ecologists, and eco-informatics experts from three universities (New Mexico State University, University of Texas El Paso, and Evergreen College) and one corporation (Microsoft) to develop, test, and automate this knowledge-learning analytics system. Two workshops will be organized to test the ability of the system to learn while using diverse datasets and to introduce the approach to a wide variety of users.

			DateOperator=&ExpDateOperator=		
129	-	REU Site: Computational Methods for Discovery Driven by Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	As participants in the University of Minnesota (UMN) REU Site program, students will engage in research that develops computational methods for scientific discovery across disciplines that are driven by big data. In this 10-week summer program, in addition to immersion in research, students will receive technical training and professional development that encourages and prepares them for a sustained career in the sciences. This includes Big Data Colloquia, Communicating Science workshops, career mentoring, and public dissemination of research findings. Towards an objective of increased participation and broader impacts, this program will bring together nationally recruited students and those from UMN and local institutions to establish a cohort with diverse academic and cultural backgrounds. A Big 10 University situated in a large urban environment, UMN has a strong research community that encourages transdisciplinary research within and outside the boundaries of the institution.Closely mentored by a member of the Computer Science and Engineering (CS&E) faculty, each student will contribute to active research that addresses open questions in computational complexity, machine learning, parallel and distributed computing, mobile and cloud computing, or graphics and visualization. A UMN REU participant might use observation data to simulate crowd behavior, analyze genomic sequence data to better understand microbial communities, develop tools to analyze chemical-genetic interaction networks, improve spatial perception in a virtual environment, develop visualization techniques to better understand massive data sets, enhance parallel distributed processing through algorithm development or by harnessing the computational power of a network of mobile devices, or use graph-based approaches to better understand climate change. The diverse research of CS&E faculty represents collaboration across the University with faculty in genetics, chemistry, climate science, neuroscience, architecture, medicine, and biomedical engineering to propel all of these disciplines and computer science towards previously unattainable insights and discoveries. The web site http://www-users.cs.umn.edu/~boley/big-data-reu/index.html has more information on this project.

			DateOperator=&ExpDateOperator=		
130	-	BIGDATA: F: DKM: Collaborative Research: PXFS: ParalleX Based Transformative I/O System for Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Recent decades have seen the development of computational science where modeling and data analysis are critical to exploration, discovery, and refinement of new innovations in science and engineering. More recently the techniques have been applied to arts, social, political and other fields less traditionally reliant on high performance computing. This innovation has grown out of realization some 20 years ago that I/O (input/output) support for high performance parallel and distributed architectures had lagged behind that of pure computational speed, and further that bring I/O up to speed was both critical, and a rather difficult problem. The core hurdle of contemporary I/O on large HPC machines relates to issues of latency in large parts caused by the deficiencies of the historical I/O model that was relevant when computers were exclusively large, centralized, single processor systems shared by many time-sharing programs. In order to improve I/O on scalability on future hardware architectures novel approaches are required.This project is conducting research on an extension of ParalleX, a new highly innovative parallel execution model. The extension provides a powerful I/O interface that allows researchers to create highly efficient data management, discovery, and analysis codes for Big Data applications. This new extension, known as PXFS, is based on HPX, an implementation of ParalleX based on C++, and OrangeFS, a high performance parallel file system. The research goal driving PXFS is to extend HPX objects into I/O space so that the objects become persistent and storage becomes another class of memory, all accessed as a single virtual address space and managed by an event driven dynamic adaptive computation environment. Critical aspects of this approach include futures-based synchronization, dynamic locality management, dynamic resource management, hierarchical name space, and an active global address space (AGAS). The overall goals of PXFS are to eliminate the division of programming imposed by conventional file system through the unification of name spaces and their management, and to minimize global synchronization in order to support asynchronous concurrency. The research methodology is to implement a Map/Reduce application framework using PXFS and evaluate its effectiveness in both performance and ease of use.This project is conducted at three major research universities involving undergraduate and graduate students, post-docs, and high-school teachers and their students. The project includes a PI from the functional genomics field acting as domain science expert in order to focus the development efforts on real world problems. Graduate students and post-docs involved in the project are trained in these areas to promote scientists who understanding both aspects of Big Data problems. The project engages under represented minorities with the

			DateOperator=&ExpDateOperator=		goal to inspire them to pursue a career in computer science or genomics. The software developed by the project is available open-source and archived using an integrated source code revision repository, wiki, and bug tracking software system in addition to code releases with accompanying documentation.
131	-	TWC: Medium: Privacy Preserving Computation in Big Data Clouds	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CNS	Privacy is critical to freedom of creativity and innovation. Assured privacy protection offers unprecedented opportunities for industry innovation, science and engineering discovery, as well as new life enhancing experiences and opportunities. The ability to perform efficient and yet privacy preserving big data computations in the Cloud holds great potential for safe and effective data analytics, such as enabling health-care applications to provide personalized medical treatments using an individual's DNA sequence, or enabling advertisers to create targeted advertisements by mining a user's clickstream and social activities, without violation of data privacy. The PrivacyGuard project is developing algorithms, systems and tools that provide end-to-end privacy guarantees over the life cycle of a data analytic job. The end-to-end privacy guarantee can be measured by how difficult one can learn about some of the original sensitive data from the sanitized data releases, the intermediate results of execution and the output of an analytic job. The ultimate goal of PrivacyGuard is to develop a methodical framework and a suite of techniques for ensuring distributed computations to meet the desired privacy requirements of input data, as well as protecting against disclosure of sensitive patterns during execution and in the final output of the computation.The PrivacyGuard project advances the knowledge and understanding of privacy preserving distributed computation from three perspectives: (1) It designs formal mechanisms to formulate a data owner's end-to-end privacy requirement for each data release, for example, by associating each data release with a well-defined usage scope to confine the set of data analytics models and algorithms that can operate on the released data. (2) It develops a suite of execution privacy guards with dual objectives: to audit and enforce privacy compliances during distributed computation against data-flow based privacy violations and to guard the compliance of input privacy. (3) It devises a proactive approach to output privacy against information leakages associated with mining output, for example, by leveraging differential privacy model to maximize the upper bound for data privacy guarantee and minimize the lower bound for data utility losses. The PrivacyGuard project is the first effort towards a practical and systematic implementation framework for ensuring the end-to-end privacy in distributed big data computations. Furthermore, by integrating the PrivacyGuard research with the curriculum development on big data systems and analytics courses at Georgia Institute of Technology, it contributes to the education and training of new generation of data scientists to be the privacy compliance advocates.

			DateOperator=&ExpDateOperator=		
132	-	Whither Turbulence and Big Data in the 21st Century? (Corsica, April 20-24, 2015)	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CBET	The proposal seeks funds to partially cover travel expenses of underrepresented young faculty and students as well as dissemination expenses related to the conference entitled, "NSF/ONR: Whither Turbulence and Big Data In the 21st Century" to be held in Corsica, France, April 20-24, 2015. World class researchers will participate in his meeting.Research activities in turbulence involve the collection of massive amounts of data from multi-scale computer simulations and/or large-scale experiments. For example, highly non-linear time dependent stochastic turbulent flow experiments routinely produce terabytes of multi-modal data (velocity, pressure, temperature, acoustics, etc.) in fractions of a second. Such computational and experimental research presents formidable big data challenges and overcoming them is indeed critical for achieving real breakthroughs. Better understanding of turbulence will lead to improved understanding of a number of engineering problems in aerospace and naval applications, biomedical applications, wind energy production, energy conservation, etc. The symposium will include breakout sessions each day to maximize interaction among participants. The organizing committee is committed to engage senior minority PhDs and junior faculty. The main goal of this workshop is to frame the challenges in big data and turbulence research fluid dynamics for the next several. Other objectives of the symposium are as follows: (i) The symposium seeks to bring awareness and basic knowledge on the state-of-the-art fluid dynamics research including: modeling, experiments, and simulations; (ii) The symposium will provide a unique cross-disciplinary environment for identifying the current issues and challenges of turbulence for the next quarter century; (iii) It will foster collaborations and students exchange with USA, Canada, Europe and Australia; (iv) It will enable the development of strategies and new initiatives for data reduction and visualization; including Cyber Infrastructure.

			DateOperator=&ExpDateOperator=		
133	-	Forecasting Innovation Pathways of Big Data & Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	SMA	Research and development (R&D) leads to the creation of new science-based technology innovations. There is a need to improve understanding of how research discoveries translate into new technologies and then develop into useful innovation. Methods used to address this concern either draw on trends in large amounts of historic information or draw on expert judgment. This project will apply advanced research techniques to enhance methods to identify trends and patterns, and to help forecast innovation pathways. Such knowledge is vital to promote scientific progress by investing judiciously in high promise R&D. It also can aid in technology management to determine how best to advance a specific field of science.This project will provide a case study to improve five analytical processes, identified as vital to improve the methodology of forecasting innovation pathways. The case to be analyzed is ?big data & analytics? ? a topic of great national importance. Figuring out how to gain advantage from large data sets will impact national scientific progress, industrial productivity, and defense. At the same time, big data poses issues of privacy and security, among others. The topic of big data was selected because it is under study by the U.S. Government Accountability Office (GAO). We anticipate sharing information on methods and findings with GAO to gain insight into ways to make our methodology more useful. . This proposal extends a previously funded SciSIP project (1064146) to address a time-limited opportunity to address the GAO case.

			DateOperator=&ExpDateOperator=		
134	-	TWC: Small: Practical Assured Big Data Analysis in the Cloud	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CNS	The use of "cloud technologies" presents a promising avenue for the requirements of big data analysis. Security concerns however represent a major impediment to the further adoption of clouds: through the sharing of cloud resources, an attack succeeding on one node can tamper with many applications sharing that node. This project explores the combination of two readily-available, practical mechanisms to holistically achieve assured cloud-based big data processing: (1) Byzantine fault tolerant replication and (2) partially homomorphic encryption. The former consists in replicating computational entities to achieve availability, and comparing their produced results to enforce integrity of results as well as isolation of suspicious components. The latter suggests leveraging the innate ability of existing "cryptosystems" to support certain specific operations on data in encrypted state in order to ensure its privacy. The project envisions an efficient application of redundant computation (replication) and redundant storage (different encryptions of same data) through a smart breakdown of programs into sub-computations and sub-datasets based on boundaries identified via program analysis. To enable that vision, the scope of Byzantine fault tolerant replication is extended beyond the present client-server scenarios to avoid significant slowdowns when applied to fine-grained parallelization of large datasets; similarly, partially homomorphic encryption is made applicable without hampering parallelism and beyond very simple programs. This project will have a high impact on software developers given the continuously increasing relevance of the cloud computing paradigm and of big data. Results will be made broadly available through scientific publications and use open-source software systems for implementation.

			DateOperator=&ExpDateOperator=		
135	-	INSPIRE Track 1: SPINTOP: Spin Torque Nano Oscillator Arrays for "Big Data" Applications	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	ECCS	Overview:This INSPIRE award is partially funded by the Electronics, Photonics, and Magnetics Devices Program in the Division of Electrical, Communications and Cyber Systems in the Directorate for Engineering; the Software and Hardware Foundations program in the Division of Computing and Communication Foundation in the Directorate for Computer & Information Science and Engineering; and the Electronic and Photonic Materials program in the Division of Materials Research in the Directorate for Mathematical and Physical Sciences.SPINTOP will explore, in detail, a new paradigm of automata computing that is based on a non-Boolean, non-Von Neumann architecture and is enabled by the electrical coupling of Spin Torque Nano-oscillators (STNOs). The project's goal is to build a simple, yet clear demonstration, of how this can achieve significant speedups in pattern recognition to match detailed simulations. This project addresses the 'Big Data' challenge with a non-conventional solution that can compute on the fly. The interdisciplinary nature of the INSPIRE program will make it possible to combine magnetic materials research with device design and test and computer architecture development involving three NSF Programs and faculty from Physics, Materials Science and Electronic and Computer Engineering at University of Virginia (UVa).Intellectual Merit:SPINTOP will be pushing the frontiers of spintronic devices, circuits and computing architectures. Key to the success of this project will be the demonstration of a novel hybrid STNO that was invented at UVa. Magnetic materials with low magnetization, low damping, high spin polarization, controlled anisotropy are key to the success of the proposed hybrid device. The proposed novel STNO combines the best features of existing devices to provide both pure sine wave output and a large voltage signal. Electrical coupling of arrays of these STNOs will be demonstrated in this project. According to detailed simulations, phase lock will occur when their individual frequencies are close - this phase locking will take place across the array rather than only nearest neighbor locking as has been previously demonstrated using spin wave coupling. This coherent phase locking is the key to the design for an associative memory for pattern recognition. SPINTOP will also develop the modeling tool used to simulate the behavior of the innovative design for STNO arrays. Further development of this tool is key for designing and implementing circuits as it will be integrated with existing codes for performing the modeling, simulation and design of hybrid CMOS-spintronic circuits.Broader Impacts:SPINTOP, if successful, will have a big impact on the scientific community by creating a new paradigm for computing, especially pattern recognition. This will help solve many of the Big Data applications that are a grand challenge today. However its impact will be much

			DateOperator=&ExpDateOperator=		broader; this project will address many of the key issues facing society today. This project is committed to fostering diversity in the workforce and will support at least one student from an underrepresented group for the duration of the project. As the INSPIRE project will fall under the umbrella of the UVA nanoSTAR Institute and the PI is also the director of nanoSTAR, this project will be closely integrated into the outreach programs and benefit from the resources of nanoSTAR. The SPINTOP faculty and students will participate in outreach visits to Virginia high schools to highlight this new technology. SPINTOP faculty and students will also participate in Nanodays, a highly interactive event held each Spring that brings over 650 local students from K-12 and their parents to UVA to experience some of the wonders of nanoscience and technology. These visitors will see a poster highlighting SPINTOP and will be invited to tour the SPINTOP laboratories. SPINTOP along with NanoSTAR will co-sponsor a club for undergraduates called NeXT (for Nano and Emerging Technologies) that hosts frequent seminars with guest speakers and will be a way for students in many of the nano-related disciplines to connect with one another. SPINTOP will leverage programs that support undergraduate research to recruit undergraduates to participate in summer research in the SPINTOP labs. In collaboration with UVA's Center for Diversity in Engineering, opportunities will be provided to several minority senior high school students to experience research in the SPINTOP facilities. SPINTOP will also have an impact on courses at UVA; there will be at least one new class developed under the auspices of this program. It will be a class on automata based computing which is the computing paradigm that the SPINTOP project will use for mapping applications at the system level. Finally, an existing class on Spintronics taught by the PI will be updated to include several lectures on spin torque oscillators and their application to computing.
136	-	MRI: Development of a Novel Computing Instrument for Big Data in Genomics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%	CNS	Proposal #: 13-37732PI(s): Lumetta, Steven S. Iyer, Ravishankar; Jongeneel, Cornelis Victor; Robinson, Gene E.; Sinha, SaurabhInstitution: University of Illinois - Urbana-ChampaignTitle: MRI/Dev.: Novel Computing Instrument for Big Data Project Proposed:This project, developing CompGen, an instrument that adopts a hardware-software co-design approach, aims to provide a- Vehicle for biologists and computer scientists to collaborate and develop new algorithms that are significantly faster and more accurate at a scale essential for handling the data deluge; - Software framework and tool set for algorithm development that support diverse data analysis and visualization; - Framework for developing accelerators and mapping to heterogeneous computational resources and hierarchical database storage. Promising technologies include emerging die-stacked and non-volatile memory technologies as well as accelerators (GPUs, FPGAs, APUs). The project brings together a multidisciplinary team of geneticists, bioinformatics specialists, computer and algorithms designers, and data mining experts. The research to be enabled includes a wide and eclectic variety of problems with direct impact on health and social issues. Some directions include understanding the impact of climate change on gene expression and ecosystems, bringing genetic analysis into medical clinics, identifying effective antibiotics, and exploring socio-genomics relations between stress, depression, and genetics among low-income African-American mothers. CompGen provides an environment that enables managing and processing genomic information and developing new algorithms. The instrument brings disruptive computing architectures and algorithmic techniques to facilitate analysis of genomic data while providing high accuracy results, resilience to errors, and scalability with growing volumes of data. It enables addressing the

			22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		challenges of scale and diversity in genomic data through the development of new algorithms, models, and statistical methods. The instrument development focuses on reduction of data volume, optimization of storage hierarchy, identification and implementation of computational primitives, data visualization, mathematical toolkit optimization, and performance and reliability assessment. These developments are expected to lead to new computational structures and hardware/software architectures that can be incorporated into hierarchical databases as well as heterogeneous processors for data analysis, compression, and optimization. Broader Impacts: In addition to serving many areas, CompGen will serve as a tool for educating students and professionals in efficient ways to process and analyze genomic data and for handling big data in general. The instrument will serve multidisciplinary classes in which students gain hands-on research experience and introductory classes that expose students to applications and tools. Existing outreach and education programs will be utilized to expose the instrument. Plans include Open House events attracting thousands of visitors, Coursera courses, and minority outreach workshops. A mentoring tool, Mytri, will be used for networking among female students. Moreover, the CompGen design will be made available to others by fundamentally changing the methods by which big datasets are handled in genomics research. To this effect, an R&D consortium of hospitals, companies, and universities has been established to help identify needs, provide sources of data, act as early adopters, and ensure that new technologies are transferred smoothly into widespread use.
137	-	Postdoctoral Fellowship: The Impact of Big Data on the Science of Metabolism	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&Award	SES	Under the direction of Hannah Landecker, post-doctoral scholar Nadine Levine, will carry out a two year laboratory ethnography focusing on the impact of big data on the science of metabolism. This study of metabolomics, the post-genomic study of the molecules and processes that make up metabolism, will serve as a case study for the role and impact of big data in biomedical research.The findings of this project will contribute to science policy discussions on the key opportunities and challenges facing the increasingly prominent field of metabolomics, as well as other fields of data-intensive research. Moreover, the findings of this project will provide collaborative and educational resources, which will enable students, researchers, and members of the public to understand how,and with what consequences for the diagnosis and treatment of disease, new understandings of metabolism are being generated with big data.

			NumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
138	-	QuBBD: Collaborative Research: Advancing mHealth using Big Data Analytics: Statistical and Dynamical Systems Modeling of Real-Time Adaptive m-Intervention for Pain	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDate	DMS	With the growing popularity of mobile phone technology, new opportunities have arisen for real-time adaptive medical intervention. The simultaneous growth of multiple "big data" sources (e.g., mobile health data, electronic health records, lab test results, genomic data) allows for the development of personalized recommendations. This award supports initiation of a collaborative research project that will generate a new mathematical model for changes in subjective pain over time in patients with chronic conditions. The model will be combined with statistical techniques to ultimately obtain optimized, continuously-updated treatment plans balancing competing demands of pain reduction and medication minimization. Those resulting personalized treatment plans will be incorporated into a currently active pilot study on mobile intervention in patients living with chronic pain due to sickle cell disease (SCD). Since nearly a quarter of patient visits to the emergency room are for conditions that could have been managed as outpatients, it is crucial to improve mobile health technologies to allow these patients to quickly recognize and receive appropriate health care information. There currently is no standard algorithm or analytical method for real-time adaptive treatment recommendations for chronic conditions like pain. Furthermore, current state-of-the-art methods have difficulty in handling continuous-time decision optimization using big data. The proposed model will consist of a dynamical systems approach using differential equations to forecast future pain levels, as well as a statistical approach tying system parameters to patient data (including reported pain levels, medication history, personal characteristics and other health records). A third key component will be the development and pilot study of a new control and optimization strategy to balance the competing demands of pain reduction and drug dosage minimization. This award is supported by the National Institutes of Health Big Data to Knowledge (BD2K) Initiative in partnership with the National Science Foundation Division of Mathematical Sciences.

			Operator=&StartDateOperator=&ExpDateOperator=		
139	-	BIGDATA: Collaborative Research: F: Big Data, It's Not So Big: Exploiting Low-Dimensional Geometry for Learning and Inference	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstr	IIS	This research will leverage ideas from algebraic and differential geometry to address core problems in modern high-dimensional and massive data science. The project will develop statistical methods and numerical tools, grounded in solid mathematical, statistical, and computational foundations, to extract low dimensional geometry from massive data with applications in clustering, data summarization, prediction, dimension reduction, and visualization. The solutions developed as part of this project can result in fundamental advances in practical applications across fields as diverse as biology, medicine, social sciences, communication networks, and engineering. In addition to internal validation via statistical and mathematical theory and simulation studies, the methods developed in the project will involve external validation via interdisciplinary applications. These applications include: (1) inference of population structure from genomic data; (2) document analysis via topic models; and (3) inference of subsets of putative gene networks relevant to drug resistance in melanoma.The research is motivated by the central premise that, even though the amount of data may be massive, a compact model can represent these data. Specifically, high-dimensional and/or massive data can be reasonably approximated by a mixture of subspaces, for which sparse representations exist. A mixture of subspaces of potentially different dimensions is a flexible, rich representation of data with nice mathematical properties that can scale to large data. There are several fundamental challenges in modeling mixtures of subspaces that will be addressed in this research: 1) the subspaces will be of different dimensions, 2) both the subspace parameters and the mixing parameters need to be inferred, 3) efficient algorithms for inference are required for both high-dimensional and massive data. The central foundational impediment in all of these challenges is that the model is a stratified space (a union of manifolds), and therefore has singularities. The key insight in this research is that there exist embeddings and representations of the model space that mitigate these singularities. These ideas are implemented as concrete Bayesian, frequentist, and numerical algorithms and models to address the real world

			ument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		examples listed above.
140	-	MCTP: Sophomore Transitions: Bridges into a Statistics Major and Big Data Research Experiences via Learning Communities	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	DMS	This Statistics Living-Learning Community (STAT-LLC) project at Purdue University addresses two transitions in the training of undergraduate statistics students: (1.) The bridge from first-year general curriculum into sophomore year Statistics major courses, and (2.) The bridge into a student's first research experience in data analysis, especially with Big Data. Each year over a five year period, 20 sophomore trainees will take part in the STAT-LLC that involve a new project-based course in big data analysis, new mentored research collaborations and thrusts, a new seminar series, professional development initiatives, and living together in a new learning community. The project leaders will schedule 3 core courses (probability theory, statistical theory, and data analysis) and the seminar series in blocks, so students take these courses together as a cohort (and still have plenty of time for individual electives), live together in a dormitory, and participate in a research experience that lasts 12 months (start of sophomore year to start of junior year). The Senior Personnel are 12 faculty from the Department of Statistics at Purdue University. Additional faculty mentors will hail from Actuarial Science, Atmospheric/Earth Science, Biology, Biomedical Engineering, Chemical Engineering, Electrical and Computer Engineering, Health and Human Sciences, and the Regenstrief Center for Healthcare Engineering and be supported by administrators, academic advisors, and recruiters. Beyond Purdue, Rose-Hulman will host the students for their annual Undergraduate Mathematics Conference. The Department of Statistical Science at Duke University and the Department of Statistics at the University of California, Los Angeles will partner with Purdue on a DataFest student competition. Assessment will be provided by Purdue's Discovery Learning Research Center, and an External Advisory Board has been assembled with colleagues from the American Statistical Association, Duke University, Indiana University, the John N. Gardner Institute for Excellence in Undergraduate Education, the Statistical and Applied Mathematical Sciences Institute, University of Minnesota, University of Wisconsin, University of South Carolina, and the Walter and Eliza Hall Institute of Medical Research, in Melbourne, Australia.Pairing the academic and residential life experiences is a known best practice for the first-year experience but is rarely used at the sophomore level as it is in the present project. Furthermore, the Statistics Living-Learning Community (STAT-LLC) will pair this model with an early, team-oriented introduction to big data analysis. We conjecture that the project will promote retention, instill confidence in the students, help propel the students through the formative second year of college, and prepare them to apply for graduate school. This holistic experience should result in more

			DateOperator=&ExpDateOperator=		students pursuing double majors: Statistics and an applied major in a data-oriented discipline. Our assessment of this comprehensive training experience should reveal new insights into the way that pairing academic coursework, research, professional development, and residential life at the sophomore level can alleviate the sophomore slump within STEM disciplines. The Senior Personnel will offer faculty training workshops at national meetings to teach colleagues how to unite the aspects of the sophomore experience in a relevant way for their campuses. The workshops will be free, to maximize participation by faculty who would not otherwise have access to such resources, networking, partnerships, and professional development opportunities. As a result, more campuses will be able to offer research experiences for sophomores in the context of a learning community. Students will communicate their experiences each spring at the Rose-Hulman Undergraduate Math Conference and the Indiana MAA Section meeting. Diversity initiatives will be supported by Purdue's Disability Resource Center; Emily Mauzy Vogel Sophomore Leadership Development Conference; HORIZONS Student Support for 1st generation students; Purdue's chapter of the Society for Advancement of Chicanos and Native Americans in Science (SACNAS); the Louis Stokes Alliance for Minority Participation (LSAMP) Indiana project; and the Science Diversity Office.
141	-	BIGDATA: Collaborative Research: F: Big Data, It's Not So Big: Exploiting Low-Dimensional Geometry for Learning and Inference	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstr	IIS	This research will leverage ideas from algebraic and differential geometry to address core problems in modern high-dimensional and massive data science. The project will develop statistical methods and numerical tools, grounded in solid mathematical, statistical, and computational foundations, to extract low dimensional geometry from massive data with applications in clustering, data summarization, prediction, dimension reduction, and visualization. The solutions developed as part of this project can result in fundamental advances in practical applications across fields as diverse as biology, medicine, social sciences, communication networks, and engineering. In addition to internal validation via statistical and mathematical theory and simulation studies, the methods developed in the project will involve external validation via interdisciplinary applications. These applications include: (1) inference of population structure from genomic data; (2) document analysis via topic models; and (3) inference of subsets of putative gene networks relevant to drug resistance in melanoma.The research is motivated by the central premise that, even though the amount of data may be massive, a compact model can represent these data. Specifically, high-dimensional and/or massive data can be reasonably approximated by a mixture of subspaces, for which sparse representations exist. A mixture of subspaces of potentially different dimensions is a flexible, rich representation of data with nice mathematical properties that can scale to large data. There are several fundamental challenges in modeling mixtures of subspaces that will be addressed in this research: 1) the subspaces will be of different dimensions, 2) both the subspace parameters and the mixing parameters need to be inferred, 3) efficient algorithms for inference are required for both high-dimensional and massive data. The central foundational impediment in all of these challenges is that the model is a stratified space (a union of manifolds), and therefore has singularities. The key insight in this research is that there exist embeddings and representations of the model space that mitigate these singularities. These ideas are implemented as concrete Bayesian, frequentist, and numerical algorithms and models to address the real world

			ument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		examples listed above.
142	-	QuBBD: Collaborative Research: Advancing mHealth using Big Data Analytics: Statistical and Dynamical Systems Modeling of Real-Time Adaptive m-Intervention for Pain	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	DMS	With the growing popularity of mobile phone technology, new opportunities have arisen for real-time adaptive medical intervention. The simultaneous growth of multiple "big data" sources (e.g., mobile health data, electronic health records, lab test results, genomic data) allows for the development of personalized recommendations. This award supports initiation of a collaborative research project that will generate a new mathematical model for changes in subjective pain over time in patients with chronic conditions. The model will be combined with statistical techniques to ultimately obtain optimized, continuously-updated treatment plans balancing competing demands of pain reduction and medication minimization. Those resulting personalized treatment plans will be incorporated into a currently active pilot study on mobile intervention in patients living with chronic pain due to sickle cell disease (SCD). Since nearly a quarter of patient visits to the emergency room are for conditions that could have been managed as outpatients, it is crucial to improve mobile health technologies to allow these patients to quickly recognize and receive appropriate health care information. There currently is no standard algorithm or analytical method for real-time adaptive treatment recommendations for chronic conditions like pain. Furthermore, current state-of-the-art methods have difficulty in handling continuous-time decision optimization using big data. The proposed model will consist of a dynamical systems approach using differential equations to forecast future pain levels, as well as a statistical approach tying system parameters to patient data (including reported pain levels, medication history, personal characteristics and other health records). A third key component will be the development and pilot study of a new control and optimization strategy to balance the competing demands of pain reduction and drug dosage minimization. This award is supported by the National Institutes of Health Big Data to Knowledge (BD2K) Initiative in partnership with the National Science Foundation Division of Mathematical Sciences.

			DateOperator=&ExpDateOperator=		
143	-	QuBBD: Collaborative Research: Advancing mHealth using Big Data Analytics: Statistical and Dynamical Systems Modeling of Real-Time Adaptive m-Intervention for Pain	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstr	DMS	With the growing popularity of mobile phone technology, new opportunities have arisen for real-time adaptive medical intervention. The simultaneous growth of multiple "big data" sources (e.g., mobile health data, electronic health records, lab test results, genomic data) allows for the development of personalized recommendations. This award supports initiation of a collaborative research project that will generate a new mathematical model for changes in subjective pain over time in patients with chronic conditions. The model will be combined with statistical techniques to ultimately obtain optimized, continuously-updated treatment plans balancing competing demands of pain reduction and medication minimization. Those resulting personalized treatment plans will be incorporated into a currently active pilot study on mobile intervention in patients living with chronic pain due to sickle cell disease (SCD). Since nearly a quarter of patient visits to the emergency room are for conditions that could have been managed as outpatients, it is crucial to improve mobile health technologies to allow these patients to quickly recognize and receive appropriate health care information. There currently is no standard algorithm or analytical method for real-time adaptive treatment recommendations for chronic conditions like pain. Furthermore, current state-of-the-art methods have difficulty in handling continuous-time decision optimization using big data. The proposed model will consist of a dynamical systems approach using differential equations to forecast future pain levels, as well as a statistical approach tying system parameters to patient data (including reported pain levels, medication history, personal characteristics and other health records). A third key component will be the development and pilot study of a new control and optimization strategy to balance the competing demands of pain reduction and drug dosage minimization. This award is supported by the National Institutes of Health Big Data to Knowledge (BD2K) Initiative in partnership with the National Science Foundation Division of Mathematical Sciences.

			ument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
144	-	QuBBD: Collaborative Research: Advancing mHealth using Big Data Analytics: Statistical and Dynamical Systems Modeling of Real-Time Adaptive m-Intervention for Pain	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstr	DMS	With the growing popularity of mobile phone technology, new opportunities have arisen for real-time adaptive medical intervention. The simultaneous growth of multiple "big data" sources (e.g., mobile health data, electronic health records, lab test results, genomic data) allows for the development of personalized recommendations. This award supports initiation of a collaborative research project that will generate a new mathematical model for changes in subjective pain over time in patients with chronic conditions. The model will be combined with statistical techniques to ultimately obtain optimized, continuously-updated treatment plans balancing competing demands of pain reduction and medication minimization. Those resulting personalized treatment plans will be incorporated into a currently active pilot study on mobile intervention in patients living with chronic pain due to sickle cell disease (SCD). Since nearly a quarter of patient visits to the emergency room are for conditions that could have been managed as outpatients, it is crucial to improve mobile health technologies to allow these patients to quickly recognize and receive appropriate health care information. There currently is no standard algorithm or analytical method for real-time adaptive treatment recommendations for chronic conditions like pain. Furthermore, current state-of-the-art methods have difficulty in handling continuous-time decision optimization using big data. The proposed model will consist of a dynamical systems approach using differential equations to forecast future pain levels, as well as a statistical approach tying system parameters to patient data (including reported pain levels, medication history, personal characteristics and other health records). A third key component will be the development and pilot study of a new control and optimization strategy to balance the competing demands of pain reduction and drug dosage minimization. This award is supported by the National Institutes of Health Big Data to Knowledge (BD2K) Initiative in partnership with the National Science Foundation Division of Mathematical Sciences.

			ument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
145	-	QuBBD: Collaborative Research: Advancing mHealth using Big Data Analytics: Statistical and Dynamical Systems Modeling of Real-Time Adaptive m-Intervention for Pain	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstr	DMS	With the growing popularity of mobile phone technology, new opportunities have arisen for real-time adaptive medical intervention. The simultaneous growth of multiple "big data" sources (e.g., mobile health data, electronic health records, lab test results, genomic data) allows for the development of personalized recommendations. This award supports initiation of a collaborative research project that will generate a new mathematical model for changes in subjective pain over time in patients with chronic conditions. The model will be combined with statistical techniques to ultimately obtain optimized, continuously-updated treatment plans balancing competing demands of pain reduction and medication minimization. Those resulting personalized treatment plans will be incorporated into a currently active pilot study on mobile intervention in patients living with chronic pain due to sickle cell disease (SCD). Since nearly a quarter of patient visits to the emergency room are for conditions that could have been managed as outpatients, it is crucial to improve mobile health technologies to allow these patients to quickly recognize and receive appropriate health care information. There currently is no standard algorithm or analytical method for real-time adaptive treatment recommendations for chronic conditions like pain. Furthermore, current state-of-the-art methods have difficulty in handling continuous-time decision optimization using big data. The proposed model will consist of a dynamical systems approach using differential equations to forecast future pain levels, as well as a statistical approach tying system parameters to patient data (including reported pain levels, medication history, personal characteristics and other health records). A third key component will be the development and pilot study of a new control and optimization strategy to balance the competing demands of pain reduction and drug dosage minimization. This award is supported by the National Institutes of Health Big Data to Knowledge (BD2K) Initiative in partnership with the National Science Foundation Division of Mathematical Sciences.

			ument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
146	-	BIGDATA: Collaborative Research: F: IA: Statistical Learning for Big Data with Random Projections	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstr	IIS	Contemporary data-driven science and engineering problems require the development of statistical methods that do not compromise statistical accuracy, yet are computationally feasible. Data quality, particularly the heterogeneity in data measurements, is a critical factor that affects statistical accuracy in the analysis of large datasets. This project will explore and demonstrate the impact and feasibility of improving computational and statistical performances simultaneously for Big Data problems with massive datasets. The research will advance the state of knowledge in predictive statistical learning with Big Data, and be extremely valuable in applications related to financial risk management or commercial operations employing recommender systems, biology, and image analysis. A key phenomenon motivating this project is the notion that some refined ensemble methods combined with random projections can simultaneously enable the fast analysis of massive data while enhancing statistical performance. Specifically, the aims of the project are: (1) Develop new classification methods based on random projections and the random forest. By defining appropriate projections, the proposed method is shown to improve statistical accuracy for massive datasets with a large number of irrelevant noisy measurements. The theoretical properties of this method will be analyzed, and an adaptive version of the algorithm developed to optimize the computational and statistical efficiency gains; (2) Propose boosting algorithms with random projections. The statistical properties, practical performance, and implementation of the proposed random projected boosting algorithms will be investigated; (3) Develop classification methods with heterogeneities. A classification method that involves the weighted bootstrap and ensemble learning to handle heterogeneity or covariate shifts in measurements in large datasets will be developed. The random projection method will be applied to improve the proposed method for high-dimensional datasets.

			ument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
147	-	A Novel Statistical Framework for Big Data Prediction	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	DMS	Recent advances in genome-wide association studies (GWAS) have led to both an increase in the size of genetic data available and identification of important genetic variants responsible for a variety of diseases. Prediction for these genetic diseases has also become of paramount importance. However, prediction for big data such as GWAS is not trivial. A key obstacle in big data prediction is identifying (perhaps a small number of) variable sets that lead to good prediction when variable dimensionality can be extremely large. The project explores why a common approach towards prediction can often fail to deliver strong prediction rates. A novel, interaction-based and prediction-oriented approach to extracting hidden information contained in big data will be investigated. To improve prediction, a new criterion to guide the selection of variable sets will be developed.Prioritizing predictivity, not significance, requires using the correct estimates of prediction rates and developing predictivity-based criteria to evaluate variable sets. The project offers a novel theoretical framework by characterizing what makes for highly predictive variable sets, and providing fundamental work towards a new criterion to identify these sets. In the framework of this research project, variable sets have theoretical (true) levels of predictivity, which can be estimated with appropriately designed sample-based measures. This framework is the first that seeks to develop estimators specific to a criterion of predictivity. Additionally, methods that encompass both marginal and joint effects will be investigated, and a candidate measure of predictivity will be studied. Four real data examples are analyzed to illustrate how final predictors found via the new approach compare to other approaches in the current literature.

			DateOperator=&ExpDateOperator=		
148	-	BIGDATA: Collaborative Research: F: IA: Statistical Learning for Big Data with Random Projections	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Contemporary data-driven science and engineering problems require the development of statistical methods that do not compromise statistical accuracy, yet are computationally feasible. Data quality, particularly the heterogeneity in data measurements, is a critical factor that affects statistical accuracy in the analysis of large datasets. This project will explore and demonstrate the impact and feasibility of improving computational and statistical performances simultaneously for Big Data problems with massive datasets. The research will advance the state of knowledge in predictive statistical learning with Big Data, and be extremely valuable in applications related to financial risk management or commercial operations employing recommender systems, biology, and image analysis. A key phenomenon motivating this project is the notion that some refined ensemble methods combined with random projections can simultaneously enable the fast analysis of massive data while enhancing statistical performance. Specifically, the aims of the project are: (1) Develop new classification methods based on random projections and the random forest. By defining appropriate projections, the proposed method is shown to improve statistical accuracy for massive datasets with a large number of irrelevant noisy measurements. The theoretical properties of this method will be analyzed, and an adaptive version of the algorithm developed to optimize the computational and statistical efficiency gains; (2) Propose boosting algorithms with random projections. The statistical properties, practical performance, and implementation of the proposed random projected boosting algorithms will be investigated; (3) Develop classification methods with heterogeneities. A classification method that involves the weighted bootstrap and ensemble learning to handle heterogeneity or covariate shifts in measurements in large datasets will be developed. The random projection method will be applied to improve the proposed method for high-dimensional datasets.

			DateOperator=&ExpDateOperator=		
149	-	EAGER: Big Data Consulting Services and Training Center	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	ACI	AbstractThis Project will support the establishment of a Big Data Consulting Services and Training Center at the University of Georgia (UGA). The goal of this center is to provide faculty and students with basic consultation and training services in order to help them more efficiently meet their data collection, processing, storage, access, sharing, curation, and publications needs. The project proposes to focus on three areas: 1) the use of high performance computing; 2) data management; and 3) cross-disciplinary coordination. Many of the elements are already in place at UGA to support data intensive research. For example, the Georgia Advanced Computing Resource Center (GACRC) and the CUDA Teaching Center already offer resources and some training on the use of high performance computing; the UGA Libraries offer services to help researchers write data management plans; and there are a variety of campus-wide and national resources that researchers can use for their data generation, analysis, and management needs. However, currently there is no clearinghouse or coordinating organization to which any researcher on campus can turn into in order to best incorporate data-intensive approaches into their research. To meet these needs, the project plans to: Form a campus-wide committee comprised of faculty, administrators, information technology (IT) staff, librarians, and students. Survey the campus in order to get a better understanding of data management needs. Visit with individuals who are working with data that are large, long-term, and/or structurally complex to learn about their needs in more detail and identify the campus and national resources that are available for big data users. The project will then develop a set of web resources that can serve as both a central resource for the campus and a starting point for further efforts to enhance UGA's cyberinfrastructure capabilities. Finally, the project will continue to conduct the popular university-wide "big data" events under the auspices of the center to aid in communicating, coordinating, and disseminating activity around data-intensive research.

			DateOperator=&ExpDateOperator=		
150	-	Rare and Weak Signals in Big Data: How to Find Them and How to Use Them	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	DMS	A research effort is proposed to create new tools for high dimensional data analysis, focusing on the very challenging regime where signals are both rare and weak. In particular, the proposer proposes to: (a). Develop graphlet screening as a new tool for high dimensional variable selection, introduce a new theoretic framework for assessing the optimality of variable selection, and show that graphlet screening achieves the optimal rate of convergence in terms of Hamming distance of the selection errors. (b). Develop a new method of spectral clustering by using the recent idea of Higher Criticism thresholding, and investigates the fundamental limits for several problems related to low-rank matrix recovery, including high dimensional clustering, sparse Principle Component Analysis, and a testing problem related to the underlying large-size covariance matrix. (c) Extend and apply the proposed methods and theory to the analysis of Big data generated in various scientific fields, including genomics and machine learning. We are often said that we are entering the era of 'Big Data', where massive datasets consisting of millions of observations are mined for associations and patterns. What is never said about this pervasive trend is that, unfortunately, the signal we are looking for is usually very rare and weak and is hard to find, and it is easy to be fooled. The project introduces new ideas, new tools, and novel theory that are appropriate for rare and weak signals in Big Data, and apply the theory and methods to various scientific fields, including genomics and machine learning.

			DateOperator=&ExpDateOperator=		
151	-	SHF: Small: Languages and Abstraction for Dynamic Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	CCF	The analysis of large datasets using computers, a.k.a., big data analytics, is emerging as an important tool in many fields, such as science and discovery, technology, health care, and commerce. The data sets used in such applications are usually dynamic: they change over time as new data becomes available. Such dynamic changes are often small, requiring similarly small but potentially important updates, because new information can be crucial in detecting a pattern or an anomaly. For example, the Internet or a social network changes dynamically as new web pages become available, new links are added, or existing links are removed. As a result of such dynamic changes, two clusters of previously disconnected web sites can become connected by the addition of a single link, indicating for example, an important news item or a security breach. Unfortunately, in many existing big-data systems, absorbing new information involves making one or more passes over the entire dataset. Such batch processing of dynamic data results in slow updates, as well as inefficiencies in the utilization of resources such as hardware and energy, by (unnecessarily) performing many subcomputations that are unaffected by changes. This project aims to lay the groundwork for the programming languages and software systems that can support the development of such applications in the real world. The work has the potential to transform the way the programmers express computations on dynamically changing big data sets, make it possible to derive new information and knowledge from big dynamic data sets by computing with them responsively and efficiently, and transform the way that we teach the design, analysis, and implementations of computations operating for dynamic data sets. The project also includes the development of undergraduate lectures on parallelism.The project aims to enable the user to express the dynamism in large data sets implicitly, without concerning themselves with how exactly the results will be updated when the data changes, e.g., which data depends on which other data, which data may need to be updated, which dependencies need to be reconstructed. Starting with an implicitly dynamic program, a software system automatically and efficiently constructs a record of the computed results and updates it as the dataset changes. To achieve this goal, the project develops abstractions, programming languages, compilers, and run-time systems. Concretely, we expect three sets of contributions: novel, powerful abstractions and cost models for writing programs that operate on dynamically changing large datasets, programming language support in the form of compilers and run-time systems for realizing such abstractions on practical hardware, and efficient algorithms and

			DateOperator=&ExpDateOperator=		implementations, to be used to evaluate the proposed and future work.
152	-	BIGDATA: Small: Big Data for Everyone	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Although big data has had a huge impact in several areas, this impact is limited by the high cost and poor quality of analyzing unstructured data, and the costs of integrating data of multiple types. Lowering these costs will bring the benefits of big data based research to many new areas. Against this background, this project aims to develop machine-learning methods that read, analyze, and integrate web-scale collections of text and other data. The project can be expected to yield fundamental advances in data integration, machine learning, natural language understanding, and automated inference. The project includes research thrusts in (1) robust semi-supervised bootstrap learning algorithms that can cope with ambiguity in text, (2) algorithms for detecting and aligning the schemas implicit in semi-structured sources relative to a shared common ontology, (3) NLP algorithms that perform deeper analysis on text to extract infrequently mentioned yet important facts, and (4) targeted reading agents capable of pursuing specific queries or conjectures based on the scientist's current focus. Anticipated results of the project include fundamental advances in each of the research thrusts and their synergistic integration into software system (NESSIE) designed to help scientists in exploring scientific hypotheses in their respective domains of interest, by supporting targeted extraction of knowledge from large amounts of textual sources in relevant areas. Broader impacts of the research include advanced techniques for extracting and organizing structured knowledge from text, and integrate the learned information with existing structured knowledge in multiple domains. The Additional broader impacts of the research include enhanced opportunities fore advanced research-based training of graduate students. The softare and data resulting from the research will be made freely available to the larger scientific community.

			DateOperator=&ExpDateOperator=		
153	-	BIGDATA: Mid-Scale: DCM: Collaborative Research: Eliminating the Data Ingestion Bottleneck in Big Data Applications	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Big-data practice suggests that there is a tradeoff between the speed of data ingestion, the ability to answer queries quickly (e.g., via indexing), and the freshness of data. This perceived tradeoff lies, for example, at the heart of the historic division between OLTP (online transaction processing) and OLAP (online analytical processing). In an OLTP database, data gets ingested quickly and the data available for querying is fresh, but analytical queries run prohibitively slowly. In an OLAP data warehouse, data is buffered for off-line indexing so that analytical queries run quickly, but by the time the data gets indexed, it is stale. This tradeoff has manifestations in the design of all types of storage systems. For example, some file-systems are optimized for reads and others for writes, but workloads generally involve a mixture of reads and writes. In this project the PIs show that this is not a fundamental tradeoff, but rather a tradeoff imposed by the choice of data structure. The PIs use write-optimized structures, an alternative to traditional indexing methodologies, to build storage systems in which this tradeoff is significantly mitigated or alleviated altogether. The performance promise of such indexing schemes follows from the PIs previous work establishing that write-optimized data structures can speed up both inserts and queries. This project addresses the remaining obstacles in the deployment of write-optimized indexes within big-data file-systems and databases. Big data imposes a new set of constraints on any storage system, and the PIs will show how write-optimized indexing can yield order-of-magnitude performance improvements at scale. In particular, this project will show that such techniques are not only applicable today but that they will scale with hardware trends, including the widespread adoption of solid-state disks (SSDs).

			DateOperator=&ExpDateOperator=		
154	-	WORKSHOP: Open Data/Private Persons: Forging a New Social Contract for Biomedicine in an Age of Genomics and Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	SES	Genomics and 'big data' approaches to health today promise a better understanding of the pathology and treatment of a wide range of diseases, from cancer to Alzheimer's. However, mining genomic and other digital health data require practices for collecting and processing large amounts of biomedical data that challenge longstanding commitments to privacy and ownership and control of one's own body. This clash of ethos of sharing with those of privacy and autonomy has sparked a public debate about who owns bodily tissues and how health-related data should be collected and shared for the public good. To bring these growing concerns to the level of national policy, we will convene a two-day long agenda-setting workshop at UC Santa Cruz. The workshop will bring together international leaders in genomics research, women?s health, informatics, healthcare, civil rights, indigenous rights, science policy and the sociology of medicine. The goal will be to innovate novel approaches to navigating the tensions between rights of privacy and consent and desires to share biomedical data.The focus of the workshop will be on how big data research infrastructures might be built that benefit diverse populations, especially Native Americans and indigenous peoples, respecting different approaches to owning and interpreting bodily tissues and genomic data. This project thus develops novel policy approaches to questions of science and justice, utilizing and enhancing the infrastructure of the UCSC Science and Justice Research Center (SJRC). The results of the research will be translated to broader contexts through publishing results in biomedical journals and the popular press. Videos of workshops and publications will be made available on the SJRC website.

			DateOperator=&ExpDateOperator=		
155	-	BIGDATA: EAGER: Deep Learning in Higher Education Big Data to Explore Latent Student Archetypes and Knowledge Profiles	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Deep Learning in Higher Education Big Data to Explore Latent Student Archetypes and Knowledge ProfilesData science techniques have revolutionized many academic fields and led to terrific gains in the commercial sector. They have to date been underutilized in solving critical problems in the US educational system, particularly in understanding Science, Technology, Engineering and Mathematics (STEM) learning and learning environments, broadening participation in STEM, and increasing retention for students traditionally underserved in STEM. The goals of the Directorate for Education and Human Resources (EHR), through the EHR Core Research program, for the Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA) program are to advance fundamental research aimed at understanding and solving these critical problems, and to catalyze the use of data science in Education Research. Deep learning is a relatively novel method in machine learning (ML) that has shown great improvements over prior ML approaches in successful classification with very little a priori definitions applied to the data. For example, in image classification, these techniques have a high rate of success at distinguishing species and particular animals. This exploratory proposal will investigate the possibilities of using this approach with educational data from Massive Open Online Courses (MOOCs) and Learning Management Systems (LMSs). Research has shown that it is often difficult to use traditional study designs with MOOC data as enrollment is open and many different types of people enroll in MOOCs. This factor has made it difficult to make these MOOCs adaptive to individual learners, an optimal approach to improving learning. The Principal Investigators will group people in MOOCs who are similar to each other to understand the different types of people taking the MOOCs so that learning activities can be tailored to them.The Principal Investigators will examine a variety of data, from the micro-level of backend data on timing to complete an exercise, or pauses between activities, to macro-level data such as course history and grades. They will use deep learning techniques to identify groups of learners with similar characteristics, the first step in making a learning environment more adaptive. The proposed research is ambitious and risky. Deep learning techniques show great promise and the Principal Investigators demonstrate that they could be extremely helpful in solving key educational challenges. The Principal Investigators have the technical and learning science expertise to carry out this ambitious endeavor. As more and larger educational datasets are developed, the field needs to expand its methodologies to learn from them. This proposal is unique in its potential for catalyzing this effort.This award is supported by the EHR

			DateOperator=&ExpDateOperator=		Core Research (ECR) program. The ECR program emphasizes fundamental STEM education research that generates foundational knowledge in the field. Investments are made in critical areas that are essential, broad and enduring: STEM learning and STEM learning environments, broadening participation in STEM, and STEM workforce development.
156	-	BIGDATA: F: DKM: Collaborative Research: PXFS: ParalleX Based Transformative I/O System for Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Recent decades have seen the development of computational science where modeling and data analysis are critical to exploration, discovery, and refinement of new innovations in science and engineering. More recently the techniques have been applied to arts, social, political and other fields less traditionally reliant on high performance computing. This innovation has grown out of realization some 20 years ago that I/O (input/output) support for high performance parallel and distributed architectures had lagged behind that of pure computational speed, and further that bring I/O up to speed was both critical, and a rather difficult problem. The core hurdle of contemporary I/O on large HPC machines relates to issues of latency in large parts caused by the deficiencies of the historical I/O model that was relevant when computers were exclusively large, centralized, single processor systems shared by many time-sharing programs. In order to improve I/O on scalability on future hardware architectures novel approaches are required.This project is conducting research on an extension of ParalleX, a new highly innovative parallel execution model. The extension provides a powerful I/O interface that allows researchers to create highly efficient data management, discovery, and analysis codes for Big Data applications. This new extension, known as PXFS, is based on HPX, an implementation of ParalleX based on C++, and OrangeFS, a high performance parallel file system. The research goal driving PXFS is to extend HPX objects into I/O space so that the objects become persistent and storage becomes another class of memory, all accessed as a single virtual address space and managed by an event driven dynamic adaptive computation environment. Critical aspects of this approach include futures-based synchronization, dynamic locality management, dynamic resource management, hierarchical name space, and an active global address space (AGAS). The overall goals of PXFS are to eliminate the division of programming imposed by conventional file system through the unification of name spaces and their management, and to minimize global synchronization in order to support asynchronous concurrency. The research methodology is to implement a Map/Reduce application framework using PXFS and evaluate its effectiveness in both performance and ease of use.This project is conducted at three major research universities involving undergraduate and graduate students, post-docs, and high-school teachers and their students. The project includes a PI from the functional genomics field acting as domain science expert in order to focus the development efforts on real world problems. Graduate students and post-docs involved in the project are trained in these areas to promote scientists who understanding both aspects of Big Data problems. The project engages under represented minorities with the

			DateOperator=&ExpDateOperator=		goal to inspire them to pursue a career in computer science or genomics. The software developed by the project is available open-source and archived using an integrated source code revision repository, wiki, and bug tracking software system in addition to code releases with accompanying documentation.
157	-	BIGDATA: F: DKM: Collaborative Research: PXFS: ParalleX Based Transformative I/O System for Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperato	IIS	Recent decades have seen the development of computational science where modeling and data analysis are critical to exploration, discovery, and refinement of new innovations in science and engineering. More recently the techniques have been applied to arts, social, political and other fields less traditionally reliant on high performance computing. This innovation has grown out of realization some 20 years ago that I/O (input/output) support for high performance parallel and distributed architectures had lagged behind that of pure computational speed, and further that bring I/O up to speed was both critical, and a rather difficult problem. The core hurdle of contemporary I/O on large HPC machines relates to issues of latency in large parts caused by the deficiencies of the historical I/O model that was relevant when computers were exclusively large, centralized, single processor systems shared by many time-sharing programs. In order to improve I/O on scalability on future hardware architectures novel approaches are required.This project is conducting research on an extension of ParalleX, a new highly innovative parallel execution model. The extension provides a powerful I/O interface that allows researchers to create highly efficient data management, discovery, and analysis codes for Big Data applications. This new extension, known as PXFS, is based on HPX, an implementation of ParalleX based on C++, and OrangeFS, a high performance parallel file system. The research goal driving PXFS is to extend HPX objects into I/O space so that the objects become persistent and storage becomes another class of memory, all accessed as a single virtual address space and managed by an event driven dynamic adaptive computation environment. Critical aspects of this approach include futures-based synchronization, dynamic locality management, dynamic resource management, hierarchical name space, and an active global address space (AGAS). The overall goals of PXFS are to eliminate the division of programming imposed by conventional file system through the unification of name spaces and their management, and to minimize global synchronization in order to support asynchronous concurrency. The research methodology is to implement a Map/Reduce application framework using PXFS and evaluate its effectiveness in both performance and ease of use.This project is conducted at three major research universities involving undergraduate and graduate students, post-docs, and high-school teachers and their students. The project includes a PI from the functional genomics field acting as domain science expert in order to focus the development efforts on real world problems. Graduate students and post-docs involved in the project are trained in these areas to promote scientists who understanding both aspects of Big Data problems. The project engages under represented minorities with the goal to inspire them to pursue a career in computer science or genomics. The software developed by the project is available open-source and archived using an integrated source code revision repository, wiki, and bug tracking

			r=		software system in addition to code releases with accompanying documentation.
158	-	From Approximate to Exact Designs with Applications to Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	DMS	Design of experiments is an integral part of the scientific process in many areas of research with a direct impact on society, such as the biological sciences, the health sciences, the social sciences, engineering, marketing, and education. A well-chosen design facilitates the collection of data that, at a minimum cost, maximizes the information for the scientific questions of interest. Many scientific studies allow for repeated use of conceptual units, so that developing tools for optimal design for these problems has great potential impact. Particularly in the realm of big data, there is much room for improvement of existing methods for design of experiments, and the tools and concepts under development in this research project have potential to lead to significant gain of information without increasing computational cost. Results from the project will be made available to researchers in other areas through easy-to-use software that implements the algorithms to be developed. Graduate students will be trained to become researchers in design of experiments. This project aims to result in a major leap forward in understanding and knowledge of optimal design of experiments. Recent work in the field has had a significant impact on the advancement of optimal crossover designs and designs for interference models for arbitrarily given covariance structures and design size configurations. However, these results have for the most part been limited to approximate designs for relatively simple models. While these results are arguably important in their own right, this project will extend methods and tools to achieve the ultimate goal of deriving exact designs for a wider spectrum of practical models. The results will be a much needed addition to our collective design toolbox. Most importantly, this project will go beyond the territory of design and apply the tools and ideas from design of experiments to subsampling problems emerging in big data with both statistical and machine learning methods under consideration. Preliminary results indicate that this is an opportune time to make these challenging but critical steps.

			DateOperator=&ExpDateOperator=		
159	-	BIGDATA: EAGER: Using Big Data to Investigate Longitudinal Education Outcomes through Visual Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	BIGDATA: Using Big Data to Investigate Longitudinal Education Outcomes through Visual AnalyticsData science techniques have revolutionized many academic fields and led to terrific gains in the commercial sector. They have to date been underutilized in solving critical problems in the US educational system, particularly in understanding Science, Technology, Engineering and Mathematics (STEM) learning and learning environments, broadening participation in STEM, and increasing retention for students traditionally underserved in STEM. The goals of the Directorate for Education and Human Resources (EHR), through the EHR Core Research program, for the Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA) program are to advance fundamental research aimed at understanding and solving these critical problems, and to catalyze the use of data science in Education Research. This Early Concept Grant for Exploratory Research (EAGER) will employ data from national representative datasets including approximately 35,000 students to investigate course taking patterns in high school and how these relate to critical outcomes such as college attendance, high school and college success, and career choices. Few investigators have attempted to answer this question with this scale of data. Therefore, this proposal will contribute significantly to the field's understanding of factors that affect success in high school and college. Building on these new insights will enable the potential to create interventions at the high school and college level based on data about what works to improve graduation and workforce outcomes. The Education Longitudinal Study of 2002-2012 (ELS:2002) and High School Longitudinal Study of 2009 (HSLS:2009) are representative samples of high school students who were tracked through college and into their early career. ELS has 15,000 students and HSLS has 21,000 students. For ELS, NCES surveyed students in 10th and 12th grade, then sophomore year of college, and then when they were approximately 26. NCES collected transcripts for high school and college, and high school and college achievement, attitudinal, participation and demographic variables. The Principal Investigator will first use visualizations to identify potential patterns of difference in outcomes and then use growth mixture modeling (GMM) and receiver operating characteristic analysis (ROC) to investigate the statistical significance of those patterns.This award is supported by the EHR Core Research (ECR) program. The ECR program emphasizes fundamental STEM education research that generates foundational knowledge in the field. Investments are made in critical areas that are essential, broad and enduring: STEM learning and STEM learning environments, broadening participation in STEM, and STEM workforce

			DateOperator=&ExpDateOperator=		development
160	-	EAGER-DynamicData: A Hierarchical Approach to Dynamic Big Data Analysis in Power Infrastructure Security	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	ECCS	This research will address a dynamic big data problem that is of urgent national interest: the need for efficient methods to diagnose faults and attacks in critical interconnected infrastructures, such as electricity power networks. Additionally, this project will investigate new methodologies to extract knowledge from the complex streams of data that come from various sensors in infrastructure systems and the models of their behavior. Results and findings in this project will be validated via industry-accredited power system simulators, and will be useful to the power industry in enhancing the safety, stability, and security of essential power infrastructure. This project will promote multi-disciplinary research involving expertise in big data analysis, machine learning, security, power systems, and control systems. This research will provide a powerful bridge between theory and real-world applications while serving as a training platform for a diverse new generation of engineers at the University of California, Riverside, one of America's most ethnically diverse research-intensive institutions.This project will foster the use of multi-resolution data-driven methods for the detection and classification of anomalies in critical dynamical infrastructures, with focus on power networks. This project has three novel, innovative, and potentially transformative technical elements: (1) A comprehensive statistical model, as an alternative to existing physics-based models, using Dynamic Bayesian Networks and Conditional Random Fields to model complex infrastructures subject to failures and attacks; (2) A hierarchical detection and classification method based upon machine learning concepts to tame and leverage the vast amount and diversity of dynamic multi-resolution data collected by spatially distributed sensors; (3) A systematic method to train and inform data-driven methodologies from model-based and analytical knowledge that come from power systems and control theory to build scalable and performing detection and classification mechanisms in power infrastructure security.

			DateOperator=&ExpDateOperator=		
161	-	SBIR Phase I: Big data Analytics Driven Adaptive Learning for STEM Education	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIP	This SBIR Phase I project will develop a big-data analytics-based adaptive online learning platform founded on the principles of adaptive learning. The online learning system will provide adaptive learning strategies with real-time learner analytics. The systems will integrate learner analytics from four dimensional aspects of learning multi-media, learning strategies, interactivity, and social interaction-- to deliver a personalized learning experience for science, technology, engineering, and math (STEM) students for significant improvement in the learning outcome. The adaptive learning software technology platforms with personalized learning strategies have demonstrated high completion and satisfaction rates for online students taking post-secondary courses. In this SBIR Phase I we propose to develop a unique data driven decision support interface that will result in real-time big data analytics for both individuals and large numbers of learners. The volume, velocity, variety, and veracity (the 4 Vs of big data) will be generated by the collection of data at individual schools first, with the potential to aggregate data from district, state, and even national levels. The big-data analytics of the learner trajectories through the adaptive learning platform will uncover patterns that can improve understanding of learner behavior in education.Big data, generated by the adaptive learning systems related to learner behavior in each learning strategy, will lead to valuable insights on efficacy of the proposed methodology and development of the product in Phase II. The learner analytics will provide the basis for intelligent feedback based on the statistical evidence. The proposed method of data driven decision process for adaptive learning is based on the real- time cross-correlation statistical analysis of the predictor variables for the an individual learner. The field trials of the proposed method will be conducted in the participating high schools. Collection of data for a group of students collected during the field trials in high schools will lead to discovery of learning patterns for the clusters of learners in each learning strategy. The proposed research will then lead to the development of an analytical model for volume, velocity, variety, and veracity of data collected at school-wide level. This data then will be used for the development of decision tree and regression analysis to find correlations (knowledge discovery) that can be used for the improvement of learning enterprise (school). The data-driven feedback to students and the group analytics for teachers will provide necessary feedback mechanism for improving competency and graduation rates for STEM education in schools and colleges.

			DateOperator=&ExpDateOperator=		
162	-	Planning Grant: I/UCRC for Computation Intensive Big Data Analytics for Multimodal Temporal Prediction, Retrieval, and Attribution	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIP	Recently, both leading technological firms and academic institutions have developed a new capability to capture and process data at a scale many orders larger than previously possible. Firms, particularly in Silicon Valley, have demonstrated how this capability, when blended with Analytics, can be exploited to solve a host of new problems of great practical value. These range from Computational (or Online) Advertising, with immense commercial value being realized, to Healthcare Analytics, which is of tremendous and critical value societally. At our Research Site at the University of California, Santa Cruz, we focus on enhancing the predictions and searches based on Variety of data (in addition to Volume and Velocity), since data such as Electronic Healthcare Records (EHRs) include not just numerical data (about vitals and labs), but also text (notations by doctors and nurses), images (X-rays etc.), video (such as a doctor?s examination), body sensors, etc. We also explore and exploit ways of identifying more informative data on-the-go, as well as identifying effective ways to speed up and slow down the rate of obtaining this informative data automatically based on need and context, to achieve superior prediction. Finally, we develop new ways of evaluating the true impact of each data type/source on the desired outcome. The fundamental discoveries will concern methods for determining the value of each type or source of data in more effective prediction (and search) of dynamic system state and intervention decisions.Our research concerns analyzing multi-type and source data for enhanced prediction, search, and decision making. Our collaborative research site focuses on research and development of novel scalable Big Data Analytics (BDA) solutions for multimodal data and streaming data suitable for various computing architectures (where the knowledge extracted from multiple modes is far greater than the sum of knowledge discovered from individual data-types); and novel interoperable solutions to combine analytics tools and solutions. Specific sub-thrusts include:1. Scalable temporal prediction/classification, including recommenders, for multi-type data, incorporating latent business processes and ontologies.2. Bayesian interactive information retrieval for massive data sets, combined with extraction incorporating latent business processes and ontologies in mining data.3. Data type and source characterization in terms of power of prediction/classification/retrieval, including causality in marketplaces and temporal aspects.4. Interoperable solutions for BDA that enable seamless and efficient knowledge discovery from multimodal data involving different solutions and tools in different languages and different APIs, and Programmable File and Storage Systems for Big Data Analytics.5. User behavioral modeling analytics and the analytics/economics of decisions and interventions.The domains and contexts we propose to explore include a subset of: system health (e.g. aviation safety, jet engine maintenance) and analytic services based on heterogeneous data and data/text mining, extraction, and retrieval for service centers (e.g. in network health or financial, sales and marketing services), principled knowledge discovery for personalized healthcare and web analytics, diagnostics and prognostics in Internet of Things.

163	-	I/UCRC FRP: Collaborative Research on Event-based Analytics for Enhanced Prognostics Design in a Big Data Environment	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIP	The proposed work seeks to investigate event-based modeling to deal with high-dimensional and heterogeneous data environments in order to enhance prognostics design with adaptive control of data collection and rapid maintenance decision-making to apply data analytics to the software development process. The event-based approach is explored given its potential to reduce temporal redundancy while preserving the machine dynamics. Event-based approaches have not been fully explored for prognostics applications with continuous signal inputs, such as sensor measurements. The proposed approach represents a paradigm shift in data modeling prognostic system design and holds the potential to help address the fundamental issues of big data in the areas of volume, velocity and variety. Results will be validated using various data collected from a fleet of electric vehicles. The outcomes of the proposed work have the potential for significant impact in the manufacturing sector in the area of prognostics and health monitoring. The resulting approach has the potential to create more efficient systems that can more rapidly adapt and respond to critical issues. The work is supported by the Industry Advisory Board as well as individual industry members of the center and has the potential to extend the centers portfolio through expansion into the area of event-driven modeling, big data reduction and mining for improved industrial efficiency. The center will involve graduate students and undergraduates in the work.
164	-	BIGDATA: F: DKM: Addressing the two V's of Veracity and Variety in Big	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	IIS	Data of questionable quality have led to significantly negative economic and social impacts on organizations, leading to overrun in costs, lost revenue, and decreased efficiencies. The issues on data reliability, credibility, and provenance have become even more daunting when dealing with the variety of data, especially data that are not directly collected by an organization, but from the third-party sources such as social media, data brokers, and crowdsourcing. To address such issues, this project aims to develop a Data Valuation Engine (DVE) that solves the critical problem of data reliability, credibility and provenance, and provides accountability and quality processes

		Data	anization=&PISta te=&PIZip=&PICo untry=&ProgOrg anization=&Prog EleCode=&Boole anElement=All& ProgRefCode=&B ooleanRef=All&P rogram=&ProgOf ficer=&Keyword =%22Big+Data% 22&AwardTitleO nly=true&Award NumberOperato r=&AwardAmou nt=&AwardInstr ument=&ActiveA wards=true&Ori ginalAwardDate Operator=&Start DateOperator=& ExpDateOperato r=		right from data acquisition. The DVE leverages and innovates techniques in estimation theory, data fusion and machine learning to fill a critical gap in data accountability and quality, thereby providing a transformative step in countering the ubiquitous data quality issues found in almost every application domain from business to environment to health to national security. The DVE will be integrated in the Hadoop ecosystem and will be agnostic to the data source, application or analytics, and provided as a hosted solution to the community. The user will interact with DVE by providing the data sources and relevant data necessary to solve a problem. The DVE in this project will be developed in a largely application-independent manner. The key challenges to develop this engine include: (i) How to generate the data quality indication labels to score data sources and the content of data based on various factors such as reliability, credibility, uncertainty and confidence? (ii) How to integrate data from various sources with different labeled scores? (iii) How to robustly evaluate the proposed engine in a broad spectrum of applications that serve as a proxy of a variety of real-world scenarios? The research plan has been designed to synergistically address the above challenges with a robust evaluation plan. Given the generality of the proposed methods, models and system, the project will potentially impact variety of applications of science, engineering, and social science and have broad environmental, economic, and health benefits. The PIs will release open source software and applicable data. The PIs will also provide a hosted DVE platform for a broad user and participant base. This project is also providing students with greater exposure to the areas of big data analytics, cloud computing, data fusion and data mining, both in courses and research experiences.
165	-	Collaborative Research: Record Linkage and Privacy-Preserving Methods for Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg anization=&PISta te=&PIZip=&PICo untry=&ProgOrg anization=&Prog EleCode=&Boole anElement=All&	SES	This research project will develop sound statistical and machine learning techniques for preserving privacy with linked data. Social entities and their patterns of behavior is a crucial topic in the social sciences. Research in this area has been invigorated by the growth of the modern information infrastructure, ease of data collection and storage, and the development of novel computational data analyses techniques. However, in many application areas relevant and sensitive information is commonly located across multiple databases. Data analysis is inherently impossible without merging databases, but at the cost of increasing the risk of a privacy violation. This research will address the problem of how to perform valid statistical inference in the presence of multiple data sources, data sharing, and privacy in the age of "big data." The investigators' new modeling construct for inference and uncertainty quantification will contribute to both statistics and the many disciplines for which statistics is a principal tool. The methods will have a wide range of applications in the social, economic, and behavioral sciences, including medicine, genetics, official statistics, and human rights violations. The investigators will collaborate with post-doctoral researcher and with graduate and undergraduate students. The statistical methods will be encapsulated in

			ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		open-source software packages, allowing off-the-shelf use by practitioners while facilitating more detailed control and extensions.This interdisciplinary research project will improve upon methods in record linkage and privacy using state-of-the-art techniques from statistics and machine learning. Record linkage is the process of merging possible noisy databases with the goal of removing duplicate entries. Privacy-preserving record linkage (PPRL) tries to identify records that refer to the same entities from multiple databases without compromising the privacy of the entities represented by these records. The research will focus on three aims: (1) development of new Bayesian methods for PPRL, where the error can be propagated exactly across the entire linkage process and into statistical inference, including new privacy measures to capture a tradeoff between utility and risk of any individual risk in a linked database; (2) development of new robust methods for realizing synthetic data releases post-linkage with differential privacy guarantees and its relaxations to address additional layers of privacy and support broader data sharing; and (3) exploration of "big data" methods such as variational inference to address scalability and latent cluster exchangeability issues existing within linkage and privacy, such that the new methods can scale to multiple and large databases. The new methods will be scalable and assess uncertainty throughout the entire linkage and privacy process and can be evaluated using Bayesian disclosure risk and Bayesian differential privacy. The project is supported by the Methodology, Measurement, and Statistics Program and a consortium of federal statistical agencies as part of a joint activity to support research on survey and statistical methodology.
166	-	AF: Small: Homological Methods for Big Enough Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleO	CCF	How do we know if big data is big enough? As algorithms make more and more decisions from data, we also need these algorithms to assure us that the decisions were well-informed, i.e. that enough data went into them. The theory of homological sensor networks, a branch of topological data analysis, was originally created to test if a collection of sensors covers a domain of interest, but the same theory can test if a data set "covers" an underlying decision space. Homological methods can complement, extend, and even replace statistical methods to give confidence in the completeness of a data set. Because they are topological, they can give robust signatures or summaries of data that are invariant to a wide range of implicit or explicit transformations. This project aims to extend the theoretical and algorithmic foundations of the homological sensor networks to be applicable in data analysis. Broader impacts include strengthening connections between theoretical computer science (TCS) and applied algebraic topology, and widening the range of data analyses to which topological methods and tools apply.The PI will train both undergraduate and graduate researchers by incorporating advanced concepts in combinatorial topology in undergraduate and graduate curricula. The PI will also educate the larger TCS and data analysis communities through expository videos and open source software.The specific aim of the proposal is to extend guarantees on homological sensor networks to apply to non-smooth sets, k-coverage, and dynamic coverage. A second specific aim is to push these algorithmic results back into the theoretical foundations of the sampling theories that underlie data analysis problems, by extending the so-called Persistent Nerve Theorem and defining new classes of near-homeomorphisms to capture the realities of unknown transformations in data while still providing theoretical guarantees. The third specific aim is to develop algorithms that extract information from

			nly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		what was traditionally called "topological noise" as simple experiments reveal that although it doesn't carry topological information, it does carry useful geometric information that may be used for classification and inference.
167	-	RAPID: Modeling Ebola Spread and Developing Decision Support System Using Big Data Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	CNS	This project aims to help address urgent public health problems (specifically the spread of the Ebola virus) of national and global significance by advancing the state of the art in computer science, big data analytics, data visualization techniques, and decision support systems. Specifically, the effort, developing computational models to predict the spread of Ebola utilizing both 'forward simulation' from a given patient and the propagation of the infection into the community and backwards, aims to trace a number of the verified infections to patient 'zero.' The work utilizes big data analytic techniques, data about underlying personal relationships, health center locations, and the known mechanisms for the spread of the Ebola virus.The project connects directly to the Florida International University (FIU)?s TerraFly system, a web-enabled system designed to aid in the visualization of spatial and remotely sensed data. The system allows users to ?fly? with fine resolution over the surface of the earth to explore various kinds of data (e.g., local information, street maps, aerial photography, satellite imagery, etc.). The Ebola spread patterns are then fed into a Decision Support System (DDS). These inputs also consist of information about social groups or individual persons. Based on spread patterns, the DSS will then calculate probabilities for a social group or a given person to get infected with Ebola. The system will be able to present data mashups to operators responding to hotline calls and field workers encountering patients and deciding about triage. The data will also be presented in report form to responsible government agencies. This time-sensitive project necessitates prompt collection and analysis of the spread of the Ebola virus in order to enable the development of the correct models.

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
168	-	Topics in Analysis of Big Data and Complex Models	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	SES	The proposed research seeks to provide researchers with new methods to study economic issues of current interest, making effective use of big datasets that are only recently available. The projects deal with issues in data preprocessing, estimation, and hypothesis testing. The emphasis is on methods with broad applicability and that can be put to practical use. The proposed research is also multidisciplinary, combining methodologies from statistics with those from computer science, while providing methods for empirical researchers from any field that uses statistical methods on massive data sets.The proposed research consists of three projects. The first project develops methods for efficient and effective analysis of big data for the purpose of understanding micro and macroeconomic phenomenon. While datasets that are terabytes in size are increasingly available, resource constraints often make it necessary to study a smaller set of observations, which raises the question about how to form subsamples. The investigator will develop methods that can efficiently use large datasets, while preserving data features valuable to economic analysis. The second project provides frequentist tools to assess sensitivity of the estimation results to model assumptions and features of the data, which will be particularly useful to assess the results from complex structural models. The third project will assess whether uncertainty is a cause or a consequence of economic fluctuations. Given that there is no ideal instrument to distinguish uncertainty shocks from real activity shocks, the investigator will develop an iterative method that would purge the unwarranted variations from a potentially invalid instrument in order to arrive at a valid instrument. This generated external IV procedure can generally be used in applications when no valid instrument is available.

169	-	WORKSHOP: MASS SPECTROMETRY BIG DATA TO KNOWLEDGE	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CHE	Mass spectrometry is a chemical analysis tool that has played an important role in the advance of cell biochemistry and large-scale methods to study biological systems. This workshop, Mass Spectrometry Big Data to Knowledge, will seek to identify new mathematical and computational frameworks to advance the use of data produced by mass spectrometry. As specific software tools and methods can be drivers and enablers of new experimental paradigms, this workshop will also consider next generation ideas, for example questions that are not possible to answer now, experimentally, but may be enabled by new mass spectrometry-based approaches. This workshop is sponsored by the Division of Chemistry, the Division of Mathematical Sciences, the Division of Behavioral and Cognitive Sciences, and the Mathematical and Physical Sciences Office of Multidisciplinary Activities. The Mass Spectrometry Big Data to Knowledge Workshop, led by Professor John Yates III of the Scripps Research Institute, will examine the current challenges in bioinformatics and their relationship to mass spectrometry technology. The areas of discussion will include: 1) Bottlenecks in data analysis- where people spend their time; 2) Assessment of current methods to translate data to knowledge, and discussion of their limitations and challenges associated with application of these methods to mass spectrometry data sets; 3) The flexibility of current algorithmic methodologies in the face of changing landscapes of both instrumental and computational technologies, and 4) Where does mass spectrometry data fit in biological studies, including the discovery of mechanisms of disease? The conclusions of this workshop will be disseminated to the scientific community through websites and a workshop report.
170	-	SBIR Phase I: Big Data Analytics for Facility Operations and	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	IIP	The broader impact/commercial potential of this Small Business Innovation Research (SBIR) Phase I project is to help owners and operators of commercial and institutional buildings to improve resource allocation by analyzing data from built infrastructure to enable smarter decision-making supported by detailed, measureable, real-time knowledge. By automatically integrating building information that is stored using various software applications and formats, this innovation enables owners and facilities managers to efficiently search for information and respond to emergency and failures, and proactively plan for operation and maintenance tasks. This innovation also applies

		Management	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		artificial intelligence to automatically conduct big data analysis and identify opportunities to improve energy efficiency and operating performance of assets and indoor environment. Organizations can not only save operating budget by reducing equipment failures and energy waste, but also improve the quality of life and productivity for occupants. This Small Business Innovation Research (SBIR) Phase I project is aimed at developing middleware technology to automatically integrate and analyze both structured and unstructured data from facilities design and operations. Facilities maintenance and operating is the longest phase in the life-cycle of buildings, accounting for more than 60% of the total cost of ownership. Owners and facilities managers are faced with the challenges of efficiently managing aging and crowded building infrastructure to extend the life of assets and control costs. However, fragmented and under-analyzed building information results in most maintenance work being conducted reactively to address problems that have already caused significant loss or waste. The vision of this innovation is to develop a fully commercialized software package to enable facilities managers to be more proactive in improving building occupant comfort, aligning limited resources where they have the most significant impact, and reducing wasted energy through optimized mechanical controls. This project aims to demonstrate the conceptual feasibility of using big data analytics and machine learning to revolutionize facilities operating and maintenance decisions. The results from this applied research will include algorithms and methods to combine structured data with field collected unstructured data into qualitative and quantitative output appropriate for improved decision making.
171	-	Fast Algorithms for Solving Big Data PDE Parameter Estimation Problems on Cloud Computing Platforms	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&	DMS	Parameter estimation problems arise in many scientific and economic disciplines, for example, in medical imaging, geophysical explorations, nondestructive testing, and economic structural estimation. Despite enormous effort put into designing efficient methods, solving parameter estimation problems is still very challenging, since the parametrized equations have to be solved repeatedly until the parameters are estimated with satisfactory accuracy. This research project aims to develop and implement efficient numerical methods for solving parameter estimation problems that involve a large number of measurements and partial differential equations. Reusable, open source software will be developed and made available to the scientific community. The techniques under development in the project will be applicable in geophysics to reduce the computational costs of large surveys that are of high economic impact, for example, in oil and gas exploration and groundwater surveys. The results from this project will also be applicable in medical imaging to reduce health care screening costs and improve diagnosis of certain diseases.Parameter estimation can be formulated as an optimization problem with constraints that are given by the parametrized partial differential equations (PDEs). The unknowns are parameters of the PDEs, which correspond to

			ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		physical properties of the object to be measured. The objective is to minimize the misfit between PDE simulations and measured data plus some regularization term. Cloud computing platforms provide access to immense computational resources at moderate costs and are thus highly attractive for solving PDE parameter estimation problems. This holds particularly for big data problems since the computational costs of the estimation are dominated by the computational costs for PDE simulations. The latter, in many cases, grows linearly with the number of data. Straightforward extensions of the currently most reliable parameter estimation algorithms to massively parallel platforms, however, lead to huge communication overhead and memory requirement. This project seeks to design alternative tailored algorithms that make efficient use of cloud platforms and are able to solve parameter estimation problem with massive amounts of data in reasonable time. The approach undertaken in this project is based on three cornerstones. First, two reduced-order modeling techniques and their combination will be investigated. The PDEs will be discretized on rather coarse rectangular meshes that are aligned to the problem domain. On these meshes, reduced order models with adaptive multiscale bases will be used. Both techniques will dramatically reduce the computational cost associated with the PDE simulations. Second, stochastic optimization methods will be designed to exploit redundancy typically present in big data sets. The goal is to reduce the required number of PDE simulations, derive parameter selection rules, and quantify uncertainty of the solution. Third, the above steps will be combined and implemented on massively parallel cloud computing platforms.
172	-	Collaborative Research: Record Linkage and Privacy-Preserving Methods for Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleO	SES	This research project will develop sound statistical and machine learning techniques for preserving privacy with linked data. Social entities and their patterns of behavior is a crucial topic in the social sciences. Research in this area has been invigorated by the growth of the modern information infrastructure, ease of data collection and storage, and the development of novel computational data analyses techniques. However, in many application areas relevant and sensitive information is commonly located across multiple databases. Data analysis is inherently impossible without merging databases, but at the cost of increasing the risk of a privacy violation. This research will address the problem of how to perform valid statistical inference in the presence of multiple data sources, data sharing, and privacy in the age of "big data." The investigators' new modeling construct for inference and uncertainty quantification will contribute to both statistics and the many disciplines for which statistics is a principal tool. The methods will have a wide range of applications in the social, economic, and behavioral sciences, including medicine, genetics, official statistics, and human rights violations. The investigators will collaborate with post-doctoral researcher and with graduate and undergraduate students. The statistical methods will be encapsulated in open-source software packages, allowing off-the-shelf use by practitioners while facilitating more detailed control and extensions.This interdisciplinary research project will improve upon methods in record linkage and privacy using state-of-the-art techniques from statistics and machine learning. Record linkage is the process of merging possible noisy databases with the goal of removing duplicate entries. Privacy-preserving record linkage (PPRL) tries to identify records that refer to the same entities from multiple databases without compromising the privacy of the entities represented by these records. The research will focus on three aims: (1) development of new Bayesian

			nly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		methods for PPRL, where the error can be propagated exactly across the entire linkage process and into statistical inference, including new privacy measures to capture a tradeoff between utility and risk of any individual risk in a linked database; (2) development of new robust methods for realizing synthetic data releases post-linkage with differential privacy guarantees and its relaxations to address additional layers of privacy and support broader data sharing; and (3) exploration of "big data" methods such as variational inference to address scalability and latent cluster exchangeability issues existing within linkage and privacy, such that the new methods can scale to multiple and large databases. The new methods will be scalable and assess uncertainty throughout the entire linkage and privacy process and can be evaluated using Bayesian disclosure risk and Bayesian differential privacy. The project is supported by the Methodology, Measurement, and Statistics Program and a consortium of federal statistical agencies as part of a joint activity to support research on survey and statistical methodology.
173	-	CAREER: Towards a Big Data Application Server Stack	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	CNS	Google's MapReduce inspired much of the Big Data Analytics work and has served as a template for open source systems like Apache Hadoop. The MapReduce programming model has wide applicability, but widespread adoption has exposed some limitations, such as the lack of support for iteration (which is common in machine learning algorithms), stream processing, graph analytics, real-time and interactive queries. Beyond the programming framework, the underlying implementation offers a template for how to scale-out massively distributed computations: break them up into small tasks that can be carried out in parallel by partitioning the underlying data, and save intermediate state to mitigate the impact of partial failures (which must be planned for when running on large clusters). The challenge then, is to build implementations of other programming frameworks (e.g., SQL and machine learning) that share the same scale-out and fault-tolerance runtime characteristics of MapReduce without imposing its limitations. Resource managers such as Apache Hadoop YARN, Google Omega and Berkeley Mesos take a first step in this direction by separating resource allocation from the details of higher-level programming models and languages. Resource managers multiplex several jobs on the same underlying machine cluster, thereby increasing utilization and fostering clean-slate software stacks. When the task executing in a container a slice of a single machine's resources (CPU/GPU, memory, disk) is finished, the container is returned to the resource manager, where it is made available to other jobs. Unlike in higher-level stacks, a container is a blank-slate process, designed to host arbitrary computations. This project prescribes further reusable software layers that capture issues like how many resources should I dedicate to a job?; what are the redundant code-pathways and can I provide them in a reusable library?; what are the right language and runtime abstractions? Exploring these questions in the context of systems like MapReduce and related SQL implementations, ML toolkits, storage systems, and messaging systems, on next generation resource managers, is the primary focus of our work.The goal is to unify a suite of large-scale data processing tasks on a single runtime layer, built on modern resource managers (the cloud operating systems). Our results will factor out commonalities in specialized systems and provide them in a single underlying runtime system, shortening the time to ?market? for the next ready-to-use Big Data toolkit, which in turn would increase the availability of such tools to the broader community. Experience gained by implementing and deploying

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		applications at scale, over next generation resource managers, could help inform critical design choices in the development of future cloud computing platforms, and hence impact a broad range of scientific, engineering, national security, healthcare and business applications. The project offers enhanced opportunities for research-based advanced training of graduate and undergraduate students, including members of groups that are currently under-represented in computer science, in databases, machine learning, and cloud computing.
174	-	BIGDATA: F: DKA: DKM: Novel Out-of-core and Parallel Algorithms for Processing Biological Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	We live in an era when vast amounts of data are being generated at a low cost in several domains of science and engineering. However, advances in analytics tools have not caught up with data generation. In particular, existing tools take too much time. A main reason is that core memories of computers cannot hold all the data to be analyzed -- most of the data have to be stored in secondary storages (SSs) such as solid state drives and (rotating) disks. Data access times from SSs are several orders of magnitude more than from core memories. Tremendous speedups can be obtained by minimizing the number of data accesses from SSs. Also, although there has been much recent research in the development of multicore and GPU algorithms for biological problems, for many of the problems only sequential in-core algorithms are known.This project is to develop novel out-of-core algorithms for biological big data (BBD) analytics. The proposed novel parallel algorithms employ various architectures including heterogeneous clusters of multicores and GPUs, to solve BBD problems. The developed novel scalable algorithms can handle petabytes of data and beyond for data mining applicable over varied datasets. This interdisciplinary project provides a new computation suite for mining voluminous biological and other data. This project provides educational opportunities to graduate and undergraduate students to get first-hand research experience in computational aspects of biological data analysis.

			DateOperator=&ExpDateOperator=		
175	-	RIDIR: Modernizing Political Event Data for Big Data Social Science Research	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	SMA	The project creates a general research platform to study civil protests, international conflict, and civil unrest using texts from Spanish, Arabic, and French, in addition to English. This expands the development of programs, data and services available for coding regional conflict and cooperation methods beyond the current English-only approaches to enable data-rich research that will advance new approaches to core questions in the social, behavioral, and economic sciences. The project includes an openly available website that allows for the extraction and reporting of conflict events across the globe as well as the identification of their causes and diffusion. The project's data and methods help make data-driven decisions about foreign policy, civil war prevention, human rights policies, and the effects of other factors such as environmental or economic policies on these phenomena. The project creates large-scale civil and inter-state conflict measures, covering multiple news sources and with a common methodology in an open framework. Using multiple news data sources reduces the biases inherent in coding from a single or small set of news sources, a common approach in the past. The project aims to facilitate the coding of more, and better, data across languages, space and time, thus facilitating the study of substantive questions in traditionally underrepresented countries, peoples, and topics. Further, usability considerations generate new software for the user interface for dealing with big data like that proposed in this research, as well as server-side optimizations that scale large datasets across a diverse set of users. The scale of the event data, covering multiple years and large-scale news databases, will generate many millions of observations over space and time. Research tools, data extraction, and other user interfaces are developed to allow the relevant research communities to have access to, queries of, and citation streams for these data. Finally, machine-coded data from news reports is validated across news sources, languages, actors, ontologies, and against human-coded gold standard records. The research and data serve as inputs for understanding the effects of climate on spatio-temporally referenced civil conflict events in Latin American, Africa, the Middle East, and worldwide.

			DateOperator=&ExpDateOperator=		
176	-	CRII: III: Discovering Complex Change Footprint Patterns on Spatio-Temporal Big Data for Urban Sustainability	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Modern urban systems are facing increasingly significant challenges in sustainable development due to environmental and societal changes such as deforestation, urban sprawl, and rapid population/traffic growth. To respond to the challenges, an essential task is to identify the footprints (i.e., where and when) of these change processes. In the meantime, many spatio-temporal big datasets (STBD) such as fine-grained environmental and climate observations and detailed public transportation records are being made available to the public. Analyzing these data for urban change footprints (CHAF) helps city planners foresee and understand potential sustainability issues. However, such analyses pose significant challenges due to the non-monotonic nature of most change processes, the large cardinality of candidate patterns in STBD, and the non-trivial tradeoff between computational efficiency and pattern quality. This project will investigate automated, efficient, and effective data mining techniques for the discovery of complex CHAF patterns in urban STBD. The research outcomes are expected to enhance the ability of current STBD analytics tools to analyze change-related patterns. The computational framework proposed can be applied to solve a broad range of other problems such as pattern discovery in video and image processing. Also, the research results will be applied to real-world datasets to discover useful urban change patterns to improve the society's understanding of sustainability. Beyond research, this project will facilitate the development of a graduate level spatio-temporal data mining course at the University of Iowa, and contribute to the training of future professionals in spatial computing. The project will also integrate activities to involve undergraduate students and students from underrepresented groups.Existing STBD analytical techniques only focus on detecting relatively simple CHAF patterns (e.g., regularly-shaped, monotonic changes), and typically report only a small number of CHAFs (e.g., the most or top-k likely changes). The research in this project will focus on the discovery of CHAF patterns that are non-monotonic temporally and irregularly-shaped spatially. The proposed techniques will also guarantee the completeness of results, i.e., report all the CHAFs in the data based on a given definition. Specifically, the following ideas will be explored in the project. (1) Designing interest measures of non-monotonic CHAFs that are statistically powerful and computation-friendly. (2) Designing algorithmic building blocks to efficiently evaluate a range of CHAF interest measure functions with similar properties (e.g., algebraic). (3) Designing a generic computational framework for sub-space enumeration that guarantees the completeness of results. In the proposed framework, the three-dimensional sub-spaces and their dominance relationships will be

			DateOperator=&ExpDateOperator=		modeled as a novel sub-cube-based directed acyclic graph (SCB-DAG). Efficient traversal and pruning strategies on the SCB-DAG will be explored to enumerate candidate CHAFs. This research will provide theoretical and experimental evaluations on real data to validate the correctness, completeness and scalability of the proposed ideas. For further information see the project web page: http://www.biz.uiowa.edu/faculty/xzhou/project/NSF_CRII/index.html
177	-	SaTC-EDU: EAGER: Big Data and Security: Educating the Next-Generation Security Analysts	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	DGE	The project develops a new curriculum to educate next-generation security analysis. The new curriculum attempts to integrate two very distinct areas, data analysis and security. In particular, new labs and projects in existing courses for security are redesigned to make students naturally exposed to learn big data processing techniques with real-world examples. For example, "large-scale malware analytics" will be provided in Information Security Lab (CS6265), and similarly "large-scale static and dynamic analysis" will be provided in "Software Analysis and Testing" (CS 6340) at Georgia Tech. To broaden our efforts, all materials and resources (e.g., new course materials, labs and related tools) will be shared as online training modules publicly available so that the educational communities can adopt the data-driven approach in their education. As a result, the number as well as the quality of security analysts can be vastly increased, and hence our government and industries can be much more effective in securing the cyberspace.

			DateOperator=&ExpDateOperator=		
178	-	III: Small: Investigating Spatial Big Data for Next Generation Routing Services	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&Start	IIS	Increasingly, location-aware datasets are of a size, variety, and update rate that exceed the capability of spatial computing technologies. This project addresses the emerging challenges posed by such datasets, which sometimes are also referred to as Spatial Big Data (SBD). SBD examples include trajectories of cell-phones and GPS devices, temporally detailed (TD) road maps, vehicle engine measurements, etc. SBD has the potential to transform society. A recent McKinsey Global Institute report estimates that personal location data could save consumers hundreds of billions of dollars annually by 2020 by helping vehicles avoid congestion via next generation routing services such as eco-routing. Eco-routing may leverage various forms of SBD to compare routes by fuel consumption or greenhouse gas (GHG) emissions rather than total distance or travel-time. To develop next-generation eco-routing services, this project innovates in three areas. Frist, Lagrangian Xgraphs, a novel concept in computer science, is explored at conceptual, logical and physical database levels to model traveler's frame of reference, a major departure from traditional binary relationship (e.g., adjacency) graphs. Second, it probes the concept of route-collections, and scalable algorithms for finding route-collections. For example, to identify a route-collection over all possible start-times of a given time-interval, the project explores a critical time point approach which divides a given time-interval into a set of disjoint sub-intervals of stationary-rankings among alternative routes. The approach is not only novel but also very important for the field. Critical time points may become a vital component of dynamic programming (DP) solutions, which would need reconsideration in the face of emerging temporally detailed SBD that violate DP assumptions about stationary ranking of alternate solutions. Third, to address the increasing diversity of SBD methods, algorithm-ensembles and flexible architectures that allow rapid integration of new data sources and routing algorithms are developed. The proposed work serves national goals for energy independence and sustainability by laying the ground work for eco-routing and other travel-related services that reduce fuel consumption and greenhouse gas emissions. By increasing the availability of SBD, the project also enhances the research infrastructure for other researchers. Educational activities include curriculum development and training of students in the emerging area of SBD and Eco-routing. Result dissemination is planned via publication in relevant peer-reviewed conferences and journals. More details are available on the project website (www.spatial.cs.umn.edu/eco-routing/).

			DateOperator=&ExpDateOperator=		
179	-	REU Site: Big Data Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	This new Research Experiences for Undergraduates (REU) site will train students in a computer science approach to data analytics, ranging from machine learning to data management, and develop research skills from primary research engagement and supplementary activities. Students will: (1) be exposed to the many applications of data analytics across computer science, (2) gain expertise via one specific project, and (3) be equipped with a technical competitive advantage, knowledge, and motivation to pursue graduate study in a STEM field. They will be deeply involved in projects that span a wide range of interests and have social appeal. This REU site will build national economic competitiveness by training and engaging a new generation of students, with a specific focus on women and on students from institutions without much access to research in STEM areas.Big data - the confluence of an ever growing supply of data generated across all fields and disciplines and increasingly powerful computational tools to process and mine these data - is widely acknowledged as one of the biggest opportunities and challenges of the 21st century. This site will give undergraduates a chance to engage in cutting-edge research in one of many computer science fields that utilizes data analysis. Each student will work with a single faculty mentor on a specific project. Students will participate in an intense, hands-on research experience associated with their own summer project, but they will enjoy a cohort community of REU students all working within and learning about data analytics. Students will learn the ways in which the intelligent use of data can solve difficult problems and improve people's quality of life. They will complete their time at the site with the skills to further the use of data analysis applied to the field of their choosing.

180	-	RET in Engineering and Computer Science Site for Machine Learning, Big Data and Computer Science Principles	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	This award establishes a new Research Experiences for Teachers site at Drexel University. The Drexel team plans to create a summer research institute in computer science for STEM high school teachers and 2-year college faculty in the City of Philadelphia and Greater Philadelphia Metropolitan Area. Teachers will spend the time performing research with faculty and students in the associated labs and initiating work on educational modules for use at their home institutions. They will prepare poster presentations and videos summarizing their projects, initial accomplishments, further expectations and descriptions of their modules. Participants will remain engaged year-round by participating in Teacher2Teacher discussion groups, facilitated by the Math Forum@Drexel. They will also visit campus for quarterly meetings, culminating in a 1-day showcase event in the spring to present results, materials and posters, and to which their colleagues and students are invited. The objectives of the program are to build partnerships between high schools, community colleges and the university, introduce teachers to cutting edge research in the computer science community, inform and excite them about computer science principles and computational thinking, produce learning materials for use in high school and community college STEM curricula, and expand the pipeline of students studying STEM and computing curricula in college.Intellectual Merit: The intellectual merit is in the strong research expertise of the participating faculty in research as well as their significant experience with pre-college education. The focus is on important themes of big data and machine learning which are areas that are current and of importance all citizens. The translation of the research experiences into modules compatible with Computer Science Principles should also add to the significant body of work linking fundamental computer science to classroom practice and applications. Broader Impacts: The project makes significant outreach to the greater Philadelphia metropolitan area, with a population exceeding 5 million, and the eighth largest school district in the USA. The Math Forum involves a substantial community of mathematics teachers and professionals, providing a variety of services and resources. Through the partnership with these organizations and dissemination through other widely used services, the team expects to provide significant impact for these activities.

181	-	Situating Big Data: Assessing Game-Based STEM Learning in Context	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DRL	This REAL project arises from the 2013 solicitation on Data-intensive Research to Improve Teaching and Learning. The intention of that effort is to bring together researchers from across disciplines to foster novel, transformative, multidisciplinary approaches to using the data in large education-related data sets to create actionable knowledge for improving STEM teaching and learning environments in the medium term and to revolutionize learning in the longer term. The project team aims to understand how to use data collected from the environment in which learning technologies are used to do the following: (1) allow automated assessment that takes the full range of classroom activities and discussions around use of the technology into account in providing customized feedback recommendations; (2) come to better understand how learning and the context in which it is happening interact; and, (3) provide theory-informed and evidence-based advice for refining learning approaches and activities. This will make it easier for teachers to manage ongoing assessment and to adapt classroom activities to learners' needs in learner-centered, project-based, and inquiry-driven learning environments. Results of this project will lay the foundations for making assessment regular, routine and ongoing and to take a fuller range of learning activities into account. This, in turn, will allow better personalization and ongoing feedback and scaffolding for learners. Results will enhance understanding of how to assess and foster not only disciplinary learning, but also disposition, identity development, and long-term participation.The PIs seek to integrate theories of situated cognition with analysis of big data. They will explore how to integrate clickstream data from technology with key forms of multimodal data describing the contexts in which the technology is being used, e.g., individual and group discourse (online and in-room), individual and curricular artifacts, classroom assessments, and school performance, to generate a data-driven methodology for: (1) understanding the learning happening in technology-rich learning environments; (2) assessing development and needs of individuals within those environments in ways that will suggest adaptations and scaffolding; and (3) investigating situated cognition. They aim to make it easier to manage ongoing assessment and to adapt classroom activities to learners' needs in learner-centered, project-based, and inquiry-driven learning environments. They will investigate how to (1) enable consideration of the full ecosystem of learning and data collected across it when assessing learning and engagement, and (2) identify what is working and not working to foster learning in a situation. They will demonstrate where and how useful data are situated in the learning ecology when learners are engaged in hands-on and discourse-rich learning activities, and how to use these data to assess effectiveness and impact of interventions. Their plan involves matching important patterns in hand-coded qualitative data to patterns of automatically collected data; this will allow them to identify the patterns in automated data collection that can be used as indicators of factors such as understanding, confusion, learning, and participation.

182	-	EAGER: Nonlinear and Data-Adaptive Compressive Sampling for Big Data Processing	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ECCS	As pervasive sensors continuously collect and record massive amounts ofhigh-dimensional data from communication, social, and biological networks,and growing storage as well as processing capacities of modern computershave provided new and powerful ways to dig into such huge quantities ofinformation, the need for novel analytic tools to comb through these "bigdata" becomes imperative. The objective of this project is to develop anovel framework for nonlinear, data-adaptive (de)compression algorithms tolearn the latent structure within large-scale, incomplete or corrupteddatasets for compressing and storing only the essential information, forrunning analytics in real time, inferring missing pieces of a dataset, andfor reconstructing the original data from their compressed renditions. The intellectual merit lies in the exploration of the fertile but largelyunexplored areas of manifold learning, nonlinear dimensionality reduction,and sparsity-aware techniques for compression and recovery of missing andcompromised measurements. Capitalizing on recent advances in machinelearning and signal processing, differential geometry, sparsity, anddictionary learning are envisioned as key enablers. Effort will be put alsointo developing online and distributed (non)linear dimensionality reductionalgorithms to allow for streaming analytics of sequential measurementsusing parallel processors. The broader impact is to contribute to the development of novel computational methods and tools useful for data inference, cleansing, forecasting, and collaborative filtering, with direct impact to statistical signal processing and machine learning applicationsto large-scale data analysis, including communication, social, andbiological networks.
183	-	CAREER: When Energy Harvesting Meets "Big Data": Designing	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	ECCS	In order to build a self-sustainable wireless sensor network, powering sensor nodes with energy harvesting devices becomes a natural and feasible solution, thanks to the recent progress on energy harvesting technology. However, wireless sensor networks usually need to collect and transmit vast amounts of data. Utilizing the random, non-uniform, and scarce harvested energy intelligently to meet the energy demand caused by "big data" is extremely challenging. The real-time latency requirement for data delivery makes the problem even more complicated. This project overcomes these challenges through strategically allocating the stochastic energy resources to collect and

		Smart Energy Harvesting Wireless Sensor Networks	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		transmit the most important sensor data, providing a reliable pipeline for data collection, transmission and analysis in energy harvesting sensor networks. The project is expected to have direct impact on the design and wide deployment of energy harvesting wireless sensor networks, with applications in radio spectrum management, environment monitoring, healthcare, surveillance, disaster relief, etc. Furthermore, the proposed work has widespread potential applications far beyond energy harvesting wireless sensor networks, such as automatic residential energy consumption scheduling in smart grid, micro-grid planning with renewable energy inputs, information extraction from astronomical amounts of data collected from large-scale participatory sensing, etc. The proposed project integrates a research agenda with a strong educational component and will serve the following educational purposes: 1) the PI will use the proposed project to stimulate and maintain undergraduate students' interests in engineering via undergraduate research, senior project design and project based learning; 2) the requested funding will be used to train graduate students via curriculum development, proper research mentoring and industry collaboration; and 3) the proposed project will be used to attract potential engineering students through various outreach activities. The goal of this project is to construct a new paradigm of sensing and transmission schemes in data-intensive energy harvesting wireless sensor networks to intelligently utilize the random, non-uniform, and scarce harvested energy with analytically provable sensing, transmission and inference performance guarantees. Two different but closely coupled approaches are proposed to achieve this goal. One is an energy-driven approach and the other is a data-driven approach. For the energy-driven approach, the statistics of the energy harvesting process are exploited to construct online sensing and transmission schemes. Two main tasks addressed in this research thrust are objective-oriented cooperative sensing scheduling policies to cope with the non-uniform energy supply in large-scale sensor networks, and delay-constrained data transmission schemes to meet the real-time latency requirement. The data-driven approach utilizes the characteristics of the underlying sensing phenomena to adaptively and strategically allocate scarce energy resources for the collection and transmission of the most important sensor data. Two specific tasks in this research thrust include a Gaussian process based framework to systematically utilize the spatial-temporal correlations in sensing fields, and adaptive sensing strategies that exploit the structured sparsity of underlying sensing signals, both under the stochastic energy constraints at sensors. The project promises to build intelligent energy harvesting wireless sensor networks with superb sensing, transmission and inference performances on a solid analytical foundation. The delay-constrained information theoretic analysis for energy harvesting communications will integrate a new set of analytical tools from renewal theory with tools in information theory, and create synergies between them. The proposed data-driven adaptive sensing approach will develop synergies between stochastic queueing control and high-dimensional data analysis. The interdisciplinary nature of the research allows us to utilize techniques from stochastic queueing control, renewal theory, information theory, machine learning and is expected to advance the understanding of those areas.

184	-	SHF: Small: Collaborative Research: Modeling and Analyzing Big Data on Peta- and Exascale Distributed Systems Supported by MapReduce Methodologies	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CCF	Current petascale platforms can perform large-scale simulations and generate massive amounts of data at unprecedented rates. These rates are expected to increase as exascale platforms are introduced. The generation of more and more data presents new challenges for scientists who struggle with the analysis, sorting, and selection of scientifically meaningful results. When very large amounts of data records are located across a large number of nodes in a distributed memory system, even a small number of comparisons can be costly or even impossible. Therefore, new methodologies are necessary to analyze large scientific datasets at scale.The goal of this project is to develop a transformative analysis method to model the properties of large scientific datasets in a distributed manner on petascale systems today and exascale systems in the future. The research activity includes (1) the design of new algorithms for encoding properties embedded in distributed data in a parallel manner by using space reduction techniques; (2) the design of new algorithms for clustering and classifying these properties by using distributed paradigms such as MapReduce; (3) the deployment of the algorithms for diverse datasets in structural biology and astronomy; and (4) the tuning of the algorithms for both result performance and accuracy on emerging storage technologies. The analysis method will provide the scientific community with infrastructures and instrumentations to identify features that can be used to predict class memberships; find recurrent patterns in datasets; and identify class memberships from a specific feature or property. By effectively and accurately capturing scientific information in a scalable manner, these infrastructures and instrumentations will break the traditional constraint of data centralization and allow scientists to overcome the difficulties associated with the fully distributed nature of the data considered.The project's educational component promotes training and learning in computational modeling and analysis techniques as well as data-intensive algorithms and platforms by involving undergraduate and graduate students in research activities and integrating big data analytics into the undergraduate curriculum at the University of Delaware. The research-based educational materials developed in this project will be made available to the scientific community through the project portal and through tutorials at XSEDE and Supercomputing (SC) conferences.

185	-	Big Data in Macro and Labor Economics: New Insights into Open Questions	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	SES	The PI will use "big data" to measure important economic variables in a unique way. The project uses a data set drawn from government administrative records; the data include over 30 years of annual earnings for a sample of millions of American men. With these data, the PI will be able to measure the year-to-year risk of changes in annual earned income. In particular, he will be able to determine whether or not common assumptions used in other economic models (for example, that the risk of an income shock is normally distributed) are correct. This is potentially transformative research; the PI is making innovative use of existing data and his results could overturn the conventional understanding of how business cycles, industry-specific factors, occupation, and skill set determine the risk of large year-to-year changes in income. The results will be broadly useful as well for economic policy. To understand how recessions affect American families, we need to understand how recessions affect earnings. Our current evidence here is based on small samples of the US population, and the data from these samples are not rich enough to answer the full range of questions.

186	-	CAREER: Scaling Forensic Algorithms for Big Data and Adversarial Environments	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Forged digital images or video can threaten reputations or impede criminal justice, due to falsified evidence. Over the past decade, researchers have developed a new class of security techniques known as 'multimedia forensics' to determine the origin and authenticity of multimedia information, such as potentially falsified images or videos. However, the proliferation of smartphones and the rise of social media have led to an overwhelming increase in the volume of multimedia information that must be forensically authenticated. Forger's capabilities have also grown dramatically, as sophisticated editing software allows forgers to perform complex manipulations of digital images and videos. Researchers have recently demonstrated that an adversarial forger can design anti-forensic attacks capable of fooling forensic algorithms. By contrast, little multimedia forensics research has focused on improving the speed at which multimedia forensics techniques operate, particularly on large data sets. This research project is focused on scaling multimedia forensic algorithms to address these new challenges that have arisen due to the evolving technical and social landscape. The research project is focusing on three main aims: (1) Scaling forensic algorithms to meet big data challenges, (2) Scaling forensic algorithms to handle complex forgeries, and (3) Scaling forensics to meet increased adversarial capabilities. To accomplish these aims, the research is drawing from a wide variety of fields such as signal processing, estimation theory, statistical hypothesis testing, machine learning, optimization theory, and game theory.

187	-	CAREER: Semiparametric and Machine Learning Approaches to Big Data Challenges in Precision Medicine	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DMS	In the era of Big Data, the goal of better patient outcomes, coupled with lower cost and burden has generated tremendous interest in precision (or personalized, individualized) medicine, which is defined as treatments targeted to the needs of individual patients on the basis of genetic, biomarker, phenotypic, or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations (Jameson and Longo, 2015).Precision medicine can be operationalized using individual's health-related metrics and environmental factors to discover individualized treatment regimes (ITRs); methodology for such discovery is an emerging field of statistics. The proposed methods are expected to bring a great impact to accelerate the discovery of new personalized treatment strategies. Therefore the proposed work is directly related with the White House Precision Medicine Initiative(https://www.whitehouse.gov/precision-medicine) as a research effort to revolutionize how to improve health and treat disease. The proposed methods are also general enough to be applied to a variety of data sources including clinical, biomarker, economic and financial data. If successful, the projects will greatly enhance the acquisition and analysis of large-scale data for the scientific and engineering communities.The main objective of this proposal is to develop cutting-edge semiparametric methods and machine learning tools to realize the promise of precision medicine. Specifically, the PI aims to: develop flexible and efficient methods for discovering optimal ITRs (Aim 1); develop a general class of optimal ITRs (Aim 2); develop optimal ITRs with high-dimensional data (Aim 3); and develop optimal ITRs under population heterogeneity (Aim 4). The proposed work contributes to both semiparametric inference and machine learning fields. Machine learning methods have rarely been studied for doubly robust estimation and optimal ITRs with high-dimensional data. The theoretical developments including driving nonasymptotic distribution, risk bounds, new empirical process technical tools are challenging. The methodologies to be developed in this project will be fundamentally important and generally applicable for studying semiparametric models in high-dimensional setting. Using semiparametric and machine learning methods for precision medicine is an emerging novel area. The integration of research and education is a key aspect of this project. New courses on statistical learning and semiparametric inference will be developed. These courses will broaden the areas of specialized training in a department that has a strong history of attracting under represented groups. The PI is expecting to stimulate interests from a diverse group of researchers in numerous fields. The PI will also reach out to the K-12 education levels by training high school teachers.

188	-	REU Site: A forest full of Big Data: the Harvard Forest Summer Research Program in Ecology 2015-2019	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DBI	This REU Site award to Harvard University's Harvard Forest is supported by the Division of Biological Infrastructure (DBI) in the Directorate for Biological Sciences (BIO) and the Division of Mathematical Sciences (DMS) in the Directorate for Mathematical and Physical Sciences (MPS). The REU program, located in Petersham, MA, will support the training of 12 students for 11 weeks during the summers of 2015- 2019. Research will focus on collecting, visualizing, analyzing, and communicating ecological "Big Data". Students will be mentored by researchers who are conducting innovative forest ecology research. The program aims to meet five overarching objectives: 1) Enhancing the ability of students to undertake high-quality interdisciplinary research; 2) Building teams of researchers in which students bring different strengths to the table, collaborate on cutting-edge projects, and find their own intellectual "voice"; 3) Encouraging students to link fundamental and applied issues in science; 4) Communicating research findings to a wide range of audiences; and 5) Cultivating the next generation of ecological scientists and educators that reflects the diversity of backgrounds and experiences of students in the United States. The Harvard Forest REU program actively recruits students from a broad spectrum of cultural, economic, educational, and ethnic diversity. The Harvard Forest's commitment to diversity was recognized with the 2014 Human Diversity Award from the Organization for Biological Field Stations. Over the next five years, the program is committed to enhancing diversity and cultivating the next generation of scientists. It is anticipated that a total of 60 students, primarily from schools with limited research opportunities, will be trained in the program. Students will learn how research is conducted, and many will present their results at scientific conferences. Students participating in collaborative groups are fully involved in the development and execution of projects on the cutting edge of ecological research. Students are also encouraged to develop independent projects that reflect their own interests and observations. Students participate in workshops on ethics, scientific writing, and presenting and analyzing data.Students maintain contact with Harvard Forest so that the impact of the Summer Research Program on their career progress can be assessed. Demographic data about students are collected through responses to emails sent via the NSF reporting system. More information is available by visiting http://harvardforest.fas.harvard.edu/other-tags/reu, or by contacting the PI (Dr. Aaron Ellison at aellison@fas.harvard.edu) or the co-PI (Dr. Andrew Richardson at arichardson@oeb.harvard.edu).
189	-	Gradient Methods for Solving Big Data (Tensor) Optimization	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	CMMI	Rapid developments in modern technologies have made large-scale statistical data sets readily available in many industrial settings, but the task of effectively turning data into useful information still remains a major challenge in a wide range of applications. One important source of the 'big data' complication stems from the way in which the data points are collected and stored. In particular, the data format in question is known as the tensor, which is a useful format, because it reflects the interconnections between various factors. Tensor data sets can be found in statistical learning, bioinformatics, consumer behavior in marketing, climate change studies, and signal and image

		Problems	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		processing. However, computationally the tensor data formats are notoriously difficult. This award supports fundamental research on the computational aspects of the above-mentioned information retrieval process. The project has a multidisciplinary research element and will positively impact engineering education.In the context of tensor data processing, a desirable operation is to compute the projection of a given data tensor onto a simpler set with a certain low complexity structure, where the 'low complexity' tensors may refer to low-rank tensors, or sparse tensors, or it may also refer to the tensors with low co-cluster numbers. Formulating tensor projection and completion problems leads to large scale non-convex -- yet algebraic -- optimization models. This research will develop a framework for the iteration complexity analysis which will enable effective first-order computational methods for tensor projection, completion and optimization models.
190	-	MRI: Acquisition of a High Performance Computing Cluster to Support Multidisciplinary Big Data Analysis and Modeling	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&	CNS	This project, acquiring a High Performance Computing (HPC) cluster--the Corpus Christi High Performance cluster (CCHP)--supports large-scale data analysis and modeling research and research training across a broad variety of science and engineering technology disciplines in coastal and environmental studies. The CCHP cluster enables research projects ranging from computer science, life science, geographical information systems (GIS), remote sensing, to atmospheric science. These include:- Higher order tensor decomposition for big data,- Population Genomics of non-model species,- Geospatial crowdsourcing for natural disaster response,- Large-scale analytics of airborne and satellite remote sensing data and UAS imagery,- Global weather and climate analysis.CCHP contains compute nodes, GPU (graphical processing unit) nodes, and shared network storage, connected through InfiniBand switches in a fat tree topology to support high bandwidth low latency data communication (critical for HPC applications) while providing massive parallel computation on graphics processor cores. Moreover, the on-board Gigabit Ethernet ports with switch connection can support large data sets transmission which enables research in real time simulation and modeling. CCHP enables exploring parallel processing to process real large data sets in the

			ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		higher order tensor decomposition, which is a basis for many data mining tasks including clustering, trend detection, and anomaly detection. This cluster is a necessary computational tool for biologists to use statistics effectively to realize the promise and power of societally relevant hypotheses in massively parallel nucleotide sequencing. The GIS researchers can utilize the CCHP cluster to advance geospatial crowdsourcing solution in case of natural disaster by providing quicker and more accurate geometric information. The cluster also enables remote sensing scientists to identify the relationship between environmental conditions, including land cover and use and rates of freshwater inflow, and attack the problems caused by the low accuracy of acquired unmanned aerial systems (UAS) images for precision agriculture. Furthermore, processing much longer time series of the global model and satellite data record, atmospheric scientists will be able to easily expand their research projects from regional to global scale.The impact of the CCHP cluster will be felt in many domains, especially on coastal and environmental studies that impacts society in general, impacting weather and climate model reliability and prediction skills in ecology, evolution, fisheries, conservation, and genetics. Moreover, curriculum materials obtained from the projects serviced by the instrumentation will enhance education and student learning at all levels, including K-12. These will impact existing courses and contribute to the creation of new ones.
191	-	MRI: Acquisition of Peta-Scale Data Storage System for Big Data Exploration in STEM Fields	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleO	ACI	This award is co-funded by EPSCoR program.This MRI project is to acquire a petascale data storage system to add to the existing high-performance computing (HPC) systems at the University of Arkansas at Little Rock (UALR). The projects concern analysis of medical images, cosmological simulations, social network analyses and protein structure prediction. A number of well-developed broadening participation activities in education and research are proposed.The project augments existing high performance computing capabilities, includes projects in different disciplines to be facilitated by cyber upgrades, provides integration of research and education through training opportunities for students at all levels, and has a potential for broadening participation and development of a diverse scientific workforce.

			nly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
192	-	BDD: Efficient and Scalable Collection, Analytics and Processing of Big Data for Disaster Applications	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	CNS	The outcomes from this project is to assist human operators in their disaster management coordination and planning, such as directing a medical physician's team to their nearest cluster of affected people in a region to administer medications as necessary, or finding a safe route for evacuation of affected people. Sensor data integrated with microblogs such as Tweets help identifying some local events and people's sentiments, which are significantly useful in handling/understanding disaster situations. It will also benefit other applications such as real-time tracking of road/driving conditions in vehicular networks.This research is conducted jointly with Osaka University in Japan, to benefit both the universities in enhancing not only their knowledge but also to learn global perspective in solving important problems. The research team is designing schemes for dynamic and collaborative data compression and multi-streams compression of multi-dimensional sensor data with error correction and recovery for addressing the energy efficiency and bandwidth limitation issues. Compression schemes exploit temporal locality and delta compression to provide better bandwidth utilization. Different methods for measuring error are designed and compared for the compressibility and actual error for variations in methods of utilizing the error tolerance. In addition, the team is developing algorithms for highly scalable indexing schemes for efficient data retrieval involving mainly range queries, top-k query, ranked-based searches and snapshot queries for multi-dimensional sensor data from different data sources to address the issue of timely dissemination. Hilbert Curve based linearization technique integrated with an overlay network is designed to (1) map multidimensional attributes onto a single dimension while preserving its data locality, and (2) to create a balanced network by associating only one node with each leaf of the virtual tree and then partition the multidimensional search space into subspaces and assign each node to a unique subspace. This allows an overlay network to start from a predefined prefix to handle data skewness. This research is also designing a scheme for using microblog messages as social sensors for efficient integration with other sensor data. We are using machine-learning techniques to match each message with its associate location based on the characteristics of the message. The results will be validated and evaluated using the sensor cloud test-bed available at Missouri S&T.

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
193	-	I-Corps: Faster than Light Big Data Analytics	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIP	Data-driven decision making is changing society. Political candidates now collate and analyze databases to identify possible campaign supporters. Couples now meet online through matchmaking websites that pair individuals based on profiles and questionnaires. Marketers now fine-tune their messages based on fine-grain market segmentation provided by extensive web and social network signals. Companiesnow develop products based on A/B testing or measured customer behavior. However, the diffusion of the basic software platforms and algorithms for analytics has been uneven. Large firms can afford to hire programmers and data scientists to run sophisticated analytics, but most organizations cannot afford the large capital and operational expenses to maintain analytics infrastructure. This leaves themout of the potentially transformative impact of data-driven decision making. This proposal expands the space of high-performance machine learning by harnessing a more general and more efficient parallel programming platform.The large capital expense stems from the lack of a turnkey solution for predictive analytics. Applications are ad-hoc, developed anew based on specific organizational requirements. The large operational expense arises from the data demand of these applications, which requires many machine-hours to produce results. The team has developed a reusable software platform that can parallelize analytics applications with orders of magnitude performance improvements and can run with a fraction of the hardware resources compared to commonly deployed commercial products. On top of this platform, the team has begun to develop a suite of state-of-the-art massively scalable machine learning algorithms. Through the course of this project, the team intends to broadly engage with potential customers of data analytics to help understand possible business models for the proposed technology.

194	-	III: Small: BigSolver: Data-Intensive Solver Support for Big Data Exploration and Mining	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	Exploratory data analysis plays a key role in data-driven discovery in a wide range of domains including science, engineering and business. This project will enable data scientists from many domains to search and explore their large data sets far easier and faster than they do today. Rather than spending a lot time to set up exploration pipelines by combining multiple software tools, users will work with a single, general purpose and more usable system. Overall, this project will enable fundamentally richer means for data exploration and lead to significant productivity improvements; it will accelerate discovery and breakthroughs in many domains such as e-commerce, finance and science. This research will be incorporated in undergraduate and graduate coursework. The outreach activities include special research and education-focused programs that are geared towards undergraduates and high-school girls. This research proposes to design and build a new prototype database system, called BigSolver, that will uniquely integrate constraint solving and data management techniques. The result will enable rich, highly-efficient means for generic ad hoc search, exploration and mining over large multidimensional data collections. BigSolver will allow Constraint Programming (CP) machinery to run efficiently inside a DBMS without the need to extract, transform and move the data. This marriage will offer the rich expressiveness and efficiency of constraint-based search and optimization provided by modern CP solvers with the ability of Database Management Systems (DBMSs) to store and query data at scale. This work will be an early yet transformative step in enriching the functionality of database systems towards new data- and search-intensive applications. The proposed effort will develop novel approaches for synopsis-based in-memory processing, speculative solving, search query optimization, parallel processing and load balancing, as well as architectural innovations, which will collectively yield performance and usability levels that far improve those of the state of the art. For further information, see the project website at: http://database.cs.brown.edu/projects/SearchLight/

195	-	SI2-SSE: Scalable Big Data Clustering by Random Projection Hashing	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	This project plans to develop a distributed algorithm for secure clustering of high dimensional data sets. Fields in health and biology are significantly benefited by data clustering scalability. Bioinformatic problems such as Micro Array clustering, Protein-Protein interaction clustering, medical resource decision making, medical image processing, and clustering of epidemiological events all serve to benefit from larger dataset sizes. The algorithm under development, called Random Projection Hash or RPHash, utilizes aspects of locality sensitive hashing (LSH) and multi-probe random projection for computational scalability and linear achievable gains from parallel speed. Furthermore, RPHash provides data anonymization through destructive manipulation of the data preventing de-anonymization attacks beyond standard best practices database security methods. RPHash will be deployable on commercially available cloud resources running the Hadoop (MRv2) implementation of MapReduce. The exploitation of general purpose cloud processing solutions allows researchers to scale their processing needs using virtually limitless commercial processing resources.The RPHash algorithm uses various recent techniques in data mining along with a new approach toward achieving algorithmic scalability on distributed systems. The basic intuition of RPHash is to combine multi-probe random projection with discrete space quantization. Regions of high density are then regarded as centroid candidates. To follow common parameterized, k-means methods, the top k regions will be selected. The focus on a randomized, and thus non-deterministic, clustering algorithm is somewhat uncommon in computing, but common for ill-posed, combinatorially restrictive problems such as clustering and partitioning. Despite theoretical results showing that k-means has an exponential worst case complexity, many real world problems tend to fair much better under k-means and other similar algorithms.

196	-	CAREER: Privacy Analytics for Users in a Big Data World	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CNS	Increasing amounts of data are being collected about users, and increasingly sophisticated analytics are being applied to this data for various purposes. Privacy analytics are machine learning and data mining algorithms applied by end-users to their data for the purpose of helping them manage both private information and their self-presentation. This research develops privacy analytics that help users answer three interconnected questions about their online persona (1) What data does the user consider sensitive, and in what contexts should one share it?; (2) What does the data say about the user; and (3) Who knows what? These privacy analytics introduce a novel, inverse data mining problem where users analyze their data to estimate the conclusions the data will produce when incorporated into larger data sets. This project designs new algorithms for quantitative and automated methods to detect privacy-related phenomena that have been observed qualitatively. These algorithms support the development of usable privacy enhancing technologies and will give users tools to cope with and manage their data in a complicated data environment. These tools will provide awareness to users about how their data is being used. These analytics will also help answer questions critical to the development of privacy law and policy.This work involves approximately twenty-five undergraduates in research activities, exposing them to research methods and privacy issues. This project also develops novel educational materials including course offerings for an interdisciplinary master's program in security and educational tools for use by the general public to bridge the digital divide.

197	-	Big Data on Small Organisms: Petascale Simulations of Data-Driven, Whole-Cell Microbial Models	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	ACI	This project aims to develop the next-generation of genome-scale, data-driven models for microbial organisms. The project will first focus on the most-studied microbe, the gram-negative bacterium Escherichia coli, due to the availability of high-throughput data, cellular organization, its significance to industry and human health. The project will take advantage of the multi-omics datasets that resulted from advances in parallel high-throughput molecular profiling over the past fifteen years, the emergence of data-driven, integrative, multi-scale models with substantial improvement of their predictive power and new techniques in machine learning, especially those related to deep learning. Accurate prediction of microbial fitness and cellular state can have profound implications to the way we test hypotheses that are directly related to health, social or economic benefits. This award will support the training of multiple undergraduate and graduate students in computational modeling and high-performance simulations of biological systems through undergraduate courses, IGEM teams and other initiatives.This project will support the generation of knowledge from the largest normalized omics compendia for the most widely used microbe that will be a boon for the development and training of the next generation of data-driven predictive methods in molecular and cellular biology. It will provide the computational resources to evaluate a state-of-the-art multi-scale model with the capacity to predict phenotypic characteristics and environmental conditions from collective omics data. This will be the first systems-level simulator that targets a specific microbe (E. coli) and will be able to simulate populations of cells with a resolution ranging from individual gene concentrations to population dynamics. To achieve that, process migration, load-balancing and strong scaling techniques have to be adopted and applied in this context, which are all novel features for the area of whole-cell modeling. The proposed HPC simulations will be intimately related to hypothesis generation and testing. The simulations will address questions related to what their phenotype and expression profiles of microbial cultures are in complex environments. In the context of systems biology, integration of these techniques has the potential of being transformative, but only if the necessary computational infrastructure able to handle these tasks is available. The Blue Waters Supercomputer with its unique architecture, large-scale simulation capabilities and professional support staff provides the ideal platform to achieve this ambitious goal.

198	-	CIF: Small: Rate Distortion Paradigms for the Big Data Era	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CCF	Data-centric technologies form an essential part of the nation's economy and infrastructure. Coincident with the evolution of these technologies, the amount of information that is being collected, stored and processed is growing at an exponential pace. A sustainable solution for managing this explosion of data should be guided in part by a fundamental understanding of how information can be stored in a succinct manner that meets utility and access requirements. This research seeks to quantify basic tradeoffs associated with information storage, utility and accessibility.Information theory provides a framework for quantifying what is and what is not possible in data compression and transmission. The traditional information-theoretic framework, however, does not readily address critical challenges inherent to modern data processing applications and large data sets. In particular, information-theoretic analyses generally ignore complexity constraints and encounter significant difficulties when distributed settings are considered. Building on recent work by the principal investigator on compression under logarithmic loss and compression for similarity queries, this project will address these two issues by investigating rate-distortion tradeoffs with the added dimensions of distributed processing and complexity constraints. Specific goals of the project include (i) characterizing the compressibility of an information source subject to the constraint that elementary queries are supported in the compressed domain with bounded complexity; and (ii) quantifying the performance of distributed processing (relative to centralized processing) for compression and inference. As a counterpart to the proposed theoretical analyses, the design of practical algorithms that are capable of achieving the established information-theoretic limits will be pursued.

199	-	CAREER: Cyberinfrastructure for Intelligent Water Supply (CIWS): Shrinking Big Data for Sustainable Urban Water	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CBET	To overcome challenges in using smart water metering as an effective tool for sustainably managing urban water supplies, an integrated research and education plan called Cyberinfrastructure for Intelligent Water Supply (CIWS) is being implemented. CIWS will investigate novel cyberinfrastructure and analytics to advance smart metering technology, enable detailed characterization of residential water use behavior, and build the scientific data and knowledge base for sustainably managing urban water supplies. CIWS will enable transitioning of existing conventional "dumb" water meters (the vast majority of water meters in use today) into low-cost (~$100) "smart" and computationally-capable devices that can collect high frequency data, use onboard processing capability to "shrink" the collected data by extracting the timing and volume of individual water end uses, and then transmit actionable data products to water managers for analysis - all without replacing or affecting the functionality of the meter. CIWS will be capable of closing critical gaps in understanding of and ability to quantify water use behavior at the household and water system level. It will also enable identification of alternative water management strategies and opportunities for water conservation and increased efficiency. CIWS and associated residential water use studies offer a way to characterize residential water use and generate new knowledge about: 1) how water use behavior varies across socio-demographic groups and neighborhood types; 2) the timing of water demand and how this information can be used by water providers to ensure water availability and efficiency, plan for related energy demand, and improve customer satisfaction; and 3) how water consumers change their behavior given detailed information about their water use. This information is critical in identifying opportunities for conservation, forecasting demand, and determining how water use patterns may change over time in response to population growth, demographic shifts, and technology improvements. CIWS will advance understanding of water use behavior, the cyberinfrastructure for smart metering, and the pool of "cyber-savvy" professionals and students capable of implementing smart metering, all of which are critical for realizing the promises of smart metering. Mentorship will be provided for a new generation of engineers and scientists who will receive training and engage in research that prepares them to leverage new cyberinfrastructure. Water users will be directly engaged in data collection and information transfer. Hundreds of USU students will engage in a campus "Water Wars" competition aimed at water conservation and sustainability using cyberinfrastructure developed by this project. Logan City and USU Facilities will be integral partners in data collection. Finally, graduate and undergraduate students will be engaged in a visualization challenge using continuous flow data from residential water meters. These efforts will involve participants from underrepresented groups in undergraduate research through targeted recruiting using existing programs at USU.

200	-	EAGER: Productive and Scalable Utilization of Manycores in Big Data Applications	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	CCF	Many important electronics applications, including medical imaging, cyber security and digital entertainment, require processing massive amount of data under limited time, power and price budgets. Modern many-core processor chips offer a promising platform for computing such data-intensive workloads, however, their utilization is hindered due to the inherent difficulty and unproductivity of parallel software development at scale. The project aims to address this critical problem by developing methodologies and tools for productive and scalable development of software that would efficiently run on thousands of processor cores. Central to the approach is development of "functionally-consistent structurally-malleable streaming specification," which enables design-time exploration of a rich software implementation space, under the constraints of hardware platform, while maintaining the application's end-to-end functional behavior. The project results are expected to improve performance, energy consumption, and application development cost in a large number of data-driven disciplines, ranging from medical imaging and cyber security to multimedia and beyond.

201	-	Extracting Knowledge from 100 years of Microstructural Images: Using Machine Vision and Machine Learning to Address the Microstructural Big Data Challenge	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DMR	The Division of Materials Research; the Civil, Mechanical, and Manufacturing Innovation Division; and the Division of Advanced Cyberinfrastructure contribute funds to this award. It supports research and education to collect, analyze, and compare data on materials from vast sources. All solid objects - from an airplane wing to a frying pan - have a microscopic structure that is usually not visible to the naked eye. This structure determines the properties of the material as a whole - whether it is strong or weak, for example. For the past century, materials scientists have studied these structures by using microscopes to take pictures (called micographs) of them. They then measure the important features seen in the microscopic images and relate those measurements to the properties of the material.Just as in personal photography, digital cameras have enabled materials scientists to take more pictures and do more with them than ever before. Moreover, older micrographs have been scanned in to digital archives. Materials scientists are now confronted with a set of images that is too large and too diverse to analyze manually. Fortunately, computer scientists have developed "machine vision" computer programs that identify similarities in large sets of images by in a sense mimicking how humans see objects. This project will gather micrographs from many sources into an open archive and use machine vision programs to search, sort, and classify them automatically without significant human intervention.By synthesizing microscopic image data at a previously impossible scale, this project creates a foundation for discovering new connections between microscopic structures and materials properties. The results will help improve current materials and even develop new ones. The data will be made available to the broader community.The Division of Materials Research; the Civil, Mechanical, and Manufacturing Innovation Division; and the Division of Advanced Cyberinfrastructure contribute funds to this award. It supports research and education to collect, analyze, and compare data on materials from vast sources. Over the past 100 years, materials scientists have made great progress in acquiring, analyzing, and comparing microstructural images. Much of this effort has been directed toward deep understanding of particular materials systems or classes of microstructures. When the catalog of possible microstructural features is known, imaging techniques can take advantage of well-defined feature characteristics to analyze microstructures with high precision. However, when the features of interest are not known a priori, these methods become intractable, inaccurate, or fail completely. Thus, typically, they are applied only to a pre-selected set of micrographs, chosen by a human expert. In contrast, the goal of this effort is to develop a general method to find useful relationships between micrographs without assumptions about what features may be present. Such an approach can leverage the explosion in digital data over the past two decades to survey the breadth of available microstructures efficiently and without significant human intervention.Capitalizing on recent advances in computer science, this project applies a subset of data science concepts - including data harvesting, machine vision, and machine learning - to advance the science of microstructure. The result will be a framework for finding connections between microstructural images within and across material systems, which will support outcomes ranging from computational tools to discovery science, including:- New open source tools for extracting micrographs and associated metadata from various digital archives, including the internet, PDF documents, and local storage media.- A comprehensive database of publicly available micrographs with traditional text-based search and novel image-

					based search functions. - Optimized, high throughput, automatic machine vision techniques to identify microstructural features that are salient to image analysis and microstructural science.- Automatic and objective machine learning systems that find relationships between microstructures in order to discover new structure-property and structure-performance connections.The goal of microstructural science is to understand the connection between microstructural features and materials properties. By developing an open-access, automatic, and objective machine learning system for finding relationships between microstructural images, this project creates a foundation for discovering new connections that may inspire deeper understanding or predictive capabilities.
202	-	CC*DNI Campus Design: Internet2 Infrastructure to Enable Research in Big Data Science and Engineering at Tennessee State University	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&Ori	ACI	This project enhances cyberinfrastructure at Tennessee State University (TSU) with high-speed access to massive and complex datasets and distributed computing infrastructure via Internet2. The two-year project accelerates outcomes of ongoing astronomy, bioinformatics and cybersecurity projects with high speed access to: instruments, such as robotic telescopes and their streaming data; high dimension datasets, such as genomic datasets; and large scale distributed computing and networking platforms. TSU is adding a dedicated network for science and engineering research traffic, upgrading the backbone network connections to the research complex housing the astronomy, bioinformatics and cybersecurity laboratories, and providing high speed access to Internet2. The enhanced infrastructure will enable TSU faculty and students to effectively engage in science and engineering research and education. The new campus network capacity enables TSU to transfer large datasets and improves quality of computational research conducted by science and engineering faculty. Specifically, researchers can connect to TSU's state-of-the art telescopes located in Southern Arizona and perform daily analysis of images of several thousand stars. Bioinformatic researchers can conduct genomic sequencing service facilities to analyze massive and complex datasets of plant tissues at Pittsburgh Supercomputing Center, Vanderbilt Technologies for Advanced Genomic, Cornell's Genomic Diversity Facility and national distributed computing platforms. Cybersecurity researchers can conduct at scale research experiments and take advantage of high speed access to computing servers and Openflow switches on the Global Environment for Network Innovations (GENI) facility. The project also advances undergraduate research, allowing students to conduct data analysis with effective use of computational resources.

			ginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		
203	-	Measuring the Ages of Stars in the Era of Big Data Time-Domain Astronomy	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	AST	James Davenport is awarded an NSF Astronomy and Astrophysics Postdoctoral Fellowship to carry out a program of research and education at Western Washington University. A key mission in astronomy is to understand the ages and histories of stars within our galaxy, yet their ages remain one of the most difficult properties to measure. When stars are young they spin rapidly and exhibit strong magnetic activity, including frequent high energy flares that can potentially impact life on nearby planets. Stars lose angular momentum over time, slowing their rotation, which should decrease this magnetic activity and generate fewer flares. This change in flare rate can be used as an approximate clock and is crucial for understanding how stars and planets evolve over time. To calibrate this clock, Davenport will use the ultra-precise data from NASA's planet-hunting Kepler mission as a training sample, finding every flare from over 200,000 stars. This work requires analyzing large volumes of data using modern computational techniques. Davenport will use this Kepler data to train students at smaller institutions core principles of data science and visualization, which are critical for work in technical fields.Using stellar magnetic activity as a clock has long been suggested as a possible means for determining stellar ages. Furthermore, planet habitability may be impacted by the evolution of the host star's magnetic activity over time. Interest in low-mass stars as exoplanet hosts has fueled a growing need for better constraints on stellar ages. With the advent of continuous monitoring from space-based missions like Kepler, we are finally able to calibrate this method. Davenport will develop the first measurement of this activity-age relationship by studying the flare rates of stars from the Kepler mission. Over 200,000 stars will be studied for flares using machine learning and time series statistical techniques. This endeavor will involve using modern data science techniques to mine large databases. The demand for such skills in the private and public sectors is growing rapidly, and Davenport will use large, dynamic, and open access astronomical datasets to teach data science methods. Gaining experience in data science for STEM majors can be challenging at smaller institutions, and data from the Kepler mission are ideal for teaching these skills. The course will also serve as a template for data science STEM training in other departments at small institutions.

204	-	Using Big Data and Visual Analytics to Investigate the Long-term, Cascading Effects of Informal STEM Learning	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	DRL	In concert with the overall strategy of the Advancing Informal STEM Learning (AISL) program to advance new approaches to, and evidence-based understanding of, the design and development of STEM learning in informal environments, Principal Investigators from Oregon State University, University of Idaho, and University of Texas at Dallas, will study a range of data in online social networks to identify evidence of the long-term impact of informal STEM education. Tracking informal learners over time to understand the impact of informal learning experiences has been a longstanding, daunting, and elusive challenge. Now, with massive amounts of data being shared and stored online, education researchers have an unprecedented opportunity to study such data and apply data analytics and visualization technologies to identify the long-term, cascading effects of informal STEM learning. Research findings will inform the design and development of a data-analysis tool for use by education practitioners to improve STEM learning experiences online, through television and film, and at informal education institutions. An independent external critical review board of learning scientists, computer scientists, engineers, informal STEM education practitioners, participating partners, broadcast media professionals, and policymakers, will ensure a robust evaluation of the research and effectiveness and utility of the data analysis tool to improve practice. A summary report for the field will be written on the scientific and practical reliability and validity of the research and data-analysis model, and the value of the work for audiences beyond informal STEM education practitioners and policymakers. The research is contemporaneously relevant, advancing innovative use of data-mining and data-analysis processes to better understand how informal learners communicate STEM learning experiences and interact with STEM content over time, across a range of social networks. Investigators will research: 1) whether learners who engage in informal STEM education experiences further their learning through discussions and sharing of information in social media networks, 2) which types of data are present in social media that are relevant for understanding the cascading impacts of learning over time, and 3) how learning may evolve independently within shared social networks, which, if discovered, could provide a predictive computational model with implications for significant impact across both formal and informal education. Investigators will employ existing and modified data crawlers to search for key terms and phrases, assess spikes and deformations in posts, queries, and blogs, and experiment with their test data to find which types or configurations of keywords or search terms deliver the most reliable and accurate results. A variety of formats will be explored to test various strategies with participating partners and practitioners. Data will be visualized to represent the following dimensions of learning: a) Interest/Affect, b) Recommendations, c) Understanding/Knowledge-Seeking, and d) Deeper Engagement.

205	-	CHS: Small: Supporting Crowdsourced Sensemaking in Big Data with Dynamic Context Slices	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	IIS	This research will investigate how crowdsourcing and computational techniques can be combined to support the efforts of an individual analyst engaged in a complex sensemaking task, such as identifying a threat to national security or determining the names of people and places in a photograph. Currently, such complex tasks are beyond the capabilities of the most advanced machine learning techniques or crowdsourcing workflows, and even trained experts struggle to perform them. Huge quantities of data are now available online, but making sense of them is challenging because human cognition, while remarkably powerful, is nevertheless a limited resource. Visual analytics tools seek to overcome this limitation by leveraging the complementary strengths of information visualization and data mining, but these tools generally assist with low-level tasks, requiring significant effort on the part of users. Crowdsourcing has emerged as a promising technique for applying human intelligence to problems computers cannot easily solve, but for crowds to assist individuals with complex sensemaking tasks, two significant challenges must be addressed. First, we must understand when crowds versus computation are more useful at each phase in the sensemaking loop. Second, we must overcome the limited time and expertise of most crowd workers to sustain deep, complex lines of inquiry.This research addresses both of these challenges through a series of four experiments. First, it will conduct a laboratory study where individuals perform complex sensemaking tasks to understand what types and amounts of context they use to make decisions, and how the sensemaking loop might be decomposed into subtasks. Second, it will conduct a series of experiments comparing crowdsourcing to automated techniques for each of the most promising sensemaking subtasks. Third, it will experiment with different crowd workflows to develop a revised sensemaking loop, optimized for the relative strengths of crowds and computation, and develop a software prototype based on this approach. At the core of the software design is the novel concept of "context slices," an innovative technique for addressing the transience of crowd workers by giving them only the information they need to complete their assigned task, allowing complex investigations to be pursued across multiple workers. The fourth experiment will evaluate this approach by comparing performance with the software to the baselines established in the first study.
206	-	CC-NIE Network Infrastructure: Advancing Network Capacity,	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrg	ACI	Powerful networks are the thread that holds together the computing and data resources that underpin modern scientific discovery. With the ever-increasing demands on bandwidth to support the analysis of the ever-growing volumes of scientific data, this project will transition the University of Wisconsin - Madison (UW-Madison) research community to a new era of networking technologies and capabilities.Software Defined Network (SDN) technologies will be employed to establish the UW-Madison science DMZ that will support up to 100Gb/sec data transfer rate to universities and research institutions across the US through Internet2. Integral to this new campus wide resource

		Efficiency, and Security for Wisconsin Big Data Research Through Improvement of campus research DMZ	anization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		will be the ability to use emerging networking protocols like OpenFlow to dynamically provision communication bandwidth between departmental resources and the campus DMZ. A new campus wide team of both central IT and research staff will facilitate the deployment and operation of these new capabilities. This Advanced Network Team (ANT) will be responsible for the integration of the new networking and software capabilities with the UW CyberInfrastructure. An important element of the ANT activity will be to monitor networking activities on the campus and guide research and development of new distributed computing technologies.Broader impacts: This will benefit all campus scholarly activities including other NSF sponsored projects, activities funded by other federal agencies, and other campus scholarship including the social sciences, humanities, and arts. The expertise gained will be disseminated throughout the UW System, the Committee for Institutional Cooperation, and the WiscNet regional network, which serves higher education, K-12, and public libraries.
207	-	Discovering Collaboration Network Structures and Dynamics in Big Data	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&	SMA	Understanding how individual scientists interact with one another and how such interaction impacts research productivity and knowledge diffusion is important for understanding the dynamics of scientific research collaboration. At the same time, information about patterns of collaboration and their consequences have implications for science policy. In quantitative research on collaboration networks, publication co-authorships and citation-linkages have been the primary source of data. As large data repositories, one of the signposts for cyberinfrastructure-enabled, data-driven science, become increasingly prevalent, however, they offer an alternative source of information about networks of scientific collaboration. This project investigates research collaboration networks emerging around one such international data repository, GenBank, and develops data products to support data-driven science policymaking and research. By utilizing this novel data source the project provides an unprecedented opportunity to validate and expand the theory of complex networks while generating rich data outputs and products to support science policy research and policymaking. This study fills a number of theoretical and methodological gaps identified by the 2008 roadmap for Science of Science Policy (SoSP), with a specific focus

			ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		on how scientific collaboration networks form and evolve. The outcomes of this study address the lack of models and tools for network analysis, visual analytics, and science mapping outlined in the 2008 roadmap for SoSP. To accomplish the data collection and processing required for this project new computational programs will be developed to parse, extract, store, transform, split, merge, and filter the data; these will be applicable to the analysis of other similar data sources for science policy and innovation research. Broader impacts. By making available dataset product prototypes the project will allow researchers, policy makers, and students to explore research networks in GenBank from longitudinal, thematic, geographical, institutional, and author dimensions. The multi-dimensional, interactive presentations of such datasets enable data-intensive science policy research and support science policymaking through filtering, sorting, associating, and visualization capabilities. The datasets and data products will be made available through an open access mechanism, so educators and undergraduate and graduate students have ample opportunities to use these resources for teaching and research. Students enrolled in Syracuse University's newly established Certificate for Advanced Study in Data Science (CAS DS) program will be able to participate in the project and gain skills in programming for data collection and processing, data quality verification, analysis, and visualization. In addition, the collaboration network analysis provides interested doctoral students an opportunity to do independent study or dissertation research. Findings from studying cyberinfrastructure-supported data sharing and knowledge diffusion is expected to advance policymakers' ability to properly assess the outcomes of federally funded research.
208	-	THE CROWD & THE CLOUD: Citizen Science, Big Data and the Democratization of Research	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleO	DRL	This media and research project will develop and study the use of new media, broadcast television, and social networks to introduce Citizen Science to a national audience, and motivate their direct involvement and participation. Project deliverables will include: four nationally-distributed public TV programs hosted by Waleed Abdalati, Director of CIREs at the University of Boulder and former NASA Chief Scientist; online videos for training and outreach of citizen science partners; digital engagement via social media; and a custom-designed application ('2nd screen app') that enables users to obtain additional informational content, share information, and connect with other viewers. The evaluation and research study will build new knowledge on how these deliverables can motivate the public to become citizen science participants.The investigators estimate the four television programs will reach approximately 80% of U.S. television households. In addition, videos and other content will be distributed through channels such as iTunes, Hulu, Netflix, and social media. Target audiences will include the general public, citizen science activists, and professional scientists. Underrepresented groups will be reached through special Google Hangouts, and professional societies such as SACNAS and AGU. The research components of the project will provide evidence on how traditional researchers respond to citizen science, and explore the deliverables' use as recruitment tools for citizen science projects and impacts on viewers' attitudes, behaviors, and skills related to citizen science. Data will be collected from multiple sources, including online surveys, in-person focus groups, and analyses of users' online postings. Retrospective surveys will be administered to explore changes in behavior regarding whether respondents have increased their interaction with professional scientists, or participated in citizen science initiatives. A quasi-experimental study will be conducted to assess the value added by the 2nd screen

			nly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=		app.

209	-	Big data analyses reveal social, emotional, and cognitive mechanisms in communication	http://www.nsf.gov/awardsearch/advancedSearchResult?PIId=&PIFirstName=&PILastName=&PIOrganization=&PIState=&PIZip=&PICountry=&ProgOrganization=&ProgEleCode=&BooleanElement=All&ProgRefCode=&BooleanRef=All&Program=&ProgOfficer=&Keyword=%22Big+Data%22&AwardTitleOnly=true&AwardNumberOperator=&AwardAmount=&AwardInstrument=&ActiveAwards=true&OriginalAwardDateOperator=&StartDateOperator=&ExpDateOperator=	BCS	Usage patterns in large-scale datasets, such as linguistic corpora collected from digital social media, can provide insights into many aspects of communication. These resources have inspired new quantitative analyses and measures, some of which originated with computer scientists, mathematicians, and physicists and others with psychologists and linguists. The technology itself is also being transformed in order to communicate the social and emotional aspects of human behavior. Innovative lines of inquiry attest to the mutual influence between technology and behavior and the transformation that ensues as the relevant disciplines come into contact.This award will bring together a panel of psychologists, linguists and computer scientists at the 2015 meeting of the American Association for the Advancement of Science in San Jose, CA. The theme of the meeting is "Innovations, Information, and Imaging." The panel will describe research and analytic techniques for developing new insights into how individuals convey and understand emotions in dialog, speech, and text. More broadly, the panel will discuss how linguistic and non-linguistic coordination arise across individuals and groups as they communicate using new technologies.
210	-	Network Query by Example	https://github.com/Aptima/pattern-matching.git	DARPA	Hadoop MapReduce-over-Hive based implementation of network query by example utilizing attributed network pattern matching. (Java)

211	-	SMILE-WIDE: A scalable Bayesian network library	https://github.com/SmileWide/main.git	DARPA	SMILE-WIDE is a scalable Bayesian network library. Initially, it is a version of the SMILE library, as in SMILE With Integrated Distributed Execution. The general approach has been to provide an API similar to the existing API SMILE developers use to build "local," single-threaded applications. However, we provide "vectorized" operations that hide a Hadoop-distributed implementation. Apart from invoking a few idioms like generic Hadoop command line argument parsing, these appear to the developer as if they were executed locally. (Java)
212	-	Active Search	https://github.com/AutonlabCMU/ActiveSearch.git	DARPA	ActiveSearch takes a collection of emails (or any dataset where a similarity can be generated between elements) and recommends related messages based on user feedback. The user provides an initial seed email then enters into a cycle where ActiveSearch provides a similar email and the user reports whether or not the email was interesting.ActiveSearch is useful for anyone navigating a large set of emails and looking for related messages on a specific topic. As it considers the similarities between emails as well as user feedback, it is an improvement in accuracy, time, and effort over basic text search or a brute force search.(Java, Perl)
213	-	skl-groups	https://github.com/dougalsutherland/skl-groups.git	DARPA	A package extending the Python machine learning toolkit scikit-learn with support for operating on sets ("groups") as features.
214	-	Blaze	https://github.com/ContinuumIO/blaze.git	DARPA	Blaze is the next-generation of NumPy. It is designed as a foundational set of abstractions on which to build out-of-core and distributed algorithms over a wide variety of data sources and to extend the structure of NumPy itself. Blaze allows easy composition of low level computation kernels (C, Fortran, Numba) to form complex data transformations on large datasets. In Blaze, computations are described in a high-level language (Python) but executed on a low-level runtime (outside of Python), enabling the easy mapping of high-level expertise to data without sacrificing low-level performance. Blaze aims to bring Python and NumPy into the massively-multicore arena, allowing it to leverage many CPU and GPU cores across computers, virtual machines and cloud services. (Python)
215	-	Bokeh	https://github.com/ContinuumIO/bokeh.git	DARPA	Bokeh (pronounced bo-Kay or bo-Kuh) is a Python interactive visualization library for large datasets that natively uses the latest web technologies. Its goal is to provide elegant, concise construction of novel graphics in the style of Protovis/D3, while delivering high-performance interactivity over large data to thin clients. (Python/JavaScript/Coffeescript)

216	-	Numba	https://github.com/numba/numba.git	DARPA	Numba is an Open Source NumPy-aware optimizing compiler for Python sponsored by Continuum Analytics, Inc. It uses the LLVM compiler infrastructure to compile Python syntax to machine code. It is aware of NumPy arrays as typed memory regions and so can speed-up code using NumPy arrays. Other, less well-typed code is translated to Python C-API calls effectively removing the "interpreter" but not removing the dynamic indirection. Numba is also not a tracing just in time (JIT) compiler. It compiles your code before it runs either using run-time type information or type information you provide in the decorator.
217	-	Circuit	https://github.com/gocircuit/circuit.git	DARPA	Circuit reduces the human development and sustenance costs of complex massively-scaled systems nearly to the level of their single-process counterparts. It is a combination of proven ideas from the Erlang ecosystem of distributed embedded devices and Go's ecosystem of Internet application development. Circuit extends the reach of Go's linguistic environment to multi-host/multi-process applications. (Go)
218	-	Darpa Open Catalog Generator	https://github.com/ericwhyne/open-catalog-generator.git	DARPA	Code and templates for building the DARPA open catalog as hosted on darpa.mil.
219	-	Escher	https://github.com/gocircuit/escher.git	DARPA	Escher is a minimal metaphor programming language that plays as a lego block for intelligent translation across foreign semantics of heterogenous technologies. (Go)
220	-	Vowpal Wabbit	https://github.com/JohnLangford/vowpal_wabbit.git	DARPA	The Vowpal Wabbit (VW) project is a fast out-of-core learning system sponsored by Microsoft Research and (previously) Yahoo! Research. Support is available through the mailing list. There are two ways to have a fast learning algorithm: (a) start with a slow algorithm and speed it up, or (b) build an intrinsically fast learning algorithm. This project is about approach (b), and it's reached a state where it may be useful to others as a platform for research and experimentation. There are several optimization algorithms available with the baseline being sparse gradient descent (GD) on a loss function (several are available). The code should be easily usable. Its only external dependence is on the boost library, which is often installed by default. (C)
221	-	USER-ALE: User Activity Logging Engine	https://github.com/draperlaboratory/user-ale.git	DARPA	Analytic Activity Logger is an API that creates a common message passing interface to allow heterogeneous software components to communicate with an activity logging engine. Recording a user's analytic activities enables estimation of operational context and workflow. Combined with psychophysiology sensing, analytic activity logging further enables estimation of the user's arousal, cognitive load, and engagement with the tool. (JavaScript)

222	-	SmallK: A high-performance library for nonnegative matrix factorization and hierarchical clustering	https://github.com/smallk/smallk.github.io.git	DARPA	SmallK is a high-performance, parallel library for nonnegative matrix factorization on both dense and sparse matrices written in C++. Implementations of several different NMF algorithms are provided, including multiplicative updating, hierarchical alternating least squares, nonnegative least squares with block principal pivoting, and a new rank2 algorithm. The library provides an implementation of hierarchical clustering based on the rank2 NMF algorithm. (C++)
223	-	Markov modulated Poisson process for event detection	https://github.com/giantoak/MMPP.git	DARPA	Markov Modulated Poisson Process for Event Detection allows R users to accurately detect unusual events and anomalies in time series of counts. (R)
224	-	LineUp	https://github.com/Caleydo/org.caleydo.vis.lineup.demos.git	DARPA	LineUp is a novel and scalable visualization technique that uses bar charts. This interactive technique supports the ranking of items based on multiple heterogeneous attributes with different scales and semantics. It enables users to interactively combine attributes and flexibly refine parameters to explore the effect of changes in the attribute combination. This process can be employed to derive actionable insights as to which attributes of an item need to be modified in order for its rank to change. Additionally, through integration of slope graphs, LineUp can also be used to compare multiple alternative rankings on the same set of items, for example, over time or across different attribute combinations. We evaluate the effectiveness of the proposed multi-attribute visualization technique in a qualitative study. The study shows that users are able to successfully solve complex ranking tasks in a short period of time. (Java)
225	-	LineUp Web	https://github.com/Caleydo/lineup.js.git	DARPA	LineUpWeb is the web version of the novel and scalable visualization technique. This interactive technique supports the ranking of items based on multiple heterogeneous attributes with different scales and semantics. It enables users to interactively combine attributes and flexibly refine parameters to explore the effect of changes in the attribute combination.

226	-	UpSet	https://github.com/VCG/upset.git	DARPA	Understanding relationships between sets is an important analysis task that has received widespread attention in the visualization community. The major challenge in this context is the combinatorial explosion of the number of set intersections if the number of sets exceeds a trivial threshold. To address this, we introduce UpSet, a novel visualization technique for the quantitative analysis of sets, their intersections, and aggregates of intersections. UpSet is focused on creating task-driven aggregates, communicating the size and properties of aggregates and intersections, and a duality between the visualization of the elements in a dataset and their set membership. UpSet visualizes set intersections in a matrix layout and introduces aggregates based on groupings and queries. The matrix layout enables the effective representation of associated data, such as the number of elements in the aggregates and intersections, as well as additional summary statistics derived from subset or element attributes. Sorting according to various measures enables a task-driven analysis of relevant intersections and aggregates. The elements represented in the sets and their associated attributes are visualized in a separate view. Queries based on containment in specific intersections, aggregates or driven by attribute filters are propagated between both views. UpSet also introduces several advanced visual encodings and interaction methods to overcome the problems of varying scales and to address scalability. (JavaScript)
227	-	SKYLARK: Randomized Numerical Linear Algebra and ML	https://github.com/xdata-skylark/libskylark.git	DARPA	SKYLARK implements Numerical Linear Algebra (NLA) kernels based on sketching for distributed computing platforms. Sketching reduces dimensionality through randomization, and includes Johnson-Lindenstrauss random projection (JL); a faster version of JL based on fast transform techniques; sparse techniques that can be applied in time proportional to the number of nonzero matrix entries; and methods for approximating kernel functions and Gram matrices arising in nonlinear statistical modeling problems. We have a library of such sketching techniques, built using MPI in C++ and callable from Python, and are applying the library to regression, low-rank approximation, and kernel-based machine learning tasks, among other problems. (C++/Python)
228	-	Immersive Body-Based Interactions	http://code.google.com/p/svnmimir/source/checkout	DARPA	Provides innovative interaction techniques to address human-computer interaction challenges posed by Big Data. Examples include: * Wiggle Interaction Technique: user-induced motion to speed visual * Tablet-based visualization controller: eye-free rapid interaction with visualizations. * Multi-touch interfaces: browsing/querying multi-attribute and geospatial data, hosted by SOLR. * Immersive Tablet Based Viewers: low-cost 3D virtual reality fly-throughs of data sets.
229	-	Apache Gora	https://git-wip-us.apache.org/repos/asf/gora.git	DARPA	The Apache Gora open source framework provides an in-memory data model and persistence for big data. Gora supports persisting to column stores, key value stores, document stores and RDBMSs, and analyzing the data with extensive Apache Hadoop MapReduce support.

230	-	Apache Nutch	http://svn.apache.org/repos/asf/nutch/	DARPA	A large scale web crawler framework that implements the search engine architecture as originally defined by Brin and Page. Nutch was started by Doug Cutting and is the predecessor to the Apache Hadoop technology. It includes parsers, a protocol framework, language detection, indexing capabilities and language and query models to fully implement a search engine.
231	-	Apache OODT	https://svn.apache.org/repos/asf/oodt/	DARPA	APACHE OODT enables transparent access to distributed resources, data discovery and query optimization, and distributed processing and virtual archives. OODT provides software architecture that enables models for information representation, solutions to knowledge capture problems, unification of technology, data, and metadata. (Java)
232	-	Apache Tika	http://svn.apache.org/repos/asf/tika/trunk/	DARPA	The Apache Tika(TM) toolkit detects and extracts metadata and structured text content from various documents using existing parser libraries.
233	-	DRAT	https://github.com/chrismattmann/drat.git	DARPA	A distributed, parallelized (Map Reduce) wrapper around Apache RAT to allow it to complete on large code repositories of multiple file types where Apache RAT hangs forever.
234	-	Khooshe	https://github.com/MBoustani/Khooshe.git	DARPA	A Big Data-Points Visualization Tool
235	-	igraph	https://github.com/igraph/xdata-igraph.git	DARPA	igraph is a collection of network analysis tools with the emphasis on efficiency, portability and ease of use. igraph can be programmed in GNU R, Python and C/C++.
236	-	Tangelo	https://github.com/Kitware/tangelo.git	DARPA	Tangelo provides a flexible HTML5 web server architecture that cleanly separates your web applications (pure JavaScript, HTML, and CSS) and web services (pure Python). This software is bundled with some great tools to get you started. (JavaScript/Python)
237	-	Information Extractor	https://github.com/mitll/MITIE.git	DARPA	Trainable named entity extractor (NER) and relation extractor. (C)
238	-	Julia	https://github.com/JuliaLang/julia.git	DARPA	Julia is a high-level, high-performance dynamic programming language for technical computing, with syntax that is familiar to users of other technical computing environments. It provides a sophisticated compiler, distributed parallel execution, numerical accuracy, and an extensive mathematical function library.

239	-	Query By Example (Graph QuBE)	https://github.com/mitll/graph-qube.git	DARPA	Graph QuBE is a tool which enables efficient pattern-of-behavior search in data containing entities transacting over time.
240	-	SciDB	http://scidb.org/forum	DARPA	Scientific Database for large-scale numerical data. Register on the forum to access the download page. (C++)
241	-	Topic	https://github.com/mitll/topic-clustering.git	DARPA	This tool takes a set of text documents, filters by a given language, and then produces documents clustered by topic. The method used is Probabilistic Latent Semantic Analysis (PLSA). (Python)
242	-	VizLinc	https://github.com/mitll/vizlinc.git	DARPA	Vizlinc is a visual analytics platform that takes as input a corpus of text documents, extracts named entities (people, locations, and organizations) and the relations between those entities from the documents, and allows a user to explore the information contained in the documents from both a high-level corpus view point and with respect to more narrow queries.(Java/Groovy)
243	-	Neon Visualization Environment	https://github.com/NextCenturyCorporation/neon.git	DARPA	Neon is a framework that gives a datastore agnostic way for visualizations to query data and perform simple operations on that data such as filtering, aggregation, and transforms. It is divided into two parts, neon-server and neon-client. Neon-server provides a set of RESTful web services to select a datastore and perform queries and other operations on the data. Neon-client is a JavaScript API that provides a way to easily integrate neon-server capabilities into a visualization, and also aids in 'widgetizing' a visualization, allowing it to be integrated into a common OWF based ecosystem. (Groovy/JavaScript)
244	-	Ozone Widget Framework	https://github.com/ozoneplatform/owf.git	DARPA	Ozone Widget Framework provides a customizable open-source web application that assembles the tools you need to accomplish any task and enables those tools to communicate with each other. It is a technology-agnostic composition framework for data and visualizations in a common browser-based display and interaction environment that lowers the barrier to entry for the development of big data visualizations and enables efficient exploration of large data sets. (JavaScript)
245	-	bigalgebra	https://r-forge.r-project.org/scm/viewvc.php/?root=bigmemory	DARPA	Bigalgebra is an R package that provides arithmetic functions for R matrix and big.matrix objects. (R)
246	-	bigmemory	http://cran.r-project.org/web/packages/bigme	DARPA	Bigmemory is an R package to create, store, access, and manipulate massive matrices. Matrices are allocated to shared memory and may use memory-mapped files. Packages biganalytics, bigtabulate, synchronicity, and

			mory/index.html		bigalgebra provide advanced functionality. (R)
247	-	flexmem	https://github.com/kaneplusplus/flexmem.git	DARPA	Flexmem is a general, transparent tool for out-of-core (OOC) computing in the R programming environment. It is launched as a command line utility, taking an application as an argument. All memory allocations larger than a specified threshold are memory-mapped to a binary file. When data are not needed, they are stored on disk. It is both process- and thread-safe. (C)
248	-	laputa	https://github.com/kaneplusplus/laputa.git	DARPA	Laputa is a Python package that provides an elastic, parallel computing foundation for the stat_agg (statistical aggregates) package. (Python)
249	-	stat_agg	https://github.com/kaneplusplus/stat_agg.git	DARPA	stat_agg is a Python package that provides statistical aggregators that maximize ensemble prediction accuracy by weighting individual learners in an optimal way. When used with the laputa package, learners may be distributed across a cluster of machines. The package also provides fault-tolerance when one or more learners becomes unavailable. (Python)
250	-	Geofin	https://github.com/phronesis-mnemosyne/geofin-open-catalog	DARPA	Geofin allows exploration of patterns in data at global, national, local, and human scales. Quickly triage locations and entities to find patterns and trends using customizable summary statistics and visualizations. (JavaScript, Elasticsearch)
251	-	Tessera	https://github.com/tesseradata/tesseradata.github.io.git	DARPA	Tessera is an open source environment for deep analysis of big data. At the front end, the analyst programs in R, and has access to the thousands of methods of statistics, machine learning, and visualization implemented in R. At the back end is a distributed parallel computational environment such as Hadoop, that enables scaling to big data. In between are three Tessera packages: datadr, Trelliscope, and RHIPE (R and Hadoop Integrated Programming Environment). These packages enable the data analyst to communicate with the back end by simple R commands, and not have to worry about the details of distributed parallel computation. Tessera is powered by a statistical approach to large complex data, Divide and Recombine (D&R). The data are parallelized, not the thousands of methods, which makes back end computation typically very nearly embarrassingly parallel, and therefore very fast. But at the same time, D&R statistical division and recombination methods ensure good statistical performance.
252	-	Data Microscopes	https://github.com/datamicroscopes/	DARPA	Data Microscopes is a collection of robust, validated Bayesian nonparametric models for discovering structure in data. Models for tabular, relational, text, and time-series data can accommodate multiple data types, including categorical, real-valued, binary, and spatial data. Inference and visualization of results respects the underlying uncertainty in the data, allowing domain experts to feel confident in the quality of the answers they receive.

					(Python, C++)
253	-	Content and Context-based Graph Analysis: NILS, Network Inference of Link Strength	https://github.com/plamenbbn/XDATA.git	DARPA	Network Inference of Link Strength will take any text corpus as input and quantify the strength of connections between any pair of entities. Link strength probabilities are computed via shortest path. (Python)
254	-	Content and Context-based Graph Analysis: PINT, Patterns in Near-Real Time	https://github.com/plamenbbn/XDATA.git	DARPA	Patterns in Near-Real Time will take any corpus as input and quantify the strength of the query match to a SME-based process model, represent process model as a Directed Acyclic Graph (DAG), and then search and score potential matches. (Python)
255	-	A vertex-centric CUDA/C++ API for large graph analytics on GPUs using the Gather-Apply-Scatter abstraction	https://github.com/RoyalCaliber/vertexAPI2.git	DARPA	Allows users to express graph algorithms as a series of Gather-Apply-Scatter (GAS) steps similar to GraphLab. Runs these vertex programs using a single or multiple GPUs - demonstrates a large speedup over GraphLab. (C++)
256	-	BayesDB	https://github.com/mit-probabilistic-computing-	DARPA	BayesDB is an open-source implementation of a predictive database table. It provides predictive extensions to SQL that enable users to query the implications of their data -- predict missing entries, identify predictive relationships between columns, and examine synthetic populations -- based on a Bayesian machine learning system in the back end. (Python)

			project/BayesDB.git		
257	-	Aggregate Micro Paths	https://github.com/Sotera/aggregate-micro-paths.git	DARPA	An analytic to help infer movement patterns from large amounts of geospatial-temporal data in a cloud environment. (Python, Scala)
258	-	Correlation Approximation	https://github.com/Sotera/correlation-approximation	DARPA	Spark implementation of the Google Correlate algorithm to quickly find highly correlated vectors in huge datasets. (Scala)
259	-	Distributed Graph Analytics	https://github.com/Sotera/distributed-graph-analytics	DARPA	Distributed Graph Analytics (DGA) is a compendium of graph analytics written for Bulk-Synchronous-Parallel (BSP) processing frameworks such as Giraph and GraphX. The analytics included are High Betweenness Set Extraction, Weakly Connected Components, Page Rank, Leaf Compression, and Louvain Modularity. (Scala, Java)
260	-	GEQE	https://github.com/Sotera/GEQE	DARPA	Geo Event Quey by Example - Leverage geo-located temporal text data in order to identify similar locations or events. (JavaScript, Python)
261	-	Graphene	https://github.com/Sotera/graphene	DARPA	Graphene is a web-based application that provides combined query, visualization, link identification and analysis, and other analytic capabilities within a single system. It allows a user the ability to intelligently search structured data from multiple data sources and can display transactional views, transactional graphs evolving over time, related-entity (shared-attribute network) graphs, drillable transaction histograms, directional transaction charts, activity plots, data export, and more. The combination of these capabilities makes it a valuable tool for analyzing any kind of data that can be manipulated to reveal transactions and relationships of any sort. (Java, JavaScript)
262	-	Newman	https://github.com/Sotera/newman	DARPA	Newman is a tool to quickly analyze and explore email using advanced analytics and visualization techniques - things not possible with traditional email applications. (JavaScript, Python)
263	-	RHIPE ARIMA	https://github.com/Sotera/rhipe-arima	DARPA	This implementation of the ARIMA (AutoRegressive Integrated Moving Average) algorithm is based on R, Hadoop, and the RHIPE (Hree-pay) framework. (R)

264	-	Track Communities	https://github.com/Sotera/track-communities.git	DARPA	An analytic for creating networks from geo-temporal track data based on time/space co-occurrence. Includes UI for visualization of communities and tracks. This tool is a synthesis of several analytic components and visualization techniques that allow a user to browse a network of communities, follow tracks of movement, and observe co-location highlights within a dynamic graph. (Python, JavaScript)
265	-	Zephyr	http://github.com/Sotera/zephyr	DARPA	Zephyr is a big data, platform agnostic Extract-Transform-Load (ETL) API, with Hadoop MapReduce, Storm, and other big data bindings. (Java)
266	-	CVXPY	https://github.com/cvxgrp/cvxpy.git	DARPA	CVXPY is a Python-embedded modeling language for convex optimization problems. It allows you to express your problem in a natural way that follows the math, rather than in the restrictive standard form required by solvers. (Python)
267	-	ECOS: An SOCP Solver for Embedded Systems	https://github.com/ifa-ethz/ecos.git	DARPA	ECOS is a lightweight primal-dual homogeneous interior-point solver for SOCPs, for use in embedded systems as well as a base solver for use in large scale distributed solvers. It is described in the paper at http://www.stanford.edu/~boyd/papers/ecos.html. (C)
268	-	LowRankModels.jl	https://github.com/madeleineudell/LowRankModels.jl.git	DARPA	LowRankModels.jl is a julia package for modeling and fitting generalized low rank models (GLRMs). GLRMs model a data array by a low rank matrix, and include many well known models in data analysis, such as principal components analysis (PCA), matrix completion, robust PCA, nonnegative matrix factorization, k-means, and many more. LowRankModels fits GLRMs using an alternating directions proximal gradient algorithm. GLRMs and algorithms for fitting GLRMs are described in detail in an associated paper, at http://www.stanford.edu/~boyd/papers/glrm.html (which also links to the code). (Julia)
269	-	SCS (Self-dual Cone Solver)	https://github.com/cvxgrp/scs.git	DARPA	Implementation of a solver for general cone programs, including linear, second-order, semidefinite and exponential cones, based on an operator splitting method applied to a self-dual homogeneous embedding. The method and software supports both direct factorization, with factorization caching, and an indirect method, that requires only the operator associated with the problem data and its adjoint. The implementation includes interfaces to CVX, CVXPY, matlab, as well as test routines. This code is described in detail in an associated paper, at http://www.stanford.edu/~boyd/papers/pdos.html (which also links to the code). (C)
270	-	imMens	https://github.com/uwdata/imMens.git	DARPA	imMens is a web-based system for interactive visualization of large databases. imMens uses binned aggregation to produce summary visualizations that avoid the shortcomings of standard sampling-based approaches. Through data decomposition methods (to limit data transfer) and GPU computation via WebGL (for parallel query processing), imMens enables real-time (50fps) visual querying of billion+ element databases. (JavaScript)

271	-	Riposte	https://github.com/jtalbot/riposte.git	DARPA	Riposte is a fast interpreter and JIT for R. The Riposte VM has two cooperative subVMs for R scripting (like Java) and for R vector computation (like APL). Our scripting code has been 2-4x faster in Riposte than in R's recent bytecode interpreter. Vector-heavy code is 5-10x faster. Speeding up R can greatly increase the analyst's efficiency. (C/R)
272	-	DeepDive	https://github.com/HazyResearch/deepdive.git	DARPA	DeepDive is a new type of knowledge base construction system that enables developers to analyze data on a deeper level than ever before. Many applications have been built using DeepDive to extract data from millions of documents, Web pages, PDFs, tables, and figures. DeepDive is a trained system, which means that it uses machine-learning techniques to incorporate domain-specific knowledge and user feedback to improve the quality of its analysis. DeepDive can deal with noisy and imprecise data by producing calibrated probabilities for every assertion it makes. DeepDive offers a scalable, high-performance learning engine. (SQL, Python, C++)
273	-	Delite	https://github.com/stanford-ppl/hyperdsl	DARPA	Delite is a compiler framework and runtime for parallel embedded domain-specific languages (DSLs). (Scala)
274	-	SNAP	https://github.com/snap-stanford/snap	DARPA	Stanford Network Analysis Platform (SNAP) is a general purpose network analysis and graph mining library. It is written in C++ and easily scales to massive networks with hundreds of millions of nodes, and billions of edges. It efficiently manipulates large graphs, calculates structural properties, generates regular and random graphs, and supports attributes on nodes and edges. (C++)
275	-	Snap.py	https://github.com/snap-stanford/snap	DARPA	Snap.py is a Python interface for SNAP. SNAP is a general purpose, high performance system for analysis and manipulation of large networks.
276	-	Lyra	https://github.com/uwdata/lyra.git	DARPA	Lyra is an interactive environment that makes custom visualization design accessible to a broader audience. With Lyra, designers map data to the properties of graphical marks to author expressive visualization designs without writing code. Marks can be moved, rotated and resized using handles; relatively positioned using connectors; and parameterized by data fields using property drop zones. Lyra also provides a data pipeline interface for iterative, visual specification of data transformations and layout algorithms. Visualizations created with Lyra are represented as specifications in Vega, a declarative visualization grammar that enables sharing and reuse. (JavaScript)
277	-	bigdata	https://bigdata.svn.sourceforge.net/svnroot/bigdata/	DARPA	Bigdata enables massively parallel graph processing on GPUs and many core CPUs. The approach is based on the decomposition of a graph algorithm as a vertex program. The initial implementation supports an API based on the GraphLab 2.1 Gather Apply Scatter (GAS) API. Execution is available on GPUs, Intel Xenon Phi (aka MIC), and multi-core GPUs.

278	-	MapGraph	http://svn.code.sf.net/p/mpgraph/code/	DARPA	MapGraph enables massively parallel graph processing on GPUs and many core CPUs. The approach is based on the decomposition of a graph algorithm as a vertex program. The initial implementation supports an API based on the GraphLab 2.1 Gather Apply Scatter (GAS) API. Execution is available on GPUs, Intel Xenon Phi (aka MIC), and multi-core GPUs.
279	-	Visualization Widgets	https://github.com/piim/xdata-visualization-widgets.git	DARPA	These visualizations were created to demonstrate the type of standalone visualization widgets that might compliment a composite dashboard display for a decision-maker. They are built using D3 and leverage relevant APIs to show the latest available data. (JavaScript)
280	-	Vega	https://github.com/trifacta/vega.git	DARPA	Vega is a visualization grammar, a declarative format for creating and saving visualization designs. With Vega you can describe data visualizations in a JSON format, and generate interactive views using either HTML5 Canvas or SVG. (JavaScript)
281	-	Aperture Tile-Based Visual Analytics	https://github.com/unchartedsoftware/aperture-tiles.git	DARPA	New tools for raw data characterization of 'big data' are required to suggest initial hypotheses for testing. The widespread use and adoption of web-based maps has provided a familiar set of interactions for exploring abstract large data spaces. Building on these techniques, we developed tile based visual analytics that provide browser-based interactive visualization of billions of data points. (JavaScript/Java)
282	-	ApertureJS	https://github.com/unchartedsoftware/aperturejs.git	DARPA	ApertureJS is an open, adaptable and extensible JavaScript visualization framework with supporting REST services, designed to produce visualizations for analysts and decision makers in any common web browser. Aperture utilizes a novel layer based approach to visualization assembly, and a data mapping API that simplifies the process of adaptable transformation of data and analytic results into visual forms and properties. Aperture vizlets can be easily embedded with full interoperability in frameworks such as the Ozone Widget Framework (OWF). (JavaScript/Java)
283	-	Influent	https://github.com/unchartedsoftware/influent.git	DARPA	Influent is an HTML5 tool for visually and interactively following transaction flow, rapidly revealing actors and behaviors of potential concern that might otherwise go unnoticed. Summary visualization of transactional patterns and actor characteristics, interactive link expansion and dynamic entity clustering enable Influent to operate effectively at scale with big data sources in any modern web browser. Influent has been used to explore data sets with millions of entities and hundreds of millions of transactions.
284	-	Uncharted Ensemble Clustering	https://github.com/unchartedsoftware/ensemble-clustering.git	DARPA	Uncharted Ensemble Clustering is a flexible multi-threaded clustering library for rapidly constructing tailored clustering solutions that leverage the different semantic aspects of heterogeneous data. The library can be used on a single machine using multi-threading or distributed computing using Spark. (Java)

285	-	BDAS		DARPA	BDAS, the Berkeley Data Analytics Stack, is an open source software stack that integrates software components being built by the AMPLab to make sense of Big Data.
286	-	BlinkDB	https://github.com/sameeragarwal/blinkdb.git	DARPA	BlinkDB is a massively parallel, approximate query engine for running interactive SQL queries on large volumes of data. It allows users to trade-off query accuracy for response time, enabling interactive queries over massive data by running queries on data samples and presenting results annotated with meaningful error bars. To achieve this, BlinkDB uses two key ideas: (1) An adaptive optimization framework that builds and maintains a set of multi-dimensional samples from original data over time, and (2) A dynamic sample selection strategy that selects an appropriately sized sample based on a query's accuracy and/or response time requirements. We have evaluated BlinkDB on the well-known TPC-H benchmarks, a real-world analytic workload derived from Conviva Inc. and are in the process of deploying it at Facebook Inc. (Scala)
287	-	Mesos	https://git-wip-us.apache.org/repos/asf/mesos.git	DARPA	Apache Mesos is a cluster manager that provides efficient resource isolation and sharing across distributed applications, or frameworks. It can run Hadoop, MPI, Hypertable, Spark, and other applications on a dynamically shared pool of nodes. (C++/Java/Python)
288	-	Shark	https://github.com/amplab/shark.git	DARPA	Shark is a large-scale data warehouse system for Spark that is designed to be compatible with Apache Hive. It can execute Hive QL queries up to 100 times faster than Hive without any modification to the existing data or queries. Shark supports Hive's query language, metastore, serialization formats, and user-defined functions, providing seamless integration with existing Hive deployments and a familiar, more powerful option for new ones. (Scala)
289	-	Spark	https://github.com/apache/spark.git	DARPA	Apache Spark is an open source cluster computing system that aims to make data analytics both fast to run and fast to write. To run programs faster, Spark offers a general execution model that can optimize arbitrary operator graphs, and supports in-memory computing, which lets it query data faster than disk-based engines like Hadoop. To make programming faster, Spark provides clean, concise APIs in Python, Scala and Java. You can also use Spark interactively from the Scala and Python shells to rapidly query big datasets. (Java/Scala/Python)
290	-	Tachyon	https://github.com/amplab/tachyon.git	DARPA	Tachyon is a fault tolerant distributed file system enabling reliable file sharing at memory-speed across cluster frameworks, such as Spark and MapReduce. It achieves high performance by leveraging lineage information and using memory aggressively. Tachyon caches working set files in memory, and enables different jobs/queries and frameworks to access cached files at memory speed. Thus, Tachyon avoids going to disk to load datasets that are frequently read. (Java)
291	-	Gunrock	https://github.com/gunrock/gunr	DARPA	Gunrock is a CUDA library for graph primitives that refactors, integrates, and generalizes best-of-class GPU implementations of breadth-first search, connected components, and betweenness centrality into a unified code

			ock.git		base useful for future development of high-performance GPU graph primitives. (CUDA C/C++)
292	-	GoFFish	https://github.com/usc-cloud/goffish.git	DARPA	The GoFFish project offers a distributed framework for storing timeseries graphs and composing graph analytics. It takes a clean-slate approach that leverages best practices and patterns from scalable data analytics such as Hadoop, HDFS, Hive, and Giraph, but with an emphasis on performing native analytics on graph (rather than tuple) data structures. This offers an more intuitive storage, access and programming model for graph datasets while also ensuring performance optimized for efficient analysis over large graphs (millions-billions of vertices) and many instances of them (thousands-millions of graph instances). (Slash/Java)
293	-	Parallel High Betweenness Nodes identification	https://github.com/usc-cloud/parallel-high-betweenness-centrality.git	DARPA	Fast MPI based high betweenness centrality identification algorithm extendible to cloud graph processing platforms such as Giraph++ or GoFFish.
294	-	Parallel Louvain Community	https://github.com/usc-cloud/parallel-louvain-modularity.git	DARPA	Fast MPI based parallel Louvain community detection algorithm easily mappable to Map Reduce.
295	-	Wings	https://github.com/IKCAP/wings.git	DARPA	WINGS is a semantic workflow system that can be used to automate data analysis processes represented as workflows of computations. A unique feature of WINGS is that its workflow representations incorporate semantic constraints about datasets and workflow components, and are used to create and validate workflows and to generate metadata for new data products. WINGS submits workflows to scalable execution frameworks such as Apache OODT to run workflows at large scale in distributed resources.
296	-	Scaling Personalized Healthcare with Big Data	http://www.nd.edu/~nchawla/papers/bdah_1.pdf	DARPA	Today the healthcare industry is undergoing one of the most important and challenging transitions to date, the move from paper to electronic healthcare records. While the healthcare industry has generally been an incrementally advancing field, this change has the potential to be revolutionarily. Using the data collected from these electronic records exciting tools such as disease recommendation systems have been created to deliver personalized models of an individual’s health profile. However despite their early success, tools such as these will soon encounter a significant problem. The amount of healthcare encounter data collected is increasing drastically, and the computational time for these applications will soon reach a point at which these systems can no longer function in a practical timeframe for clinical use. This paper will begin by analyzing the performance limitations of

					the personalized disease prediction engine CARE (Collaborative Assessment and Recommendation Engine). Next it will detail the creation and performance of a new single patient implementation of the algorithm. Finally this work will demonstrate a novel parallel implementation of the CARE algorithm, and demonstrate the performance benefits on big patient data.

14 APPENDIX C – CPS/IOT Projects

No	Acronym / No.	Project title	Implemented in (EU or US)	Topic/application areas	Funding Body	Related programme, if any	Website	Abstract	Funding (in €, $ converted as $1=€0.90)
1	Flex4Apps	Platform for Application and Infrastructure Flexibility in Cyber-Physical Systems	EU	CPS	EUREKA /ITEA		No website	The convergence of cloud, communication and IoT infrastructure plus the trend towards virtual applications (e.g. migrating software to the cloud), creates new challenges for application developers and infrastructure providers. The resulting systems are complex with dynamic resources hiding possible problems. This creates a requirement for flexible monitoring and optimization methods. The Flex4Apps project addresses the challenges of monitoring and optimizing large, distributed, cyber-physical systems. The goal of the project is to provide a solution to manage the high data volumes and complexity of system monitoring whilst disturbing the target system as little as possible. Market domains addressed by Flex4Apps include cloud computing, telecommunication infrastructure and IoT. Reducing costs in these global, hundreds of billions of Euros, markets would result in significant savings. In addition the targeted markets, such as telecommunication networks, are important enablers for other sectors of business. Major outcomes expected from Flex4Apps are: - Scalable Platform for	14,097,000.00

								monitoring applications and cyber-physical systems to detect and characterize “problems”, embedding cloud technology efficiency and open APIs, - Disruptive Methods and Tools for new in kind data analysis, offline weak point assessment, adaptive performance optimisation and visualisation, and - Real-world deployment of Flex4Apps in four industrial use case scenarios.	
2	OPENCPS	Open Cyber-Physical System Model-Driven Certified Developm ent	EU	CPS	EUREKA /ITEA		[http://https //www.ope ncps.eu](http://https//www.opencps.eu)	Cyber-physical systems put increasing demands on reliability, usability and flexibility while, at the same time, lead time and cost efficiency are essential for industry competitiveness. Tools and environments for model-based development of cyber-physical systems are becoming increasingly complex and critical for the industry: tool interoperability, vendor lock-ins, and tool life-cycle support are some of the challenges. The project focuses on interoperability between the standards Modelica/UML/FMI, improved execution speed of (co-)simulation, and certified code generation.	11,477,000.00
3	M2MGrids	Smart M2M Grids – M2M Internet for dynamic M2M Informati on Business ecosyste m	EU	CPS, IoT	EUREKA /ITEA		No website	The Smart M2M grids project is focused on creating enablers for a dynamic cyber-physical information business ecosystem connecting the physical world with the business processes of companies in real-time. The first goal is to connect physical world sensors, actuators and various embedded devices and machines (physical M2M objects) with IT systems automatically/ semi-automatically by applying and extending horizontal open standards based M2M infrastructures for communication and services. The second goal is to enable information management for embedded and distributed application for smart interaction with physical M2M objects and IT back-office systems. The third goal is enabling smart information exchange between selected business cases related to energy,	20,708,000.00

								buildings, transportation and consumer M2M products and services to make the future world smart, smooth and secure for consumers/prosumers. The resulting system is aimed at boosting transfer towards a more sustainable society and a novel real-time service economy within selected industrial business cases.	
4	SITAC	Social Internet of things: Apps by and for the Crowd	EU	IoT	EUREKA /ITEA		No website	The Internet is expected to ultimately interconnect billions of people and trillions of devices. Over recent years, ‘Web-of-Objects’ and ‘Internet-of-Things’ initiatives have emerged, primarily aimed at machine-to-machine or device-to-device interactions using standard communications protocols. As yet, their rate of adoption in commercial products and services is fairly low. The challenge undertaken by the SITAC project is to create a unifying architecture and ecosystem – comprising platforms, tools and methodologies – which enables seamless connection and co-operation of many types of network-connected entities, whether systems, machines, devices or humans equipped with handheld devices. The project will innovate through the use of the ‘social networking’ paradigm to facilitate and unify interactions both between people and devices and between devices. It will propose a distributed framework for enabling the web-based service representation of smart spaces and objects they include.	15,788,000.00
5	RTIPA	Real-Time Internet Platform Architectures	EU	IoT	EUREKA /ITEA		http://www.hitech-projects.com/euprojects/rtipa	Needs, architecture, protocols and services for the Internet of the future, via an innovative network architecture for real-time streaming of multimedia data	42,167,000.00

6	sCorPiuS	European Roadmap for Cyber-Physical Systems in Manufacturing	EU	CPS	H2020	FoF	http://www.scorpius-project.eu/	sCorPiuS is a Support Actions aiming at fostering the European leadership in the field of Cyber--Physical Systems for the Factory of the Future, by identifying and prioritizing the most relevant research and innovation challenges in engineering and manufacturing as well as by building consensus through the involvement of all the stakeholders along the whole value chain. In particular the sCorPiuS objectives are: to carry out an extensive analysis of the current framework and state--of--the art of CPS technologies, applications in industry, supplier and users as well the most relevant RD&I initiatives at global level; to increase the understanding of how CPS technologies can improve engineering and manufacturing performance; to analyze current development and needs in order to identify possible gaps and areas that need to be further researched; to identify technological challenges and drivers in line with the needs of the vision for the future of engineering and manufacturing; to develop, with the support of academic and industrial experts, a reference roadmap addressing the advances of CPSs (in terms of technology and implementation) that could also support the definition of further research streams in this area; to increase knowledge and awareness (i.e. consensus building) about the use and impacts of CPSs in engineering and manufacturing by fostering a cross-sectorial dialogue involving potential users, experts and technology providers.	610,000.00

7	CP-SETIS	Towards Cyber-Physical Systems Engineering Tools Interoperability Standardisation	EU	CPS	H2020	CPS Cluster	http://cp-setis.eu/	"CPS require multiple engineering competences across various engineering disciplines. The development of such systems is a huge challenge, also because of the heterogeneity of engineering tools involved in development platforms across the development lifecycle. In order to overcome this challenge, past and on-going EU research projects have developed the basis for an International Open Standard for Development Tool Interoperability, the so called Interoperability Specification (IOS). However, due to lack of coordination, current IOS related activities, especially with respect to its standardization and possible extensions, are un-coordinated, endangering the huge financial effort that has been put into the IOS and the chance to establish it as a formal Standard. The main goal of CP-SETIS is to conceive and set up a sustainable organizational structure as a platform joining all stakeholders, to coordinate all IOS related activities, especially the formal standardization and further extensions of the IOS. This organizational structure will be implemented within existing, sustainable European organizations, like the ARTEMIS-IA Working Groups or similar, that are open to all stakeholders. CP-SETIS will ensure the support of all stakeholders for this structure, its operational rules, its implementation within existing structures and, most important, their commitment to coordinate all IOS related activities within this structure. In this way, CP-SETIS will secure the huge effort – both in terms of manpower as well as in terms of financial support – that has been put into the IOS, furthering and enabling the setup of the IOS as a formal standard, and enabling the enormous innovation potential of the IOS both, for innovations in interoperable tools with new	700,000.00

								functionalities needed for future generations of Cyber-Physical Systems, and for innovations in future CPS themselves. CP-SETIS will also use lessons learned during this process to update Standardization Research Agendas."	
8	EuroCPS	European Network of competencies and platforms for Enabling SME from any sector building Innovative CPS products to sustain demand for European manufacturing	EU	CPS	H2020	SmartAnythingEverywhere	https://www.eurocps.org	SMEs play a key role in European economies; they constitute the largest business block and provide the bulk of employment. They generate most of the innovative ideas for ICT and CPS-enabled IoT products, two areas which represent an inflection point for innovators and industry in Europe. EuroCPS is an ambitious project that aims to arm Europe with a network of design centres that will boost and initiate synergies between SMEs, major CPS-platforms, and CPS-competency providers, to capture the emerging markets for IoT products. EuroCPS will 1. Leverage the existing regional ecosystem across the full value chain (from micro-electronics, smart systems, and CPS, to high value added products and services) and range of expertise and competencies to provide innovators from any sector with an easy path to build innovative CPS-enabled systems. 2. Act as a one-stop-shop that provides a critical mass of technologies and competencies by facilitating user-supplier partnerships across value-chains and regions. Hence the typical development time of innovative for CPS applications will be significantly decreased through the ease of access to the platforms, and coaching by the competence partners within EuroCPS. 3. Link software, system and nano-electronic industries along the full CPS value chain to demonstrate a new cooperation model evidenced by experiments initiated and led by innovators that translate the rich pool of	8,186,834.00

								ideas from end users into implementation of CPS applications made in Europe. EuroCPS gathers major European system suppliers and world class research centres and technology providers, all rooted in the top European regional ecosystems. Based on strong foundations in European and national initiatives, EuroCPS will, through pan European collaboration and knowledge exchange and access to the strong value chain in this strategic sector, significantly reduce development time and certification efforts, thus putting Europe at the cutting edge of CPS development and implementation.	
9	U-Test	Testing Cyber-Physical Systems under Uncertainty: Systematic, Extensible, and Configurable Model-based and Search-based Testing Methodologies	EU	CPS	H2020	CPS Cluster	www.u-test.eu	"Uncertainty is intrinsic in Cyber-Physical Systems (CPSs) due to novel interactions of embedded systems, networking equipment, cloud infrastructures, and humans. CPSs have become predominant in critical domains and necessitate the implementation of proper mechanisms to deal with uncertainty during their operation at an acceptable cost avoiding unwarranted threats to its users and environment. One way to guarantee the correct implementation of such mechanisms is via automated and systematic Model-Based Testing (MBT)—a way of improving dependability. U-Test will improve the dependability of CPSs by defining extensible MBT frameworks supporting holistic MBT of CPSs under uncertainty in a cost-effective manner. More specifically our objectives are: 1) Provide a comprehensive and extensible taxonomy of uncertainties classifying uncertainties, their properties, and relationships; 2) An Uncertainty Modelling Framework (UMF) to support modelling uncertainties at various levels relying on exiting modelling/testing standards; 3) Defining an intelligent way to evolve	

								uncertainty models developed using UMF towards realistic unknown uncertainty models using search algorithms (e.g., Genetic Algorithms); 4) Generating cost-effective test cases from uncertainty and evolved models. U-Test consortium encompasses domain experts from various facets of CPSs, i.e., software, embedded systems, distributed systems, and cloud infrastructure. We have chosen two case studies from diverse domains including Handling Systems and Geo Sports to assess the cost-effectiveness of U-Test. The solutions will be integrated into two key commercial tools available in the market: ModelBus/Fokus!MBT and CertifyIt. Moreover, the solutions will be deployed into the actual practise in addition to standardization to achieve a wider impact within Logistics, Geo Sports, and Healthcare domains and further facilitate interoperability among tools and technologies."	
10	TAPPS	Trusted Apps for open CPS	EU	CPS	H2020	CPS Cluster	www.tapps-project.eu	Open and smart cyber-physical systems (CPS) are considered to be the next revolution in ICT with enormous economic potential enabling novel business models for integrated services and products. In many areas of CPS devices, there is a strong trend towards open systems, which can be extended during operation instantly adding functionalities on demand. The main goal of the TAPPS (Trusted Apps for open CPS) project is the development of a platform for CPS Apps, which can also access and modify device internals. Therefore, the solution will address all necessary layers from hardware over software to an app store concept always ensuring security and full real-time support for the applications. The extensibility and the pervasive trusted environment of TAPPS are important differentiators that will enable new market extensions to keep pace with user expectations and latest technology.	3,885,484.00

								As current, rich execution platforms for apps are limited in security, the project will develop a parallel, real-time Trusted Execution Environment (TEE) for highly-trusted CPS Apps. The TEE is located separately from existing the execution environment inside the System Control Units and exploits functionalities provided by the novel hardware-, processor- and network-centric security mechanisms as well as a hypervisor for virtualization. Furthermore, TAPPS will provide and validate an end-to-end solution for development and deployment of trusted apps, including an App Store and a model-based tool chain for trusted application development including verification tools. This multi-level trusted Apps platform and tool chain are matured and validated in health and automotive application domains using industrial, realistic use cases paving the way for future exploitation in further demanding application domains.	
11	CPSELabs	CPS Engineering Labs - expediting and accelerating the realization of cyber-physical systems	EU	CPS	H2020	SmartAnythingEverywhere	www.cpse-labs.eu	"Smart cyber-physical systems (CPS) are considered to be the next revolution in ICT with lots of game-changing business potential for integrated services and products. Mastering the engineering of complex and trustworthy CPS is key to implementing CPS-based business models. Current CPS, however, are often engineered and maintained at very high cost and sometimes with unknown risks, and recent technological progress from R&D projects is not readily available to most innovators. The CPS Engineering Labs (CPSE Labs) therefore equips innovators - businesses, researchers, and students - with CPS engineering infrastructure, knowledge, and tools for realizing novel CPS-based products and services, with the explicit goal of expediting and accelerating the realization of smart CPS. The CPSE Labs build upon existing R&D centres - in Madrid, Munich, Oldenburg, Newcastle, Stockholm, and Toulouse - and turn these already excellent regional	7,437,655.00

								clusters into world-class hotspots for CPS engineering. The design centers develop and maintain a common strategic innovation agenda for building up novel and complete CPS value chains. Based on this strategy the CPSE Labs build up and maintain a portfolio of added-value experiments. Experiments are focused and fast-track and they have a clear innovation objective; they build upon results and achievements from large-scale national and European projects on the rigorous design of embedded systems and CPS. Experience gained from experiments, validation results, and best practices, cross-cutting engineering principles that underpin the integration of cyber and physical elements of CPS are continuously integrated and disseminated by the CPSE Labs. The CPSE Labs' marketplace provides an open forum for sharing platforms, architectures, and software tools for the engineering of dependable and trustworthy CPS. The ultimate goal is to establish a CPS engineering framework which sets a world-wide standard."	
12	CPS-SUMMIT	Transatlantic CPS Summit	EU	CPS	H2020	CPS Cluster	http://cps-vo.org/group/cps-summit	"Cyber-physical systems (CPS) are a core enabling technology for securing economic leadership in embedded systems and ICT, having an enormous social and economic importance, and making decisive contributions to societal challenges. The EU and the US face common challenges to push forward the limits of the science for engineering CPS, creating a favorable environment for strategic and pre-competitive collaboration. The Transatlantic CPS Summit is an ambitious 18-month support action with the goal of facilitating and creating an enduring and sustainable collaboration campaign on CPS research and development between Europe and the US. The support action achieves its overall aim by means of:	181,250.00

								1. Identifying and evaluating possible R&D cooperations between Europe and the US; 2. Investigating and promoting implementation of opportunities for cooperation; 3. Preparing a roadmap for R&D cooperation on CPS engineering between the EU and US together with recommendations for action; 4. Presenting final results to interested stakeholders (e.g. public bodies, industry, academic researchers) on both sides of the Atlantic. To achieve the above, the project mobilises an outstanding multidisciplinary consortium of 7 EU partners and 5 US partners and brings together recognized CPS researchers across the EU and the US in a series of CPS Summit Workshops."	
13	Road2CPS	Strategic action for future CPS through roadmaps, impact multiplication and constituency building	EU	CPS	H2020	CPS Cluster	www.road2cps.eu	"The miniaturisation of sensing, actuating, and computing components together with the increasing number of interacting systems in strongly connected environments, and the growing complexity of such systems have triggered a paradigm shift. CPS concepts address challenges for system implementation such as increasing complexity and flexibility. These challenges and the need to optimise performance and comply with essential requirements like safety and security raise many questions that are partially addressed by current research in areas such as transport, health, production, smart grids and smart cities already. Nevertheless, there is still a huge gap between theoretical concepts, technical developments, and successful application, as well as considerable differences with regard to propagation and maturity of CPS between application domains and along the value chain. Strategic action is needed to bring the relevant stakeholders together and to facilitate mutually beneficial collaborations between	832,893.50

								them. Road2CPS has been conceived to respond to this situation by • Analysing impact from past and ongoing projects, identifying gaps and bridging efforts towards impact multiplication • Developing technology, application and innovation strategy roadmaps for CPS to serve as a catalyst for early adoption of CPS technologies • Enhancing CPS implementation and exposing exploitation opportunities via case studies • Developing recommendations for future research priorities and implementation strategies • Building a CPS constituency – bringing together key players into targeted task forces to contribute to the Road2CPS action plan This will provide European organisations with the direction required to establish their future visions of CPS environments, supporting their efforts to stay at the forefront of new developments and preparing them for future challenges in the industrial application of CPSs; thus reinforcing the leading position of the European industry in CPS."	
14	COSSIM	A Novel, Comprehensible, Ultra-Fast, Security-Aware CPS Simulator	EU	CPS	H2020	CPS Cluster	www.cossim.org	One of the main problems the CPS designers face is “the lack of simulation tools and models for system design and analysis”. This is mainly because the majority of the existing simulation tools for complex CPS handle efficiently only parts of a system while they mainly focus on the performance. Moreover, they require extreme amounts of processing resources and computation time to accurately simulate the CPS nodes’ processing. Faster approaches are available, however as they function at high levels of abstraction, they cannot provide the accuracy required to model the exact behavior of the	2,882,030.00

								system under design so as to guarantee that it meets the requirements in terms of performance and/or energy consumption. The COSSIM project will address all those needs by providing an open-source framework which will a) seamlessly simulate, in an integrated way, both the networking and the processing parts of the CPS, b) perform the simulations orders of magnitude faster, c) provide much more accurate results especially in terms of power consumption than existing solutions, d) report more CPS aspects than any existing tool including the underlying security of the CPS. COSSIM will achieve the above by developing a novel simulator framework based on a processing simulation sub-system (i.e. a "full-system simulator") which will be integrated with a novel network simulator. Furthermore, innovative power consumption and security measurement models will be developed and incorporated to the end framework. On top of that, COSSIM will also address another critical aspect of an accurate CPS simulation environment: the performance as measured in required simulation time. COSSIM will create a framework that is orders of magnitude faster, while also being more accurate and reporting more CPS aspects, than existing solutions, by applying hardware acceleration through the use of field programmable gate arrays (FPGAs), which have been proven extremely efficient in relevant tasks.	
15	UnCoVerCPS	Unifying Control and Verification of Cyber-Physical	EU	CPS	H2020	CPS Cluster	http://cps-vo.org/group/UnCoVerCPS	The proposed research effort provides methods for a faster and more efficient development process of safety- or operation-critical cyber-physical systems in (partially) unknown environments. Cyber-physical systems are very hard to control and verify because of the mix of discrete dynamics (originating from computing elements) and continuous dynamics	4,932,902.25

		Systems						(originating from physical elements). We present completely new methods for de-verticalisation of the development processes by a generic and holistic approach towards reliable cyber-physical systems development with formal guarantees. In order to guarantee that specifications are met in unknown environments and in unanticipated situations, we synthesise and verify controllers on-the-fly during system execution. This requires to unify control and verification approaches, which were previously considered separately by developers. For instance, each action of an automated car (e.g. lane change) is verified before execution, guaranteeing safety of the passengers. We will develop completely new methods, which are integrated in tools for modelling, control design, verification, and code generation that will leverage the development towards reliable and at the same time open cyber-physical systems. Our approach leverages future certification needs of open and critical cyber-physical systems. The impact of this project is far-reaching and long-term: UnCoVerCPS prepares the EU to be able to develop critical cyber-physical systems that can only be realised and certified when uncertainties in the environment are adequately considered. This is demonstrated by applying our ground-breaking methods to automated vehicles, human-robot collaborative manufacturing, and smart grids within a consortium that has a balanced participation of academic and industrial partners.	
16	AXIOM	Agile, eXtensible, fast I/O Module for the cyber-	EU	CPS	H2020	CPS Cluster	www.axiom-project.eu	"We are entering the Cyber-Physical age, in which both objects and people will become nodes of the same digital network for exchanging information. Therefore, in our imaginary, the general expectation is that "things" or systems will become somewhat smart as people, allowing rapid and close interactions not only	3,945,937.50

		physical era						system-system, but also human-system, system-human. More scientifically, we expect that such Cyber-Physical Systems (CPSs) will at least react in real-time, have enough computational power for the assigned tasks, consume the least possible energy for such task (energy efficiency), scale up through modularity, allow for an easy programmability across performance scaling and exploit at best existing standards at minimal costs. The whole set of these expectations impose scientific and technological challenges that need to be properly addressed. The AXIOM project (Agile, eXtensible, fast I/O Module) aims at researching new software/hardware architectures for CPSs to meet the above expectations. The technical approach aims at solving fundamental problems to enable easy programmability of multi-core multi-board systems through the open-source OmpSs programming model, leveraging Distributed Shared Memory (DSM) inspired concepts across the modules. The OmpSs will allow accelerating functions through an FPGA (Agility). In particular, to the best of our knowledge, this is the first time that DSM will be effectively demonstrated on an embedded modular system (eXtensibility). Modular scalability will be possible thanks to a fast interconnect that will enrich the module. To this aim, an innovative modular ARM-based board with enhanced capabilities for interfacing with the physical world will be designed and demonstrated in key scenarios such as Smart Video-Surveillance and Smart Living/Home (Domotic)."	

17	TAMS4CPS	Trans-Atlantic Modelling and Simulation For Cyber-Physical Systems	EU	CPS	FP7	CPS Cluster	www.tams4cps.eu	"Smart systems, in which sophisticated software/hardware is embedded in physical systems are part of everyday life. From simple products with embedded decision making software to massive systems in which hundreds of systems, each with hundreds or thousands of embedded processors, interoperate, the growth of cyber-physical systems (CPS) is likely to accelerate. For Europe to benefit from this expansion, while avoiding the pitfalls that such complexity creates, there must be advances in the modelling and simulation (M&S) of CPS. Collaborative research with the US will be an opportunity to advance European M&S capabilities for CPS. The overall aim of TAMS4CPS is to lay the foundations for concrete EU-US collaboration in modelling and simulation for cyber-physical systems. To achieve this, Loughborough and Newcastle Universities (M&S) will work with Steinbeis Innovation (road mapping) and leading researchers in the field at top US universities to create: - A strategic research and collaboration agenda, endorsed by researchers in EU and US - A set of test cases for model developers to perform collaborative evaluation - A state of the art web-based report to act as a baseline for collaborative research Taking a consultative approach, we will engage industry and academic researchers and M&S users in workshops and web-based meetings to prioritise M&S research challenges and to create a constituency of future collaboration partners for pre-competitive research in the Artemis themes, of: - Architectures principles and models for safe secure Cyber-Physical Systems - Systems design, modelling and virtual engineering for Cyber-Physical Systems	399,649.75

								- Autonomous adaptive and cooperative of Cyber-Physical Systems - Computing platforms and energy management for Cyber-Physical Systems To which is added the exploitation and enabling theme of: - Integration of socio/legal/governance models within modelling frameworks Thus, this project directly addresses European priorities in CPS."	
18	INTO-CPS	INtegrated TOol chain for model-based design of CPSs	EU	CPS	H2020	CPS Cluster	www.into-cps.au.dk	"The aim of the INTO-CPS project is to create an integrated tool chain for comprehensive model-based design of Cyber-Physical Systems (CPSs). The tool chain will support the multidisciplinary, collaborative modelling of CPSs from requirements, through design, down to realisation in hardware and software. This will enable traceability at all stages of the development. INTO-CPS will support the holistic modelling of CPSs, allowing system models to be built and analysed that would otherwise not be possible using standalone tools. We will integrate existing industry-strength tools with high Technology Readiness Levels (TRL 6–9) in their application domains. The solution will be based centrally around Functional Mockup Interface (FMI)-compatible co-simulation. The project focuses on the pragmatic integration of these tools, making extensions in areas where a need has been recognised. The tool chain will be underpinned by a well-founded semantic foundations that ensures the results of analysis can be trusted. The tool chain will provide powerful analysis techniques for CPSs, including connection to SysML; generation and static checking of FMI interfaces; model checking; Hardware-in-the-Loop (HiL) and Software-in-the-Loop	7,956,804.25

								(SiL) simulation, supported by code generation. The tool chain will allow for both Test Automation (TA) and Design Space Exploration (DSE) of CPSs. The INTO-CPS technologies will be accompanied by a comprehensive set of method guidelines that describe how to adopt the INTO-CPS approach, lowering entry barriers for CPS development. The tool chain will be tested with case studies in railways, agriculture, building and automotive. The consortium has 4 academic and 7 industrial partners. The industrial partners comprise both tool vendors and case study owners. The INTO-CPS technology will enable experimenting with design alternatives enabling radical innovation where the overall concept is right first time, even when hardware prototypes does not yet exists."	
19	EoT	Eyes of Things	EU	CPS	H2020	CPS Cluster	http://eyesofthings.eu/	Vision, our richest sensor, allows inferring big data from reality. Arguably, to be "smart everywhere" we will need to have "eyes everywhere". Coupled with advances in artificial vision, the possibilities are endless in terms of wearable applications, augmented reality, surveillance, ambient-assisted living, etc. Currently, computer vision is rapidly moving beyond academic research and factory automation. On the other hand, mass-market mobile devices owe much of their success to their impressing imaging capabilities, so the question arises if such devices could be used as "eyes everywhere". Vision is the most demanding sensor in terms of power consumption and required processing power and, in this respect, existing mass-consumer mobile devices have three problems: 1) power consumption precludes their 'always-on' capability, 2) they would have unused sensors for most vision-based applications and 3) since they have been designed for a definite purpose (i.e. as cell phones, PDAs and "readers") people will not consistently use them for	3,734,830.00

								other purposes. Our objective in this project is to build an optimized core vision platform that can work independently and also embedded into all types of artefacts. The envisioned open hardware must be combined with carefully designed APIs that maximize inferred information per milliwatt and adapt the quality of inferred results to each particular application. This will not only mean more hours of continuous operation, it will allow to create novel applications and services that go beyond what current vision systems can do, which are either personal/mobile or 'always-on' but not both at the same time. Thus, the "Eyes of Things" project aims at developing a ground-breaking platform that combines: a) a need for more intelligence in future embedded systems, b) computer vision moving rapidly beyond academic research and factory automation and c) the phenomenal technological advances in mobile processing power.	
20	IMMORTAL	Integrated Modelling, Fault Management, Verification and Reliable Design Environment for Cyber-Physical Systems	EU	CPS	H2020	CPS Cluster	www.h2020-immortal.eu	"In IMMORTAL, a consortium of leading European academic and industrial players aim at combining their expertise in developing an integrated, cross-layer modelling based tool framework for fault management, verification and reliable design of dependable Cyber-Physical Systems (CPS). Recently, the world has seen emerging CPS modelling frameworks addressing various design aspects such as control, security, verification and validation. However, there have been no considerations for reliability and automated debug aspects of verification. The main aim is to fill this gap by introducing reliable design and automated system debug into CPS modelling. To reach this aim, the project will develop a cross-layer CPS model spanning device (analogue and digital), circuit, network architecture, firmware and software layers. In addition, a holistic fault model for fundamentally	

								different error sources in CPSs (design bugs, wear-out and environmental effects) in a uniform manner will be proposed. Moreover, IMMORTAL plans to develop fault management infrastructure on top of the reliable design framework that would allow ultra-fast fault detection, isolation and recovery in the emerging many-core based CPS networked architectures that are expected to be increasingly adopted in the coming years. As a result, the project will enable development of dependable CPSs with improved reliability and extended effective life-time, ageing and process variations. In line with the expected impacts of the Call, the project will have a significant impact in development time as well as maintenance costs of dependable cyber-physical systems. The tool framework to be developed will be evaluated on a clearly specified real-world use-case of a satellite on-board-computer. However, since the results are more general and applicable to many application domains, including avionics, automotive and telecommunication, demonstration of the framework tools will be applied to CPS examples from other domains as well."	
21	SAFURE	SAFety and secURity by design for interconnected mixed-critical cyber-physical	EU	CPS	H2020	CPS Cluster	https://safure.eu/	SAFURE targets the design of cyber-physical systems by implementing a methodology that ensures safety and security 'by construction'. This methodology is enabled by a framework developed to extend system capabilities so as to control the concurrent effects of security threats on the system behaviour. The current approach for security on safety-critical embedded systems is generally to keep subsystems separated, but this approach is now being challenged by technological evolution towards openness, increased communications and use of multi-core architectures. The objectives of SAFURE are to (1) implement a holistic	5,231,375.00

		systems						approach to safety and security of embedded dependable systems, preventing and detecting potential attacks; (2) to empower designers and developers with analysis methods, development tools and execution capabilities that jointly consider security and safety; (3) to set the ground for the development of SAFURE-compliant mixed-critical embedded products. The results of SAFURE will be (1) a framework with the capability to detect, prevent and protect from security threats on safety, able to monitor from application level down to the hardware level potential attacks to system integrity from time, energy, temperature and data threats; (2) a methodology that supports the joint design of safety and security of embedded systems, assisting the designer and developers with tools and modelling languages extensions; (3) proof-of concept through 3 industrial use cases in automotive and telecommunications; (4) recommendations for extensions of standards to integrate security on safety-critical systems; (5) specifications to design and develop SAFURE-compliant products. The impact of SAFURE will help European suppliers of safety-critical embedded products to develop more cost and energy-aware solutions. To ensure this impact, a community will be created around the project. SAFURE comprises 7 industrial manufacturers, 4 leading universities and research centres and 1 SME.	
22	BEinCPPS	Business Experiments in Cyber Physical Production Systems	EU	CPS, IoT	H2020		www.beincpps.eu	"BEinCPPS Innovation Action aims to integrate and experiment a CPS-oriented Future Internet-based machine-factory-cloud service platform firstly intensively in five selected Smart Specialization Strategy Vanguard regions (Lombardia in Italy, Euskadi in Spain, Baden Wuertemberg in Germany, Norte in Portugal, Rhone Alpes in France), afterwards extensively in all European regions, by involving local competence	7,999,485.75

								centers and manufacturing SMEs. The final aim of this Innovation Action is to dramatically improve the adoption of CPPSs all over Europe by means of the creation, nurturing and flourishing of CPS-driven regional innovation ecosystems, made of competence centers, manufacturing enterprises and IT SMEs. The BE in CPPS project stems upon three distinct pillars: • A FI-based three-layered (machine-factory-cloud) open source platforms federation, integrated from state-of-the-art R&I advances in the fields of Internet of Things, Future Internet and CPS / Smart Systems and able to bi-directionally interoperate data pertaining to the machine, the factory and the cloud levels. • A pan-European SME-oriented experimentation ecosystem. In a first phase of the project, the five Champions will provide requirements to the platforms integrators. In a second phase, an Open Call for IT SMEs developers (applications experiments) will award 10 third parties. In a final third phase, the extended platform will be instantiated and deployed in additional 10 third parties equipment experiment SMEs. • A well-founded method and toolbox for Innovation management, where an existing TRL-based methodology for KETs technology transfer will be enriched by a CPPS certification, education and training programme for young talents and experienced blue collar workers and by a well-founded three-fold (objectives-variables-indicators) method for results assessment and evaluation."	
23	MANTIS	Cyber Physical System based Proactive Collabora	EU	CPS	H2020 - ECSEL	CPS Cluster	www.mantis-project.eu	The overall concept of MANTIS is to provide a proactive maintenance service platform architecture based on Cyber Physical Systems that allows to estimate future performance, to predict and prevent imminent failures and to schedule proactive maintenance. Maintenance is no longer a necessary evil that costs what it costs, but	9,791,974.05

		tive Maintenance							an important function that creates additional value in the business process as well as new business models with a stronger service orientation. Physical systems (e.g. industrial machines, vehicles, renewable energy assets) and the environment they operate in, are monitored continuously by a broad and diverse range of intelligent sensors, resulting in massive amounts of data that characterise the usage history, operational condition, location, movement and other physical properties of those systems. These systems form part of a larger network of heterogeneous and collaborative systems (e.g. vehicle fleets or photovoltaic and windmill parks) connected via robust communication mechanisms able to operate in challenging environments. MANTIS consists of distributed processing chains that efficiently transform raw data into knowledge while minimising the need for bandwidth. Sophisticated distributed sensing and decision making functions are performed at different levels in a collaborative way, ranging from local nodes to locally optimise performance, bandwidth and maintenance; to cloud-based platforms that integrate information from diverse systems and execute distributed processing and analytics algorithms for global decision making. The research addressed in MANTIS will contribute to companies' assets availability, competitiveness, growth and sustainability. Use cases will be the testing ground for the innovative functionalities of the proactive maintenance service platform architecture and for its future exploitation in the industrial world. Results of MANTIS can be utilised directly in several industry areas and different fields of maintenanance.	

24	C2NET	Cloud Collaborative Manufacturing Networks	EU	CPS	H2020 - ECSEL		c2net-project.eu	"The goal of C2NET Project is the creation of cloud-enabled tools for supporting the supply network optimization of manufacturing and logistic assets based on collaborative demand, production and delivery plans. C2NET Project will provide a scalable real-time architecture, platform and software to allow the supply network partners: to master complexity and data security of the supply network, to store product, process and logistic data, to optimize the manufacturing assets by the collaborative computation of production plans, to optimize the logistics assets through efficient delivery plans and to render the complete set of supply chain management information on the any digital mobile device (PC, tablets, smartphones, …) of decision makers enabling them to monitor, visualize, control, share and collaborate. The Project results will be: i) the C2NET Data Collection Framework for IoT-based continuous data collection from supply network resources; ii) the C2NET Optimizer for the optimization of manufacturing and logistics assets of the supply network by the collaborative computation of production, replenishment and delivery plans; iii) the C2NET Collaboration Tools for providing support to the collaborative processes of the supply network, and iv) the C2NET Cloud Platform (C2NET CPL) to integrate the data module, the optimizers and the collaborative tools in the cloud. C2NET will be designed to comprehensively cover the entire supply chain considering all stages of manufacturing, distribution and sales to supply a product to market. Different actors in the supply network as plant managers, planners, carriers, shop floor workers, shop assistants or customers are potential users of the services that will be offered by C2NET. A distinguishing feature of these services is to	7,126,494.75

								have complete visibility and real-time status of the entire supply chain at all times looking for an optimal response to maximize both local and global benefit."	
25	CRYSTAL	CRitical sYSTem engineering AcceLeration	EU	CPS	ARTEMIS-JU	CPS Cluster	www.crystal-artemis.eu		
26	EMC2	Embedded Multi-Core Systems for Mixed Criticality Applications in Dynamic and Changeable Real-time Environments	EU	CPS	ARTEMIS-JU		www.artemis-emc2.eu	EMC² – 'Embedded Multi-Core systems for Mixed Criticality applications in dynamic and changeable real-time environments' is an ARTEMIS Joint Undertaking project in the Innovation Pilot Programme 'Computing platforms for embedded systems' (AIPP5). Embedded systems are the key innovation driver to improve almost all mechatronic products with cheaper and even new functionalities. They support today's information society as inter-system communication enabler. A major industrial challenge arises from the need to face cost efficient integration of different applications with different levels of safety and security on a single computing platform in an open context. EMC² finds solutions for dynamic adaptability in open systems, provides handling of mixed criticality applications under real-time conditions, scalability and utmost flexibility, full scale deployment and management of integrated tool chains, through the entire lifecycle. The objective of EMC² is to establish Multi-Core technology in all relevant Embedded Systems domains. EMC² is a project of 99 partners of embedded industry and research from 19 European countries with an effort of about 800 person years and a total budget of about	95,130,000.00

								100 million Euro.	
27	ARROWHEAD	Service Interoperability enabling collaborative automation	EU	CPS	ARTEMIS-JU		www.arrowhead.eu	Arrowhead is addressing efficiency and flexibility at the global scale by means of collaborative automation for five application verticals. That means production (manufacturing, process, energy), smart buildings and infrastructures, electro-mobility and virtual market of energy.'Our vision is to enable collaborative automation by networked embedded devices. The grand challenges are enabling the interoperability and integrability of services provided by almost any device. We assume that a service-based approach will be the feasible technology that enables collaborative automation in an open-network environment connecting many embedded devices. The success of Arrowhead technology depends not only on the technology but also on the capability to create and pursue innovations supported by the core of our technology. If successful the approach is expected to strongly contribute to very significant reduction, 75 % or more, in the design and engineering efforts for the predicted multi-billion networked devices. Our society is facing both energy and competitiveness challenges. These challenges are tightly linked and require new dynamic interactions between energy producers and energy consumers, between machines, between systems, between people and systems, etc. Cooperative automation is the key for these dynamic interactions and is enabled by the technology developed around the Internet of Things and Service Oriented Architectures.	70,390,000.00
28	ACCUS	Adaptive Cooperative Control in Urban	EU	CPS	ARTEMIS-JU		http://projectaccus.eu/	Adaptive Cooperative Control in Urban (sub) Systems like traffic, energy, and outdoor lighting are managed by self-contained embedded systems. ACCUS aims to establish long-term sustainable urban areas and develop	8,870,000.00

		(sub) Systems						the necessary infrastructure to enable the sharing of urban subsystem services, demonstrating this through selected use-cases in a real-life urban setting.	
29	eScop	Embedded systems for Service-based control of Open Manufacturing and Process Automation	EU	CPS	ARTEMIS-JU		www.escop-project.eu	Embedded systems Service-based Control for Open manufacturing and Process automation, or E-SCOP, aims to overcome the current drawbacks for the shop-floor control level, thus improving the state of the art of the overall production control system architecture. This is achieved by introducing an innovative approach based on the combination of embedded systems, ontology-based knowledge management and service-oriented architecture.	5,820,000.00
30	CPSoS	Towards a European Roadmap on Research and Innovation in Engineering and Management of Cyber-physical Systems of Systems	EU	CPS, SoS	FP7	SoS Cluster	http://www.cpsos.eu	CPSoS is a 30-month Support Action that will provide an exchange platform for Systems of Systems (SoS) related projects and communities. It will focus on the challenges posed by the engineering and the operation of technical systems in which computing and communication systems interact with large complex physical systems. Its approach will be simultaneously integrative, aiming at bringing together knowledge from different SoS related communities and applications driven.It will bridge between the different approaches to the design, analysis and control of systems of systems that are pursued by different communities in theory and applications and relate the methods and tools proposed for dealing with SoS to key application domains which are important for Europe's competiveness as well as for the well-being of its citizens. The project will examine in depth application-specific issues, capture cross-industry and cross-application findings and propose new avenues for SoS analysis, design and control, towards a science	

								of systems of systems and a European R&I agenda on SoS, involving different scientific communities, application domain experts, end-users and vendors.Three Working Groups will be created: on SoS in Transportation and Logistics, Physically connected SoS, and Tools for SoS Engineering and Management. Public events will be organised in conjunction with the Working Group meetings, floor will be given to SoS-related projects to expose their outcomes. The contributions will be summarized in a concise strategic policy document 'European Research and Innovation Agenda on Cyber-physical Systems of Systems', supported by a set of in-depth technical papers, and presented at a symposium 'Cyber-physical Systems of Systems Meeting Societal Challenges'. The project will prepare the EU stakeholders for extracting a competitive advantage from the recent and the future developments in the area of SoS. 	
31	CyPhERS	Cyber-Physical European Roadmap and Strategy	EU	CPS, SoS	FP7	SoS Cluster	www.cyphers.eu	The ongoing integration of software-intensive embedded systems and global communication networks into Cyber-Physical Systems (CPS) is considered to be the next revolution in ICT with a lot of game-changing business potential and novel business models for integrated services and products. CPS will be a core enabling technology for securing economic leadership in embedded systems/ICT, having an enormous social and economic importance, and making decisive contributions to societal challenges. Europe is well positioned to meet this merging of the physical and virtual worlds. However, to effectively address the associated challenges, a strategic agenda is needed to ensure Europe's competitiveness. This Support Action will systematically survey, analyse, and evaluate the economic, technical, scientific, and societal significance of Cyber-Physical Systems for Europe. The project will	535,000.00

								develop an integrated strategic CPS research agenda for Europe, and derive comprehensive recommendations for action that will cover the identification and prioritization of research areas, support measures for both horizontal and vertical cooperation, and an outline of possible research partnerships, and will address questions of research funding as well as the issues of training, standardization and policies.	
32	ADVANCE	Advanced Design and Verification Environment for Cyber-physical System Engineering	EU	CPS	FP7		www.fp7-advance.eu	The overall objective of ADVANCE is the development of a unified tool-based framework for automated formal verification and simulation-based validation of cyber-physical systems. Unification will be achieved through the use of a common formal modelling language supported by methods and tools for simulation and formal verification. An integrated tool environment will provide support for construction, verification and simulation of models. The delivered methods and tools will overcome significant deficiencies in current practices in cyber-physical systems engineering that make verification and validation hugely costly and time consuming. The ADVANCE consortium consists of six strong and complementary partners representing a combination of leading European industrial players in systems engineering along with academic partners with internationally leading expertise in formal verification and simulation tools. Systerel and the Univs. of Düsseldorf and Southampton will lead the development of novel methods and tools while Alstom and Critical Software will apply these to the engineering of intelligent transport and energy systems. Selex ES, as the end user of the energy system development, will bring industrial and commercial experience to the exploitation of the methods and tools developed. We will exploit recent advances in technology for high-level formal modelling (Event-B & Rodin) with strong support	2,550,000.00

								for formal verification; The Rodin tool will be further strengthened and augmented with novel approaches to multi-simulation and testing. ADVANCE will reinforce European scientific excellence and technological leadership in the design and operation of large-scale complex systems, improve industrial competitiveness through strengthened capabilities in advanced embedded systems, in monitoring, control and optimisation of large-scale complex systems, in areas like energy, transport, and production, and in engineering of large-scale systems. In particular, the outcome of Case Study 1 will be to improve safety in the railway domain for dynamic trusted railway interlocking, and the outcome of Case Study 2 will be to have an impact on the efficiency of energy distribution in the emerging smart grid market, in which Selex ES has already a market presence. For the railway case study, an experienced safety and certification expert from the Alstom RAMS team will contribute to the identification of safety requirements in the formal model and to the assessment of the compliance of the ADVANCE process to certification requirements. For the smart grid case study, Selex ES are providing expert input into the formal modeling of Low Voltage Networks which will result in a smart grid solution which not only is energy efficient but can also avert transformer failure, thereby ensuring more reliable and cost-effective energy supply to meet future user demand.	
33	Road4FAME	Development of a Strategic Research and Innovation	EU	CPS, SoS	FP7	CPS Cluster	http://road4fame.eu/	In the global economy, European manufacturers face increased competitive pressure and continuous innovation is required to remain competitive. ICT offer a seemingly boundless source of innovation, promising to deliver precisely the capabilities European manufacturing businesses require to remain competitive and to furthermore enable them to run their business in	859,968.00

		Roadmap for Future Architectures and Services for Manufacturing in Europe and Derivation of Business Opportunities						a sustainable way. However, the potential of ICT in manufacturing is far from being fully exploited. Towards making ICT an enabler for innovation in manufacturing, employing suitable IT architectures and services is of fundamental importance. In manufacturing, however, these are in many cases not in place because a) decision makers are not aware of the most suitable architectures and services, b) decision makers may be sceptical towards novel paradigms, c) novel paradigms may not be tailored to meet the demanding manufacturing requirements or d) promising paradigms may indeed not have been discovered for manufacturing. The Road4FAME project has been conceived to respond to this situation by developing a holistic research and innovation roadmap for architectures and services, aligned with the concrete needs and requirements of manufacturing. In the roadmapping process, promising paradigms and concepts shall be evaluated with respect to their applicability and utility as architectures and services in manufacturing. The roadmap will serve to guide the R&D of architectures and services to match precisely the requirements of European manufacturing businesses. Furthermore, the Road4FAME project will yield strategy recommendations and outline business opportunities tailored to decision makers of European manufacturing businesses enabling them to harness the potential of ICT through the right architectures and services. Lastly, the Road4FAME project will contribute to building a constituency united by the commonly faced challenges and their common quest for the right architectures and services to address these.	

34	DYMASOS	Dynamic Management of Physically Coupled Systems of Systems	EU	CPS, SoS	FP7	SoS Cluster	http://www.dymasos.eu/	The well-being of the citizens in Europe depends on the reliable and efficient functioning of large interconnected systems, such as electric power systems, air traffic control, railway systems, large industrial production plants, etc. Such large systems consist of many interacting components, e.g. generation plants, distribution systems, and large and small consumers. The subsystems are usually managed locally and independently, according to different policies and priorities. The dynamic interaction of the locally managed components gives rise to complex behaviour and can lead to large-scale disruptions as e.g. black- outs in the electric grid.Large interconnected systems with autonomously acting sub-units are called systems of systems. DYMASOS addresses systems of systems where the elements of the overall system are coupled by flows of physical quantities, e.g. electric power, steam or hot water, materials in a chemical plant, gas, potable water, etc. Within the project, new methods for the distributed management of large physically connected systems with local management and global coordination will be developed.The approaches explored are: Modelling and control of large systems analogously to the evolution of the behaviour of populations in biological systems; Market-like mechanisms to coordinate independent systems with local optimisation functions; Coalition games where agents that control the subsystems dynamically group to pursue common goals.The properties of the distributed management and control techniques for large systems of systems are investigated theoretically and validated in large-scale simulations of case studies provided by industrial partners in the fields of electric grid management and industrial production management.From the experience gained in the case studies, conclusions will be drawn about the suitability	2,700,000.00

								of the proposed management and control mechanisms for certain classes of systems of systems.The expected technical outcomes of the project are: • Innovation in distributed management methods for complex interconnected systems •Progress in methods for the rigorous analysis and validation of systems of systems; •Improvements in the management of electric grids and of large production complexes; •Tools for the engineering of management systems for systems of systems; •Identification of technology gaps in advanced management and coordination methods.The developed coordination methods will lead to improved system stability and lower resource consumption in industrial production, and in electric-power generation and distribution. This will result in a reduction of the CO2-emissions, higher competitiveness of the European industry and lower prices for the customers. Thus, DYMASOS will contribute to the goal of a greener and more competitive Europe.	
35	Road2SoS	Development of strategic research and engineering roadmaps in Systems of Systems Engineering and related case	EU	CPS, SoS	FP7	SoS Cluster	http://road2sos-project.eu/	The emerging System-of-Systems (SoS) concept describes the large scale integration of many independent self-contained systems to satisfy global needs or multiple requests. The increasing number of interacting (mostly embedded) systems in our strongly connected society and industry as well as the growing overall complexity of systems have triggered a paradigm shift and the need to enhance the classical view of Complex System Engineering towards SoS Engineering. SoS find their application in many highly relevant areas to our society: air-traffic control, urban transport, smart grids and integrated multi-site production, among others. The inherent complexity of SoS as well as the need to meet optimised performance for both the individual systems and SoS, raise many questions which are the object of important research efforts nowadays.	991,987.00

		studies						The project Road2SoS will develop advanced research and engineering roadmaps to identify future RTD and Innovation (RTD&I) strategies for Europe in the field of SoS Engineering in four key domains: •Distributed energy generation and smart grids •Integrated multi-site industrial production •Multi-modal traffic control •Emergency & crisis management The roadmaps will comprise the latest high-level scientific results and will identify trends in research and development and associate them to future product and application visions.They will also be integrated in the socio-economic context, via workshops and case studies in order to facilitate the transfer and integration of European RTD\I results and developments from the SoS Engineering sector to relevant stakeholders. The project will help European organisations to stay at the front of these new developments and to be prepared for future challenges in the industrial applications of SoS in order to reinforce the leading position of the European industry in Complex System Engineering.	
36	AMADEOS	Architecture for Multi-criticality Agile Dependable Evolutionary Open System-of-Systems	EU	CPS, SoS	FP7	SoS Cluster	http://amadeos-project.eu/	The objective of this research proposal is to bring time awareness and evolution into the design of System-of-Systems (SoS), to establish a sound conceptual model, a generic architectural framework and a design methodology, supported by some prototype tools, for the modeling, development and evolution of time-sensitive SoSes with possible emergent behaviors. Special emphasis is placed on evolution, emergence, dependability (e.g. safety, availability) and security, considering embedded devices and the cloud as the execution platform. The concept of evolution will be addressed from two complementary perspectives, considering both long-term evolution and short-term unexpected changes (e.g., failures) in the constituent systems. The project starts with a study of fielded	2,481,999.00

								industrial SoSs, where the handling of time and the evolution aspects will be in the center of the analysis, in the domains of disaster management, transport, and smart grid applications. The following development of the conceptual model, the architectural framework, the design methodology and some extensions to UML-based tools will form the core of the project work. In place of the traditional guarantees that were the target for more closed and static systems, the architectural framework will be based on the concept of guaranteed best adaptation under the given constraints, sometimes just monitoring how the environment evolves, and influencing how the SoS takes mitigating actions. The viability of the framework will be validated on a case study of a CPS, a small smart grid application, where guaranteed responsiveness, evolution, dependability and security are essential requirements. The research is based on the in-depth experience of some of the key researchers of the consortium in the fields of architecture design, real-time systems, dependability, security and the development of large systems-of-systems in such diverse domains as disaster management, the transport sector, and energy distribution.	
37	Local4Global	SYSTEM-OF-SYSTEMS THAT ACT LOCALLY FOR OPTIMIZING GLOBALLY	EU	CPS, SoS	FP7	SoS Cluster	http://local4global-fp7.eu	Today's Technical Systems of Systems (TSoS) such as transport, traffic and energy management systems require the deployment of an expensive-to-deploy and operate sensor and communication infrastructure. Moreover, they need a very time/effort-consuming modelling, analysis and control design procedure in order to achieve an efficient performance. On the contrary, Natural Systems of Systems (NSoS) such as the human brain, animal herds (swarms), teams of interacting/cooperating humans or animals achieve a highly efficient, elegant and supreme functionality	2,515,864.00

								without the need of an expensive infrastructure as they primarily rely on local information between neighbouring systems and, most importantly, they do not need any modelling, analysis or control design tools to achieve such a functionality. If the powerful attributes of NSoS were possible to be transferred and embedded into TSoS, this would lead not only to more efficient TSoS operations but, most importantly, to TSoS that are significantly easier, safer and more economical to design, deploy and operate. This is actually the main objective of Local4Global: to develop, test and evaluate a new groundbreaking, generic and fully-functional methodology/system for controlling TSoS which - as in the NSoS case - optimizes the TSoS performance at the global level without the need of deployment and operation of an expensive sensor and communication infrastructure and, most importantly, without the need for the use of elaborate and time/effort consuming modelling, analysis and control design tools. By embedding in TSoS attributes currently found only in NSoS, Local4Global's ambition is to develop a system that can be embedded in every day TSoS operations, produce substantial savings and Quality-of-Service improvements with the requirement of using the minimum possible infrastructure and minimum installation/operation effort. The economic and societal impact and consequences of the availability of such a system will be tremendous in literally any activity of everyday life: for instance, drivers/travellers will spent significantly less time for commuting, building occupants will see their energy bills significantly reduced and, most importantly, energy consumption and pollution will be substantially reduced. Furthermore, Local4Global application will not be only limited to areas and systems where no sophisticated control is currently employed	

								(due to the requirement for an elaborate infrastructure). It will also be of great significance to areas and systems where, despite that the infrastructure is there, current control and management systems 'cannot do the job'. The Local4Global advances will lead to a fully-functional and ready-to-use system (Local4Global final product) - delivered in the form of an embedded, web-based, 'plug-and-play' software system for generic TSoS. This system will be deployed and extensively tested in 2 real-life TSoS Use Cases, a Traffic TSoS Use Case and a Building TSoS Use Case.	
38	AGILE	RAPIDLY-DEPLOYABLE, SELF-TUNING, SELF-RECONFIGURABLE, NEARLY-OPTIMAL CONTROL DESIGN FOR LARGE-SCALE NONLINEAR SYSTEMS	EU	CPS	FP7	SoS Cluster	not available	The inability of existing theoretical and practical tools to scaleably and efficiently deal with the control of complex, uncertain and time-changing large-scale systems, not only leads to a effort-, time- and cost-consuming deployment of Large-Scale Control Systems (LSCSs), but also prohibits the wide application of LSCS in areas and applications where LSCSs could potentially have a tremendous effect in improving system efficiency and Quality of Services (QoS), reducing energy consumption and emissions, and improving the day-to-day quality of life. Based on recent advances of its partners on convex design for LSCSs and robust and efficient LSCS self-tuning, the AGILE project aims at developing and evaluating an integrated LSCS-design methodology, applicable to large-scale systems of arbitrary scale, heterogeneity and complexity and capable of: - Providing proactive, arbitrarily-close-to-optimal LSCS performance; - Being intrinsically self-tuneable, able to rapidly and efficiently optimize LSCS performance when short- medium- and long-time variations affect the large-scale system; - Providing efficient, rapid and safe fault-recovery and LSCS re-configuration; and, - Achieving all the above, while being scalable and modular. To ease implementation and	1,299,094.00

								deployment of the AGILE system in existing open-architecture SCADA/DCS infrastructures, a set of open-source interfacing tools will be developed. The integrated LSCS design system to be developed within AGILE along with the interfaces will be extensively tested and evaluated into two real-life large-scale Test Cases (a 20-junction urban traffic network and a large-scale energy-controlled building) possessing a rich variety of design and performance characteristics, extremely complex nonlinear dynamics, highly stochastic effects, uncertainties and modeling errors, as well as reconfiguration and modular design requirements.	
39	DANSE	Designing for Adaptability and evolutioN in System of systems Engineering	EU	CPS, SoS	FP7	SoS Cluster	http://danse-ip.eu/	DANSE focuses on the development of a new methodology to support evolving, adaptive and iterative System of Systems (SoS) life-cycle models based on a formal semantics for SoS inter-operations and supported by novel tools for analysis, simulation, and optimisation.	
40	COMPASS	Comprehensive Modelling for Advanced Systems of Systems	EU	CPS, SoS	FP7	SoS Cluster	http://www.compass-research.eu/	The target of our research is the integration of well-founded engineering notations, methods and tools to support developers in building models of SoSs and analysing the global SoS-level properties of these models. These integrated techniques are intended to allow the comparison of alternative architectures and allocations of responsibilities to constituent subsystems. It is vital that the techniques we develop are accessible to a wide range of developers, so they must provide different levels of description, starting with a graphical view in a notation such as SysML that is easy for	

								stakeholders to understand. We plan to extend SysML and link it to an underlying formal notation, built from established formalisms, extended with SoS-specific aspects. This level will be accessible to stakeholders trained in the formal notation, which we call the COMPASS Modelling Language (CML). CML's formality means that many kinds of analysis can be conducted and some will also be presentable at the SysML level. The overall approach is presented in the diagram below. Note that CML extensions could be developed for other modelling languages, not only SysML.	
41	T-Area-SoS	Trans-Atlantic Research and Education Agenda in System of Systems	EU	CPS, SoS	FP7	SoS Cluster	https://www.tareasos.eu/	The 24-month T-AREA-SoS Support Action addresses ICT-2011.3.3, target outcome g) 'to analyse international research agendas to prepare concrete joint R&D initiatives for international collaboration, particularly with the USA in the area of System of systems (SoS)'. The concept for this proposal is to exploit the established networks of the participants and associates (based in EU and US) including the extant activities they lead within the IEEE and INCOSE in the area of SoS to explore and evolve SoS R& D themes and priorites for FP7/FP8 and other international programmes that will lead to outcomes that address societal needs, with exemplars across a wide variety of sectors including non-traditional aspects of Energy, Transport, and Production, using Engineering of SoS (i.e. SoSE) as a mainstream discipline for the management of large complex systems. It is a basic premise of this proposal that SoS engineering includes, as a central component, the consideration of societal needs and issues within the management of large, socially-significant system of systems. It is also understood that SoSE is an emerging discipline that deals with ultra large systems that include many heterogeneous systems that may be independently owned and/or operated, distributed,	489,000.00

								evolutionary in nature and which exhibit emergent behaviours. he outputs from this Support Action will be a strategic research agenda that will create the environment for the development of concrete research initiatives through which the EU and the US which will collaborate to enhance existing research programmes and set the scene for future programmes. These outputs will be supported by an analysis of the state of the art and high level definition of research requirements in SoSE, and a thesaurus to enable concepts to be shared across industrial sectors and technical disciplines. Through the creation of a SoSE research agenda commonly agreed by EU and US stakeholders with initiatives to be embedded in future FP7/FP8 programmes the aim of this proposal to increase European competitiveness in, and improve the societal impact of, the development and management of large complex systems in a range of sectors.	
42	I4MS	ICT Innovation for Manufacturing SMEs	EU	CPS	FP7	EC I4MS Initiative	http://i4ms.eu/i4ms/i4ms.php	I4MS (ICT Innovation for Manufacturing SMEs) is the initiative promoted by the EC to support the European leadership in manufacturing through the adoption of ICT technologies. In fact, Europe's competiveness in that sector depends on its capacity to deliver highly innovative products, where the innovation often originates from advances in ICT. Therefore, I4MS aims at promoting leading edge technologies, developed in FP7 large ICT projects, in the following areas: Robotics HPC cloud based simulation services Laser based applications Intelligent sensor-based equipment The initiative is dedicated to the whole European manufacturing industry with special focus on SMEs to	77,000,000.00

								address their needs to access to technology, infrastructures and new markets. I4MS will operate through two main instruments:	
43	CLOUT	ClouT: Cloud of Things for empowering the citizen clout in smart cities	EU	IoT	FP7	AIOTI - IERC- European Research Cluster on the Internet of Things	http://clout-project.eu/	ClouT's overall concept is leveraging the Cloud Computing as an enabler to bridge the Internet of Things with Internet of People via Internet of Services, to establish an efficient communication and collaboration platform exploiting all possible information sources to make the cities smarter and to help them facing the emerging challenges such as efficient energy management, economic growth and development. ClouT will provide infrastructures, services, tools and applications that will be reused by different city stakeholders such as municipalities, citizens, service developers and application integrators, in order to create, deploy and manage user-centric applications taking benefit of the latest advances in internet of things and cloud domains. ClouT with its user-centric approach, will also offer to end-users the possibility of creating their own Cloud services and share them with other citizens. ClouT will have the following major outputs: i) a smart city infrastructure with a near to infinity processing and storage capacity of data from trillions of things and people that are integrated via virtual services in the Cloud while keeping their universal interoperability; ii) a set of platform level tools and services aiming at facilitating IoT application development, deployment and supervision iii) secure data access and processing mechanisms that can handle big data acquired from the heterogeneous sources in quasi real-time; iv) innovative city applications and field trials in four pilot cities: Santander and Genova in Europe, Mitaka and Fujisawa in Japan. ClouT will reduce costs and time to develop and deploy new applications by using new flavours of public-private	1,499,967.00

								partnerships inspired by the Cloud models, supporting win-win strategies for all stakeholders. ClouT's ultimate goal is to join the forces and create a long-lasting synergy for smart city initiatives between Europe and Japan.	
44	VITAL	Virtualized programmable InTerfAces for innovative cost-effective IoT depLoyments in smart cities	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	http://vital-iot.eu/	Internet-of-Things (IoT) applications are currently based on multiple architectures, standards and platforms, which have led to a highly fragmented IoT landscape. This fragmentation is evident in the area of smart cities, which typically comprise several technological silos (i.e. IoT systems that have been developed and deployed independently). Nowadays there is a pressing need to remove these silos in order to allow cities to share data across systems and coordinate processes across domains, thereby essentially improving sustainability and quality of life. In response to this need, VITAL will realize a radical shft in the development, deploment and operation of IoT applications, through introducing an abstract virtualized digital layer that will operate across multiple IoT architectures, platforms and business contexts. Specifically, VITAL will provide platform and business context agnostic access to Internet-Connected-Objects (ICO). Moreover, it will research virtualized filtering, complex event processing (CEP) and business process management mechanisms, which will be operational over a variety of IoT architectures/ecosystems. The mechanisms will compromise the diverse characteristics of the underlying ecosystems, thereby boosting interoperability at the technical and business levels. VITAL will also provide development and governance tools, which will leverage the project's interfaces for	2,695,000.00

								virtualized access to ICOs.VITAL will allow solution providers to (re)use a wider range of data steams, thereby increasing the scope of potential applications. It will also enable a more connected/integrated approach to smart city applications development, which will be validated in realistic deployments in London and Istanbul. The partners will contribute and adapt a host of readily available urban infrastructures, IoT platforms and novel IoT applications, which will ease the accomplishment of the project's goals based on an optimal value for EC money.	
45	SOCIOTAL	Creating a socially aware and citizen-centric Internet of Things	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	http://sociotal.eu/	SOCIOTAL addresses a crucial next step in the transformation of an emerging business driven Internet of Things (IoT) infrastructure into an all-inclusive one for the society by accelerating the creation of a socially aware citizen-centric Internet of Things. It will close the emerging gap between business centric IoT enterprise systems and citizen provided infrastructure. SOCIOTAL will establish an IoT eco-system that puts trust, user control and transparency at its heart in order to gain the confidence of everyday users and citizens. By providing adequate socially aware tools and mechanisms that simplify complexity and lower the barriers of entry it will encourage citizen participation in the Internet of Things. This will add a novel and rich dimension to the emerging IoT ecosystem, providing a wealth of opportunities for the creation of new services and applications that address true societal needs and allow the improvement of the quality of life across European cities and communities.SOCIOTAL builds on the foundations of emerging IoT architectures and introduces the following innovative key target outcomes, ensuring that privacy and trust are deeply embedded in the resulting architecture:1) A governance, trust and reputation framework combining a set of innovative enablers that	2,811,000.00

								addresses the challenges of massive crowd-sourced IoT infrastructure2) A privacy-preserving context-sensitive communication framework for IoT devices with adequate security enablers3) A detailed understanding of technological and socio-economic barriers for citizen participation in an IoT4) An intuitive environment inspired by social media tools that provides increased awareness and control and empowers citizens to easily manage access to IoT devices and information, while allowing IoT enabled citizen centric services to be created through open community APIs5) Services piloted in two cities demonstrating the value of SOCIOTAL to real word communities 	
46	RERUM	REliable, Resilient and secUre IoT for sMart city applicatio ns	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	https://ict-rerum.eu/	RERUM will develop, evaluate, and trial an architectural framework for dependable, reliable, and secure networks of heterogeneous smart objects supporting innovative Smart City applications. The framework will be based on the concept of 'security and privacy by design', addressing the most critical factors for the success of Smart City applications.The technical approach aims to enable tighter integration of Internet of Things (IoT) technology in the Smart City domain. The ultimate target is to provide innovative applications that will improve the citizens' quality of life. RERUM will design mechanisms to manage resources in an efficient and effective manner, ensuring confidentiality, authenticity, integrity, and availability of data gathered by IoT objects. The RERUM approach is scenario-independent, hence applicable to the entire gamut of future Smart City applications. To achieve these goals RERUM will:• analyse industrial and Smart City application requirements for steering the overall system design,• develop a new, technically innovative framework to interconnect a large number of heterogeneous software and hardware smart objects,	3,497,000.00

								using novel approaches, e.g. based on virtualization and cognitive radio,• increase the energy efficiency of IoT to provide sustainable infrastructures in the economic and ecologic sense,• model the trustworthiness of the IoT and enable secure and reliable exchange of trusted information,• enable secure bootstrapping and self-monitoring of networks to detect and mitigate security events,• perform proof-of-concept experiments and real-world trials in two Smart City environments to assess the project results with respect to technical feasibility and user acceptance, and• assess the portability and its 'scenario-agnostic' characteristics by cross-evaluating the trials' results.To vertically address these aspects RERUM unites a consortium of complementary stakeholders covering all areas of the IoT-based Smart Cities paradigm.	
47	COSMOS	Cultivate resilient smart Objects for Sustainable city applicatiOnS	EU	IoT	FP7	AIOTI - IERC- European Research Cluster on the Internet of Things	http://iot-cosmos.eu/	In a world of multi-stakeholder information and assets provision on top of millions of real-time interacting and communicating things, COSMOS aims at enhancing the sustainability of smart city applications by allowing IoT based systems to reach their full potential. COSMOS will enable things to evolve and act in a more autonomous way, becoming more reliable and smarter. Things will be able to learn based on others experiences, while situational knowledge acquisition and analysis will make things aware of conditions and events affecting their behaviour. Adaptive selection approaches will manage the uncertainty and volatility introduced due to real-world dynamics. COSMOS will integrate decentralized management mechanisms in IoT based systems allowing applications to exploit millions of things. Socially-enriched coordination will consider the role and participation scheme of things in and across networks. Management decisions and runtime adaptability will be based on things security, trust, administrative, location,	3,188,000.00

								relationships, information, and contextual properties. COSMOS will facilitate IoT based systems with end-to-end security and privacy, with hardware-coded security, approaches for security and privacy on storage and with the introduction of an innovative concept, the concept of Privelets for IoT services. COSMOS will deliver data and information management mechanisms to handle the exponentially increasing 'born digital' data. Extended complex event processing and social media technologies will extract only the valuable knowledge from the information flows, while workload-optimized data object stores will facilitate efficient storage by also exploring the interplay between storage and analytics on networks of data objects. COSMOS enables smart city IoT applications to take full advantage of its technologies, through 3 representative scenarios: Smart Heat and Electricity Management (London), Eco-Context Awareness for Smart Public Transport (Madrid) and IoT Business Eco-System (Taipei)	
48	CITY PULSE	Real-Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	http://www.ict-citypulse.eu/page/	CityPulse provides innovative smart city applications by adopting an integrated approach to the Internet of Things and the Internet of People. The project will facilitate the creation and provision of reliable real-time smart city applications by bringing together the two disciplines of knowledge-based computing and reliability testing.	

49	ALMANAC	Reliable Smart Secure Internet Of Things For Smart Cities	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	http://www.almanac-project.eu/	ALMANAC will develop a service delivery platform with corresponding technologies that integrates Internet of Things (IoT) edge networks (termed capillary networks) with Telco's metro access networks thus enabling an integrated Smart City Information System for green and sustainable Smart City applications.The Smart City Platform (SCP) collects, aggregates, and analyses real-time or near real-time data from appliances, sensors and actuators, smart meters, etc. deployed to implement Smart City processes. ALMANAC aims at achieving pervasiveness by defining a capillary radio network providing local Machine-to-Machine (M2M) connectivity to smart things. The SCP allows decision support and implements intelligent control of the devices through the capillary networks with a M2M management platforms.The key element of the platform is a middleware based on a SOA architecture supporting semantic interoperability of heterogeneous resources, devices, services and data management. The middleware supports various Smart City applications by leveraging on a communication network built dynamically by federating private and public networks.The SCP supports end-to-end security and privacy. It can also integrate services that, although being natively external to the SCP itself, enrich the set of data and information used by the Smart City applications.The technological development will be driven by Smart City requirements from the City of Torino. Two specific applications (waste management and water supply) have been selected for proof-of-concept implementation and evaluation.The project will also develop a business model framework based on the concept of dynamic value constellations of actors based on public-private partnerships.The Consortium brings together research groups with strong academic	2,995,000.00

								background, SMEs, industries and a big European Smart City. The project will leverage on the work done by several partners in previous FP6, FP7 projects and PPP, ARTEMIS projects.	
50	UNIFY-IoT	Supporting Internet of Things Activities on Innovation Ecosystems	EU	IoT	H2020	AIOTI - IERC-European Research Cluster on the Internet of Things	https://www.hipeac.net/network/projects/6825/unify-iot/	Coordination and Support Action to stimulate the collaboration between IoT projects, between the potential IoT platforms and support these in sustaining the IoT ecosystems developed by focusing on complementary actions, e.g., fostering and stimulating acceptance of IoT technology as well as the means to understand and overcome obstacles for deployment and value creation.	1,000,000.00
51	VICINITY	Open virtual neighbourhood network to connect intelligent buildings and smart objects	EU	IoT	H2020	AIOTI - IERC-European Research Cluster on the Internet of Things	http://www.vicinity-h2020.eu/	"The lack of interoperability is considered as the most important barrier to achieve the global integration of IoT ecosystems across borders of different disciplines, vendors and standards. Indeed, the current IoT landscape consists of a large set of isolated islands that do not constitute a real internet, preventing the exploitation of the huge potential expected by ICT visionaries. To overcome this situation, VICINITY presents a virtual neighborhood concept, which is a decentralized, bottom-up and cross-domain approach that resembles a social network, where users can configure their set ups, integrate standards according to the services they want to use and fully control their desired level of privacy. VICINITY then automatically creates technical interoperability up to the semantic level. This allows users without technical background to get connected to the vicinity ecosystem in an easy and open way, fulfilling the consumers needs. Furthermore, the combination of services from different domains together with privacy-respectful user-defined share of information, enables	7,499,007.50

								synergies among services from those domains and opens the door to a new market of domain-crossing services. VICINITY's approach will be demonstrated by a large-scale demonstration connecting 8 facilities in 7 different countries. The demonstration covers various domains including energy, building automation, health and transport. VICINITY's potential to create new, domain-crossing services will be demonstrated by value added services such as micro-trading of DSM capabilities, AI-driven optimization of smart urban districts and business intelligence over IoT. Open calls are envisioned in the project to integrate further, preferably public, IoT infrastructures and to deploy additional added value services. This will not only extend the scale of VICINITY demonstration, but also efficiently raise the awareness of industrial communities of VICINITY and its capabilities."	
52	INTER-IoT	Interoperability of Heterogeneous IoT Platforms	EU	IoT	H2020	AIOTI - IERC-European Research Cluster on the Internet of Things	http://www.inter-iot-project.eu/	"INTER-IoT project is aiming at the design, implementation and experimentation of an open cross-layer framework and associated methodology to provide voluntary interoperability among heterogeneous Internet of Things (IoT) platforms. The proposal will allow developing effectively and efficiently smart IoT applications, atop different heterogeneous IoT platforms, spanning single and/or multiple application domains. The overall goal of the INTER-IoT project is to provide a interoperable framework architecture for seamless integration of different IoT architectures present in different application domains. Interoperability will be provided at different levels: device, network, middleware, services and data. The two application domains and use cases addressed in the project and in which the IoT framework will be	7,329,850.00

								applied are m-health and port transportation and logistics. The project outcome may optimize different operations (e.g. increasing efficiency in transportation time; reducing CO2 emission in a port environment; improving access control and safety; improving remote patient attendance and increase the number of subject that surgery units can assist using the mobile devices with the same resources; reducing time spent in hospitals premises or reduce the time dedicated to the assistance activities carried out directly at the surgery with advantage for subjects in charge and also benefits those waiting, i.e. reduction of the waiting list) in the two addresses domains, but it may be extended to other application domains in which there is a need to interconnect different IoT architectures already deployed. The project may deal with interoperability at different layers."	
53	symbIoTe	Symbiosis of smart objects across IoT environments	EU	IoT	H2020	AIOTI - IERC-European Research Cluster on the Internet of Things	https://www.symbiote-h2020.eu/	Connected smart objects have invaded our everyday life across multiple domains, e.g. home withautomation solutions, assisted living with sensors and wearables to monitor personal activities, smart transportation and environmental monitoring. IoT is evolving around a plethora of vertically isolated platforms, each specifically suited to given scenarios and often adopting non-standard, sometimes fully proprietary, protocols to control the variety of sensors, actuators and communication elements. symbIoTe comes to evolve this fragmented environment and provides an abstraction layer for a unified control view on various IoT platforms and sensing/actuating resources. symbIoTe designs and develops an IoT orchestration middleware capable of unified and secure access to physical and virtualized IoT resources; hierarchical and orchestrated discovery and control across multiple IoT platforms; federation of IoT controllers and resources	7,104,827.50

								for cooperative sensing/actuation tasks; seamless roaming of smart objects across smart spaces. symbIoTe builds its orchestration middleware on top of existing standards for protocols and interfaces, plus a number IoT platforms both proprietary (i.e. developed by its industrial partners) and from open source (e.g. OpenIoT). This unique set of backgrounds and foreground can result in a significant step forward in horizontal integration and federation of IoT domains. Five use cases with real large scale deployments have been selected to validate our vision in representative smart spaces: home/residence, educational campus, stadium, mobility and yachting. Engagement with real users is key in our validation process. With its research, symbIoTe can enable innovative business models for a large set of stakeholders of the IoT value chain, and particularly SMEs and new entrants in the IoT market. The consortium includes direct beneficiaries of these impacts, including small and large industry with IoT business and renowned research performers.	
54	bIoTope	Building an IoT OPen innovation Ecosystem for connected smart objects	EU	IoT	H2020	AIOTI - IERC- European Research Cluster on the Internet of Things	http://biotope.cs.hut.fi/	"The Internet of Things (IoT) brings opportunities for creating new services and products, reducing costs for societies, increasing the service level for the citizens in a number of areas, and changing how services are sold and consumed. Despite these opportunities, current information system architectures create obstacles that must be addressed for leveraging the full potential of IoT. One of the most critical obstacles are the ‘vertical silos’ that shape today’s IoT because they constitute a serious impediment to the creation of cross-domain, cross-platform and cross-organisational applications and services. Those silos also hamper developers from producing new added value across multiple platforms due to the lack of interoperability and openness. bIoTope provides the necessary Standardized Open APIs	7,848,160.00

								for enabling horizontal interoperability between silos. Such horizontal interoperability makes it possible to develop Systems of Systems where cross-domain information from platforms, devices and other information sources can be accessed when and as needed. bIoTope-enabled Systems can seamlessly exploit all available information, which makes them smart in the sense that they can take or propose the most appropriate actions depending on the current User's or Object's Context/Situation, and even learn from experience. bIoTope capabilities lay the foundation for open innovation ecosystems where companies can innovate both by the creation of new software components for IoT ecosystems, as well as create new Platforms for Connected Smart Objects with minimal investment. Large-scale pilots implemented in smart cities will provide both social, technical and business proofs-of-concept for such IoT ecosystems. This is feasible because the bIoTope consortium combines unique IoT experience, commercial solution providers and end-users, thus ensuring the high quality and efficiency of the results and implementations."	
55	Be-IoT	The business engine for IoT pilots: Turning the Internet of things in Europe into an economic ally successful	EU	IoT	H2020	AIOTI - IERC- European Research Cluster on the Internet of Things	not available	"The project IoTBe will be proposed as a support action run by the consortium with the partners etventure, Fraunhofer IAIS, Faubourg Numérique, Acreo Swedish ICT and Challengy Israel. Our vision is to build a broad and vibrant ecosystem around the pilot projects that increase the collaboration between them, generates economic impact through new innovative business models and creates trust in the internet of things by transparent information about social challenges such as privacy and security implications. We will reach this vision by focussing on following five objectives: - We will support the collaboration and knowledge	1,041,073.10

		and socially accepted vibrant ecosystem						exchange between pilots and other relevant EU-projects (e.g. FIWARE). - We will build the bridge between pilots and relevant stakeholders (e.g. potential customers such as European small and medium sized enterprises (SMEs), entrepreneurs and developers, but also researchers and policy makers) and thus expand the ecosystem further. - We will set the ground for upcoming business building activities by creating awareness and also by facilitating and fostering societal acceptance (e.g. by running a variety of innovation activities). - Building on the above mentioned developments we will set the ground for the development of concrete new customer-oriented businesses based on the emerging pilots. Those business models will be derived through a proven systematic user centric ideation and validation process increasing the market acceptance and success rate of these business models. The emerging new businesses shall have a substantial economic impact in Europe. - We will pave the way for future advances within IoT: We will write a whitepaper that explains, describes and collects the most promising future standards from pilots and distribute these to relevant parties. We will use the consortium existing academies for education activities (e.g. Fraunhofer academy or etventure Berlin School of Digital Business)."	
56	WAZIUP	Open Innovation Platform for IoT-Big Data in Sub-Sahara	EU	IoT	H2020	AIOTI - IERC-European Research Cluster on the Internet of Things	not available	"The WAZIUP project, namely the Open Innovation Platform for IoT-Big Data in Sub-Saharan Africa is a collaborative research project using cutting edge technology applying IoT and Big Data to improve the working conditions in the rural ecosystem of Sub-Saharan Africa. First, WAZIUP operates by involving farmers and breeders in order to define the platform specifications in focused validation cases. Second, while	2,799,662.50

		Africa						tackling challenges which are specific to the rural ecosystem, it also engages the flourishing ICT ecosystem in those countries by fostering new tools and good practices, entrepreneurship and start-ups. Aimed at boosting the ICT sector, WAZIUP proposes solutions aiming at long term sustainability. The consortium of WAZIUP involves 7 partners from 4 African countries and partners from 5 EU countries combining business developers, technology experts and local Africa companies operating in agriculture and ICT. The project involves also regional hubs with the aim to promote the results to the widest base in the region."	
57	ARMOUR	Large-Scale Experiments of IoT Security Trust	EU	IoT	H2020	AIOTI - IERC-European Research Cluster on the Internet of Things	https://team.inria.fr/eva/h2020-armour/	"The Internet-of-Things (IoT) is rapidly heading for large scale meaning that all mechanisms and features for the future IoT need to be especially designed and duly tested/certified for large-scale conditions. Also, Security, Privacy and Trust are critical elements of the IoT where inadequacy of these is a barrier to the deployment of IoT systems and to broad adoption of IoT technologies. Suitable duly tested solutions are then needed to cope with security, privacy and safety in the large scale IoT. Interestingly, world-class European research on IoT Security & Trust exists in European companies (especially SME) and academia where even there are available technologies that were proven to work adequately in the lab and/or small-scale pilots. More, unique experimental IoT facilities exist in the EU FIRE initiative that make possible large-scale experimentally-driven research but that are not well equipped to support IoT Security & Trust experiments. But notably, Europe is a leader in IoT Security & Trust testing solutions (e.g. RASEN toolbox, ETSI Security TC, etc.) that can be extended to large-scale testing environments and be integrated in FIRE IoT testbeds for supporting experimentations.	1,999,558.75

								The ARMOUR project is aimed at providing duly tested, benchmarked and certified Security & Trust technological solutions for large-scale IoT using upgraded FIRE large-scale IoT/Cloud testbeds properly-equipped for Security & Trust experimentations. To this, ARMOUR will: (1) Enhance two outstanding FIRE testbeds (> 2700nodes; ~500users) with the ARMOUR experimentation toolbox for enabling large-scale IoT Security & Trust experiments; (2) Deliver six properly experimented, suitably validated and duly benchmarked methods and technologies for enabling Security & Trust in the large-scale IoT; and (3) Define a framework to support the design of Secure & Trusted IoT applications as well as establishing a certification scheme for setting confidence on Security & Trust IoT solutions."	
58	BIG IoT	Bridging the Interoperability Gap of the Internet of Things	EU	IoT	H2020	AIOTI - IERC-European Research Cluster on the Internet of Things	http://big-iot.eu/	"The objective of the BIG IoT project is to ignite really vibrant Internet of Things (IoT) ecosystems. We will achieve this by bridging the current interoperability gap between the vertically integrated IoT platforms and by creating marketplaces for IoT services and applications. Despite various research and innovation projects working on the Internet of Things, no broadly accepted professional IoT ecosystems exist. The reason for that are high market entry barriers for developers and service providers due to a fragmentation of IoT platforms. The goal of this project is to overcome these hurdles by Bridging the Interoperability Gap of the IoT and by creating marketplaces for service and application providers as well as platform operators. We will address the interoperability gap by defining a generic, unified Web API for smart object platforms, called the BIG IoT API. The establishment of a marketplace where platform, application, and service providers can monetize their assets will introduce an incentive to	7,999,882.50

								grant access to formerly closed systems and lower market entry barriers for developers. The BIG IoT consortium is well suited to reach the outlined goals, as it comprises all roles of an IoT ecosystem: resource providers (e.g., SIEMENS, SEAT), service and application developers (e.g., VODAFONE, VMZ), marketplace providers (e.g., ATOS), platform providers (e.g., BOSCH, CSI, ECONAIS), as well as end users connected through the public private partnerships of WAG and CSI or the user-focused information services that VMZ provides for the city of Berlin. The major industry players cover multiple domains, including mobility, automotive, telecommunications, and IT services. Four university departments will help to transfer the state of the art into the state of the practice and solve the open research challenges. This consortium will mobilise the necessary critical mass at European level to achieve the goals and to reach the ireach the impacts set out in this project."	
59	SMARTIE	Secure and sMArter ciTIes data management	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	http://www.smartie-project.eu/	The Internet of the Future will be an essential part of the knowledge society and will provide new information-based business. The usage of the Internet of Things for large-scale, partially mission-critical systems creates the need to address trust and security functions adequately. The vision of SMARTIE (Secure and sMArterciTIEs data management) is to create a distributed framework for IoT based applications sharing large volumes of heterogeneous information. This framework is envisioned to enable end-to-end security and trust in information delivery for decision-making purposes following data owner's privacy requirements. New challenges identified for privacy, trust and reliability are: • Provide trust and quality-of-information in shared information models to enable re-used across many applications.• Provide secure	3,286,000.00

								exchange of data between IoT devices and consumers of their information.• Provide protection mechanisms for vulnerable devices. SMARTIE will address these challenges within the context of Smart Cities. A smart city controller handling data for the city must show that the information collected from different devices are communicated and stored in a secure way. Privacy-protection and access control to the data and objects is necessary to convince data owners to share information and to protect the city infrastructure. SMARTIE envisions a data-centric paradigm with the 'information management and services' plane as a unifying umbrella, which will operate above heterogeneous network devices and data sources and will provide advanced secure information services.The feasibility and utility of SMARTIE will be tested in real environments with real users of the city infrastructures. The two application areas Transport and Energy will be considered; both are key infrastructures of cities. The tests will involve the cities Frankfurt an der Oder (Germany), Belgrade (Serbia) and Murcia (Spain).	
60	SMART-ACTION	Innovating Beyond IoT	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	https://www.smart-action.eu	With the increasing miniaturization of communication devices and the proliferation of technologies such as smart phones or sensor networks, the concept of an Internet of Things (IoT) is becoming a reality, especially with their integration into Smart Cities. The interdisciplinary nature of IoT work requires understanding, coordinating, supporting and engaging not only ICT, but also other areas such as biotechnology or nanotechnologies, that provide the right context in which IoT concepts can be embedded to provide solutions that can benefit society at large. SMART-ACTION will support the development of strategic research agendas and serve as an enabler for the dissemination and further integration of results into	

								future research and industrial developments, while coordinating international efforts. The main objective of SMART-ACTION is to provide support for International Road Mapping of the Research Agenda and Policies of IoT as part of the Future Internet Technological Development and to increase the uptake of European industry and businesses related to IoT, serving the Lisbon Agenda. SMART-ACTION has several strategic goals:(i) To consolidate the current research and technology transfer activities in Europe and to make sure the efforts performed by individual research groups generate an optimal effect on others research areas. (ii) To extend the reach of IoT technologies beyond Europe and to make sure that Asia, the US, Australia and Africa are all involved in the process of defining an overall strategy for the Internet of Things by providing input to workshops organised as part of SMART-ACTION. (iii) To make sure that companies are also involved in the process to make sure that roadblocks and potential problems are identified and solved as early as possible. We will do this with a keen eye on IoT policy aspects that emerge in the context of a number of relevant policy areas by keeping track of these developments and highlighting IoT aspects that may need to be considered in those policy areas. Following this strategy, we are confident that paths for the transfer of technology and, in general, the creation of new business in areas related to the Internet of Things will be encouraged.	

61	FITMAN	Future Internet Technologies for MANufacturing	EU	IoT	FP7	AIOTI - IERC-European Research Cluster on the Internet of Things	http://www.fitman-fi.eu/	According to the 2010 EC Competitiveness Report, Manufacturing is still the driving force of Europe's economy, contributing over € 6553 billion in GDP and providing more than 30 million jobs. It covers more than 25 different industrial sectors, largely dominated by SMEs, and generates annually over € 1535 billion (42%) worth of value added services. The mission of the FITMAN (Future Internet Technologies for MANufacturing industries) project is to provide the FI PPP with a set of industry-led use case trials in the Smart, Digital and Virtual Factories of the Future domains, in order to test and assess the suitability, openness and flexibility of FI-WARE Generic Enablers, this way contributing to the social-technological-economical-environmental-political sustainability of EU Manufacturing Industries. In order to accomplish the mission statement, the FITMAN project will deliver: • One FITMAN Generic Platform for Manufacturing Industries, as a collection of several Generic Enablers Implementations belonging to most of the identified technological Chapters of FI-WARE project; • One generic and flexible Trials Verification and Validation Framework, encompassing concepts, methods and tools for a technical and business assessment of the eleven Trials • One open-to-all FITMAN Phase III Package, to support FI-WARE PPP Phase III objective 1.8, Expansion of Use Cases, by providing access to FITMAN Reports and Prototypes for Phase III preparation and implementation • Three FITMAN Specific Platforms for Smart, Digital and Virtual Factories, as a collection of several Specific Enablers Implementations belonging to the background of FITMAN beneficiaries and specifically derived from previous RTD projects in the Factories of the Future and Future Internet Enterprise Systems research • Eleven FITMAN Trials Platforms as	12,890,000.00

								instantiation of the selected Generic and Specific Enablers for 11 industry-driven multi-sectorial Trials • Eleven FITMAN Trial Experimentations by deploying the FITMAN Trials Platforms in realistic Smart-Digital-Virtual Factories IT and business cases, as well as assess and evaluate the achieved results: i. Smart Factories Trials: TRW (LE) automotive supplier – Safe & Healthy Workplace, PIACENZA (SME) textile/clothing – Cloud Manufacturing, COMPLUS (SME) LED smart lighting – Collaborative Production, WHIRLPOOL (LE) white goods manufacturer – Mobile workforce. ii. Digital Factories Trials: VOLKSWAGEN (LE) automotive manufacturer – PLM ramp-up for reduced Time to Market , AGUSTAWESTLAND (LE) aeronautics manufacturer – Training services for blue collar workers, CONSULGAL (SME) construction – As-designed vs. As-built Interoperability, AIDIMA (SME) furniture – Mass Customised Production. iii. Virtual Factories Trials: APR (SME) plastic industry – Collaboration valorisation, TANet (SME) manufacturing resource management – Networked Business Innovation, GEOLOC (SME) Machinery for wood industry – Project-based Collaboration.	
62	CuteLoop	Customer in the Loop: Using Networked Devices enabled Intelligence for Proactive Customers	EU	IoT	FP7	AIOTI - IERC- European Research Cluster on the Internet of Things	www.cuteloop.eu	The strategic objective of CuteLoop is to explore how Intelligent Networked Devices such as enhanced RFID-based systems, can be used to effectively 'integrate customers within an Integrated Enterprise' and with this provide an important step towards 'real' Integrated, Real Time Enterprise. Such integrated real time enterprise, having customers as integrated drivers, needs, on one side, highly flexible and dynamic business interconnections to react dynamically and agile, on the other side a highly intensive and just-in-time exchange of knowledge/experience among Large Enterprises (LEs), SMEs and customers. CuteLoop intends to explore how	2,495,438.00

		Integration as Drivers of Integrated Enterprise						to radically improve the interaction of diverse actors in the integrated enterprise, specifically including customers as an integral part of these complex relationships, while focusing on the usage of 'Networked Devices Enabled Intelligence' to realise distributed and autonomous control of business processes. CuteLoop is aiming at realisation of a holistic approach on: - an innovative architecture by integration of event-driven and SOA based principles, - intelligent and agile agents combined with an event-driven architecture, - decentralised and self-evolving approach for assuring security and trust as well as supporting a customer oriented privacy of data - new interaction models and patterns for the real time enterprise. CuteLoop intends to address these problems in an SME driven integrated enterprise scenario (focusing specifically upon small and micro enterprises), which is, due to its high complexity and requested flexibility, the most critical scenario from both technical and organisational/business points of view. CuteLoop is focusing on small and micro enterprises in two sectors: construction and food industry.	
63	BUTLER	uBiquitous, secUre inTernet-of-things with Location and contExt-awaReness	EU	IoT	FP7		www.iot-butler.eu	Recent ICT advances are bringing to reality a world where sensors, actuators and smart portable devices are interconnected into an Internet-of-Things (IoT) ecosystem reaching 50 Billion devices by 2015.The IoT major challenges are, from a systemic viewpoint, smart resource management and digital security; and from a user/service perspective, the pervasiveness (uniformity of performance anytime and anywhere) and awareness (inversely proportional to the degree of knowledge required from users).BUTLER will be the first European project to emphasise pervasiveness, context-awareness and security for IoT. Through a consortium of leading Industrial, Corporate R&D and Academic partners with	9,708,000.00

								extensive and complementary know-how, BUTLER will integrate current and develop new technologies to form a 'bundle' of applications, platform features and services that will bring IoT to life.For this purpose, BUTLER will focus on:a)\tImproving/creating enabling technologies to implement a well-defined vision of secure, pervasive and context-aware IoT, where links are inherently secure (from PHY to APP layers) applications cut across different scenarios (Home, Office, Transportation, Health, etc.), and the network reactions to users are adjusted to their needs (learned and monitored in real time).b)\tIntegrating/developing a new flexible smartDevice-centric network architecture where platforms (devices) function according to three well-defined categories: smartObject (sensors, actuators, gateways), smartMobile (user's personal device) and smartServers (providers of contents and services), interconnected over IPv6.c)\tBuilding a series of field trials, which progressively integrate and enhance state-of-the-art technologies to showcase BUTLER's secure, pervasive and context-aware vision of IoT.In addition to these R&D innovations, BUTLER and its External Members Group will also aggregate and lead the European effort in the standardisation and exploitation of IoT technologies.	
64	PPP Future Internet (FIWARE projects)		EU	IoT	FP7		https://www.fi-ppp.eu/	The Future Internet Public-Private Partnership (FI-PPP) is a European programme for Internet innovation. It is aimed at accelerating the development and adoption of Future Internet technologies in Europe, advancing the European market for smart infrastructures and increasing the effectiveness of business processes through the Internet. It follows an industry-driven, user-oriented approach that combines R&D on network and communication technologies, devices, software, service and media	270,000,000.00

								technologies; and their experimentation and validation in real application contexts. It brings together the demand and supply sides, and it involves users early into the research lifecycle. The platform technologies will be used and validated by many actors, in particular by small and medium-sized companies and public administrations. Thus, the effectiveness of business processes and infrastructures supporting applications in areas such as transport, health or energy can be increased; leading to the creation of innovative business models that strengthen the competitive position of the European industry in sectors such as telecommunication, mobile devices, software and services, and content provision and media. The whole process has been divided into three stages.	
65	IoT Large Scale Pilots	Smart living environments for ageing well	EU	IoT	H2020	Large-scale pilot project	https://ec.europa.eu/digital-single-market/en/news/horizon-2020-work-programme-2016-2017-internet-things-large-scale-pilots	LSPs are goal driven initiatives that will propose Internet of Things(IoT) approaches to specific real-life industrial/societal challenges addressed in the IoT Focus Area of the Horizon 2020 Work Program 2016-2017. The areas include smart living environments for ageing well, smart farming and food security, wearables for smart ecosystems, reference zones in EU cities and Autonomous vehicles in a connected environment. These Large Scale Pilots seek to involve stakeholders from supply side to demand side, and contain all the technological and innovation elements and tasks related to the use, application and deployment; the development of the technology, testing and integration activities will also be part of the proposed LSPs. The IoT LSPs are also accompanied by coordination and support actions to ensure the smooth and efficient cooperation and management of the various activities of the Focus Area as well as to support cross fertilisation of the various pilots for technological and validation issues	20,000,000.00

								of common interest across the various use cases. In order to strengthen IoT Research and Innovation, an additional call is foreseen in 2017; this call is focused on sophisticated platform architectures for smart objects, embedded intelligence, and smart networks. Along with this additional call international joint IoT calls with Japan, South Korea, China and Brazil are associated to the IoT Focus Area in the Work Programme 2016-17. More information on the additional calls on IoT will follow.	
66	IoT Large Scale Pilots	Smart Farming and Food Security	EU	IoT	H2020	Large-scale pilot project	https://ec.europa.eu/digital-single-market/en/news/horizon-2020-work-programme-2016-2017-internet-things-large-scale-pilots	LSPs are goal driven initiatives that will propose Internet of Things(IoT) approaches to specific real-life industrial/societal challenges addressed in the IoT Focus Area of the Horizon 2020 Work Program 2016-2017. The areas include smart living environments for ageing well, smart farming and food security, wearables for smart ecosystems, reference zones in EU cities and Autonomous vehicles in a connected environment. These Large Scale Pilots seek to involve stakeholders from supply side to demand side, and contain all the technological and innovation elements and tasks related to the use, application and deployment; the development of the technology, testing and integration activities will also be part of the proposed LSPs. The IoT LSPs are also accompanied by coordination and support actions to ensure the smooth and efficient cooperation and management of the various activities of the Focus Area as well as to support cross fertilisation of the various pilots for technological and validation issues of common interest across the various use cases. In order to strengthen IoT Research and Innovation, an additional call is foreseen in 2017; this call is focused on sophisticated platform architectures for smart objects, embedded intelligence, and smart networks. Along with	30,000,000.00

								this additional call international joint IoT calls with Japan, South Korea, China and Brazil are associated to the IoT Focus Area in the Work Programme 2016-17. More information on the additional calls on IoT will follow.	
67	IoT Large Scale Pilots	Wearables for smart ecosystems	EU	IoT	H2020	Large-scale pilot project	https://ec.europa.eu/digital-single-market/en/news/horizon-2020-work-programme-2016-2017-internet-things-large-scale-pilots	LSPs are goal driven initiatives that will propose Internet of Things(IoT) approaches to specific real-life industrial/societal challenges addressed in the IoT Focus Area of the Horizon 2020 Work Program 2016-2017. The areas include smart living environments for ageing well, smart farming and food security, wearables for smart ecosystems, reference zones in EU cities and Autonomous vehicles in a connected environment. These Large Scale Pilots seek to involve stakeholders from supply side to demand side, and contain all the technological and innovation elements and tasks related to the use, application and deployment; the development of the technology, testing and integration activities will also be part of the proposed LSPs. The IoT LSPs are also accompanied by coordination and support actions to ensure the smooth and efficient cooperation and management of the various activities of the Focus Area as well as to support cross fertilisation of the various pilots for technological and validation issues of common interest across the various use cases. In order to strengthen IoT Research and Innovation, an additional call is foreseen in 2017; this call is focused on sophisticated platform architectures for smart objects, embedded intelligence, and smart networks. Along with this additional call international joint IoT calls with Japan, South Korea, China and Brazil are associated to the IoT Focus Area in the Work Programme 2016-17. More information on the additional calls on IoT will	15,000,000.00

								follow.	
68	IoT Large Scale Pilots	Reference zones in EU cities	EU	IoT	H2020	Large-scale pilot project	https://ec.europa.eu/digital-single-market/en/news/horizon-2020-work-programme-2016-2017-internet-things-large-scale-pilots	LSPs are goal driven initiatives that will propose Internet of Things(IoT) approaches to specific real-life industrial/societal challenges addressed in the IoT Focus Area of the Horizon 2020 Work Program 2016-2017. The areas include smart living environments for ageing well, smart farming and food security, wearables for smart ecosystems, reference zones in EU cities and Autonomous vehicles in a connected environment. These Large Scale Pilots seek to involve stakeholders from supply side to demand side, and contain all the technological and innovation elements and tasks related to the use, application and deployment; the development of the technology, testing and integration activities will also be part of the proposed LSPs. The IoT LSPs are also accompanied by coordination and support actions to ensure the smooth and efficient cooperation and management of the various activities of the Focus Area as well as to support cross fertilisation of the various pilots for technological and validation issues of common interest across the various use cases. In order to strengthen IoT Research and Innovation, an additional call is foreseen in 2017; this call is focused on sophisticated platform architectures for smart objects, embedded intelligence, and smart networks. Along with this additional call international joint IoT calls with Japan, South Korea, China and Brazil are associated to the IoT Focus Area in the Work Programme 2016-17. More information on the additional calls on IoT will follow.	15,000,000.00

69	IoT Large Scale Pilots	Autonomous vehicles in a connected environment	EU	IoT	H2020		https://ec.europa.eu/digital-single-market/en/news/horizon-2020-work-programme-2016-2017-internet-things-large-scale-pilots	LSPs are goal driven initiatives that will propose Internet of Things(IoT) approaches to specific real-life industrial/societal challenges addressed in the IoT Focus Area of the Horizon 2020 Work Program 2016-2017. The areas include smart living environments for ageing well, smart farming and food security, wearables for smart ecosystems, reference zones in EU cities and Autonomous vehicles in a connected environment. These Large Scale Pilots seek to involve stakeholders from supply side to demand side, and contain all the technological and innovation elements and tasks related to the use, application and deployment; the development of the technology, testing and integration activities will also be part of the proposed LSPs. The IoT LSPs are also accompanied by coordination and support actions to ensure the smooth and efficient cooperation and management of the various activities of the Focus Area as well as to support cross fertilisation of the various pilots for technological and validation issues of common interest across the various use cases. In order to strengthen IoT Research and Innovation, an additional call is foreseen in 2017; this call is focused on sophisticated platform architectures for smart objects, embedded intelligence, and smart networks. Along with this additional call international joint IoT calls with Japan, South Korea, China and Brazil are associated to the IoT Focus Area in the Work Programme 2016-17. More information on the additional calls on IoT will follow.	20,000,000.00

70	1543656	The Science of Activity-Predictive Cyber-Physical Systems (APCPS)	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1543656	This project aims to design algorithmic techniques to perform activity discovery, recognition, and prediction from sensor data. These techniques will form the foundation for the science of Activity- Prediction Cyber-Physical Systems, including potential improvement in the responsiveness and adaptiveness of the systems. The outcome of this work is also anticipated to have important implications in the specific application areas of health care and sustainability, two priority areas of societal importance. The first application will allow for health interventions to be provided that adapt to an individual's daily routine and operate in that person's everyday environment. The second application will offer concrete tools for building automation that improve sustainability without disrupting an individual's current or upcoming activities. The project investigators will leverage existing training programs to involve students from underrepresented groups in this research. Bi-annual tours and a museum exhibit will reach K-12 teachers, students and visitors, and ongoing commercialization efforts will ensure that the designed technologies are made available for the public to use. Deploying activity-predictive cyber-physical systems "in the wild" requires a number of robust computational components for activity learning, knowledge transfer, and human-in- the-loop computing that are introduced as part of this project. These components then create cyber physical systems that funnel information from a sensed environment (the physical setting as well as humans in the environment), to activity models in the cloud, to mobile device interfaces, to the smart grid, and then back to the environment. The proposed research centers on defining the science of activity-predictive cyber-physical systems, organized around the following	990,000.00

								thrusts: (1) the design of scalable and generalizable algorithms for activity discovery, recognition, and prediction; (2) the design of transfer learning methods to increase the the ability to generalize activity-predictive cyber-physical systems; (3) the design of human-in-the-loop computing methods to increase the sensitivity of activity-predictive cyber-physical systems; (4) the introduction of evaluation metrics for activity-predictive cyber-physical systems; and (5) transition of activity-predictive cyber-physical systems to practical applications including health monitoring/intervention and smart/sustainable cities.	
71	1329738	Integrated Sensing and Control Algorithms for Computer-Assisted Training	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1329738	This project will result in fundamental physical and algorithmic building blocks of a novel cyber-physical for a two-way communication platform between handlers and working dogs designed to enable accurate training and control in open environments (eg, disaster response, emergency medical intervention). Miniaturized sensor packages will be developed to enable non- or minimally-invasive monitoring of dogs' positions and physiology. Activity recognition algorithms will be developed to blend data from multiple sensors. The algorithms will dynamically determine position and behavior from time series of inertial and physiological measurements. Using contextual information about task performance, the algorithms will provide duty-cycling information to reduce sensor power consumption while increasing sensing specificity. The resulting technologies will be a platform for implementation of communication. Strong interactions among computer science, electrical engineering, and veterinary science support this project. Work at the interface between electrical engineering and computer science will enable increased power efficiency and specificity of sensing in the detectors;	926,000.00

								work at the interface of electrical engineering and veterinary behavior will enable novel physiological sensing packages to be developed which measure behavioral signals in real time; Project outcomes will enable significant advances in how humans interact with both cyber and physical agents, including getting clearer pictures of behavior through real time physiological monitoring. Students are part of the project and multidisciplinary training will help to provide development of the Cyber-Physical Systems pipeline. Project outreach efforts will include working with middle school children, especially women and under-represented minorities, presentations in public museums that will promote public engagement and appreciation of the contribution of cyber-physical systems to daily lives. The goal of each outreach activity is to encourage both interest and excitement for STEM topics, demonstrating how computer science and engineering can lead to effective and engaging cyber-physical systems.	
72	FORCES (1239166)	Foundations of Resilient CybEr-Physical Systems	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1239166	This NSF Cyber-Physical Systems (CPS) Frontiers project "Foundations Of Resilient CybEr-physical Systems (FORCES)" focuses on the resilient design of large-scale networked CPS systems that directly interface with humans. FORCES aims to provide comprehensive tools that allow the CPS designers and operators to combine resilient control (RC) algorithms with economic incentive (EI) schemes. Scientific Contributions The project is developing RC tools to withstand a wide-range of attacks and faults; learning and control algorithms which integrate human actions with spatio-temporal and hybrid dynamics of networked CPS systems; and model-based design to assure semantically	3,240,000.00

								consistent representations across all branches of the project. Operations of networked CPS systems naturally depend on the systemic social institutions and the individual deployment choices of the humans who use and operate them. The presence of incomplete and asymmetric information among these actors leads to a gap between the individually and socially optimal equilibrium resiliency levels. The project is developing EI schemes to reduce this gap. The core contributions of the FORCES team, which includes experts in control systems, game theory, and mechanism design, are the foundations for the co-design of RC and EI schemes and technological tools for implementing them. Expected Impacts Resilient CPS infrastructure is a critical National Asset. FORCES is contributing to the development of new Science of CPS by being the first project that integrates networked control with game theoretic tools and the economic incentives of human decision makers for resilient CPS design and operation. The FORCES integrated co-design philosophy is being validated on two CPS domains: electric power distribution and consumption, and transportation networks. These design prototypes are being tested in real world scenarios. The team's research efforts are being complemented by educational offerings on resilient CPS targeted to a large and diverse audience.	

73	1544887	Collaborative Research: Dynamic Methods of Traffic Control that Impact Quality of Life in Smart Cities	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1544887	In the recent past the term "Smart Cities" was introduced to mainly characterize the integration into our daily lives of the latest advancements in technology and information. Although there is no standardized definition of Smart Cities, what is certain is that it touches upon many different domains that affect a city's physical and social capital. Smart cities are intertwined with traffic control systems that use advanced infrastructures to mitigate congestion and improve safety. Traffic control management strategies have been largely focused on improving vehicular traffic flows on highways and freeways but arterials have not been used properly and pedestrians are mostly ignored. This work proposes to introduce a novel hierarchical adaptive controls paradigm to urban network traffic control that will adapt to changing movement and interaction behaviors from multiple entities (vehicles, public transport modes, bicyclists, and pedestrians). Such a paradigm will leverage several key ideas of cyber-physical systems to rapidly and automatically pin-point and respond to urban arterial congestion thereby improving travel time and reliability for all modes. Safety will also be improved since advanced warnings actuated by the proposed cyber-physical system will alert drivers to congested areas thereby allowing them to avoid these areas, or to adapt their driving habits. Such findings have a tangible effect on the well-being, productivity, and health of the traveling public. The primary goal is to create a Cyber-Control Network (CCN) that will integrate seamlessly across heterogeneous sensory data in order to create effective control schemes and actuation sequences. Accordingly, this project introduces a Cyber-Physical architecture that will then integrate: (i) a sub-network of heterogeneous sensors, (ii) a decision control substrate,	1,094,000.00

								and (iii) a sub-actuation network that carries out the decisions of the control substrate (traffic control signals, changeable message signs). This is a major departure from more prevalent centralized Supervisory Control And Data Acquisition (SCADA), in that the CCN will use a hierarchical architecture that will dynamically instantiate the sub-networks together to respond rapidly to changing cyber-physical interactions. Such an approach allows the cyber-physical system to adapt in real-time to salient traffic events occurring at different scales of time and space. The work will consequently introduce a ControlWare module to realize such dynamic sub-network reconfiguration and provide decision signal outputs to the actuation network. A secondary, complementary goal is to develop a heterogeneous sensor network to reliably and accurately monitor and identify salient arterial traffic events. Other impacts of the project include the integration of the activities with practitioners (e.g., traffic engineers), annual workshops/tutorials, and outreach to K-12 institutions.	
74	1330110	Distributed Sensing, Learning and Control in Dynamic Environments	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1330110	The objective of this project is to improve the performance of autonomous systems in dynamic environments, such as disaster recovery, by integrating perception, planning paradigms, learning, and databases. For the next generation of autonomous systems to be truly effective in terms of tangible performance improvements (e.g., long-term operations, complex and rapidly changing environments), a new level of intelligence must be attained. This project improves the state of robotic systems by enhancing their ability to coordinate activities (such as searching a disaster zone), recognize objects or people, account for uncertainty, and "most important" learn, so the system's performance is continuously improving. To do	

								this, the project takes an interdisciplinary approach to developing techniques in core areas and at the interface of perception, planning, learning, and databases to achieve robustness. This project seeks to significantly improve the performance of cyber-physical systems for time-critical applications such as disaster monitoring, search and rescue, autonomous navigation, and security and surveillance. It enables the development of techniques and tools to augment all decision making processes and applications which are characterized by continuously changing operating conditions, missions and environments. The project contributes to education and a diverse engineering workforce by training students at the University of California, Riverside, one of the most diverse research institutions in US and an accredited Hispanic Serving Institution. Instruction and research opportunities cross traditional disciplinary boundaries, and the project serves as the basis for undergraduate capstone design projects and a new graduate course. The software and testbeds from this project will be shared with the cyber-physical system research community, industry, and end users. The project plans to present focused workshops/tutorials at major IEEE and ACM conferences. The results will be broadly disseminated through the project website. For further information see the project website at: http://vislab.ucr.edu/RESEARCH/DSLC/DSLC.php	

75	SAM (1136141)	Making Cloud Computing Sense, Act, and Move	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1136141	This project takes the paradigm of cloud computing developed in the cyber-world and puts it into the physical world, to create a cyber-physical computing cloud, SAM-C. Unlike conventional cloud computing, SAM-C servers move in space, meaning, they are vehicles with physical constraints. The server vehicles also have sensors and actuators to create a way to re-organize mobile sensor networks in the paradigm of cloud computing. The project envisions an industry offering sensing as a service provided by the cloud. To enable this new service, the project extends the virtual machine instance fundamental to cloud computing with one new property -- virtual speed. The research terms this augmented entity a virtual vehicle and develops the theories, algorithms and protocols to realize many virtual vehicles over a network of real vehicle servers at scale. This research impacts the cloud computing industry by providing tools and theories it can use to leverage the many possible mobile server hosts in our society, e.g., cars, planes, people, and emerging vehicles like Unmanned Air Vehicles or drifters. It impacts mobile sensor networks by shifting them from an artifact built for an application to a service provided by a cloud. The inter-disciplinary research team spanning computer science, civil, and mechanical engineering impacts graduate and undergraduate teaching in systems, computer science, and control theory. The project guides K-12 students to build simple electric airplanes with sensors for greenhouse gas measurement thereby introducing young users of computation to cyber-physical cloud computing.	990,000.00

76	1239085	Correct-by-Design Control Software Synthesis for Highly Dynamic Systems	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1239085	This CPS Frontiers project addresses highly dynamic Cyber-Physical Systems (CPSs), understood as systems where a computing delay of a few milliseconds or an incorrectly computed response to a disturbance can lead to catastrophic consequences. Such is the case of cars losing traction when cornering at high speed, unmanned air vehicles performing critical maneuvers such as landing, or disaster and rescue response bipedal robots rushing through the rubble to collect information or save human lives. The preceding examples currently share a common element: the design of their control software is made possible by extensive experience, laborious testing and fine tuning of parameters, and yet, the resulting closed-loop system has no formal guarantees of meeting specifications. The vision of the project is to provide a methodology that allows for complex and dynamic CPSs to meet real-world requirements in an efficient and robust way through the formal synthesis of control software. The research is developing a formal framework for correct-by-construction control software synthesis for highly dynamic CPSs with broad applications to automotive safety systems, prostheses, exoskeletons, aerospace systems, manufacturing, and legged robotics. The design methodology developed here will improve the competitiveness of segments of industry that require a tight integration between hardware and highly advanced control software such as: automotive (dynamic stability and control), aerospace (UAVs), medical (prosthetics, orthotics, and exoskeleton design) and robotics (legged locomotion). To enhance the impact of these efforts, the PIs are developing interdisciplinary teaching materials to be made freely available and disseminating their work to a broad audience.	900,000.00

77	1136146	Design Science for CPS	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1136146	A CPS is a system in which computer-based (?cyber?) technology is combined with all kinds of physical systems, such as planes and robotic-surgeons. CPSs require integration (in industry and academia) of different types of knowledge from many different domains. CPSs are built from often inaccurate, undependable components, and operate in harsh and unpredictable environments. The cyber domain, interfaces, and the physical domain are tightly interwoven and networked (distributed) and hence cannot be designed and optimized individually. The goal of this project is to create a general CPS design-science that makes the design of every CPS simpler, faster, and more dependable, while at the same time reducing the cost and the required expertise level. This project gives rise to a unified theory that can allow for specification, modeling, design, optimization, and verification of CPSs on different levels of design abstraction and different steps of projection, even across boundaries between varied technologies. The project does bridge the gap between the continuous-time physical domain and the discrete timed cyber system. This project has a broad and profound impact in scientific, engineering, industrial, and academic communities. By enabling a fundamentally efficient design of CPSs, the most limiting bottleneck in design technology is eliminated, paving the way for many new applications and jobs with significant economic and social impact. This project contributes to the on-line educational endeavors currently underway, allowing cross education in different disciplines of complex CPS and speeding up development of new CPS programs in engineering and computer science.	1,350,000.00

78	1136174	Foundations of Secure Cyber Physical Systems	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1136174	Cyber-physical systems regulating critical infrastructures, such as electrical grids and water networks, are increasingly geographically distributed, necessitating communication between remote sensors, actuators and controllers. The combination of networked computational and physical subsystems leads to new security vulnerabilities that adversaries can exploit with devastating consequences. A synchronized attack on the interdependent network components and physical plants can create complex and new security vulnerabilities that cannot be addressed by securing the constituent systems individually. This project takes a holistic view by utilizing the properties of physical systems to design new secure protocols and architectures for cyber-physical systems (CPS) through a unified conceptual framework, which uses models for the physical system and the communication/computation network to define precise attack models and vulnerabilities. These mathematical models are used to design algorithms and protocols with provable operational security guarantees, thus enabling the design of more trustworthy architectures and components. The algorithms, protocols, and architectures are validated on CPS testbeds targeting building, automobile, and smart-grid applications. Additionally, the research is being integrated into the curriculum via the creation of novel coursework combining the underlying control, information theory, cryptography, and embedded system concepts. By improving the protection of critical cyber-physical infrastructure against emerging threats, this research is expected to provide direct socio-economic benefits, ranging from individual organizations to a national scale. The inter-disciplinary team of this project will integrate teaching and curriculum development with the	1,620,000.00

								research, contributing to the training of a new generation of engineers well versed in the design of trustworthy cyber-physical systems.	
79	VO-CyPhER (0931632)	CPS:VO: Virtual Organization for Cyber-Physical Research	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=0931632	This NSF award provides support for a CPS Virtual Organization. The National Science Foundation established the Cyber-Physical Systems (CPS) program with the vision of developing a scientific and engineering foundation for routinely building cyber-enabled engineered systems in which cyber capability is deeply embedded at all scales, yet which remain safe, secure, and dependable -- "systems you can bet your life on." The CPS challenge spans essentially every engineering domain. It requires the integration of knowledge and engineering principles across many computational and engineering research disciplines (computing, networking, control, human interaction, learning theory, as well as mechanical, chemical, biomedical, and other engineering disciplines) to develop a "new CPS system science." Achieving such an ambitious goal is challenging. The objective of the CPS "virtual organization" (CPS-VO) project is to actively build and support the multidisciplinary community needed to underpin this new research discipline and enable international and interagency collaboration on CPS. In support of the CPS-VO, Vanderbilt University will work with the community to develop strategies and mechanisms to: (i) facilitate and foster interaction and exchange among CPS researchers across a broad range of institutions, programs and disciplines, (ii) enable sharing of knowledge generated by CPS research with the broader engineering and scientific communities, sharing and integrating experimental tools, platforms and simulators among researchers and stakeholders, (iii) facilitate and foster collaboration and information	3,600,000.00

								exchange between CPS researchers and industry and (iv) facilitate international collaboration on CPS research.	
80	1035655	Science of Integration for Cyber-Physical Systems	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1035655	The objective of this research is to develop new foundations of composition in heterogeneous systems, to apply these foundations in a new generation of tools for system integration, and to validate the results in experiments using automotive and avionics System-of-Systems experimental platforms. The approach is to exploit simplification strategies: to develop theories, methods, and tools to assist in inter-layer decoupling. Intellectual merit. The research program has three focus areas: (1) theory of compositionality in heterogeneous systems, (2) tools and tool architectures for system integration, and (3) systems/experimental research. The project develops and deploys theories and methods for inter-layer decoupling that prevent or decrease the formation of intractable system-wide interdependences and maintain compositionality at each layer for carefully selected, essential system properties. Compositionality in tools is sought by exploring semantic foundations for model-based design. Systems/experimental research is conducted in collaboration with General Motors Global R&D (GM) and focuses on electric car platforms. Broader impact. The project is contributing to the cost effective development and deployment of many safety and security-critical cyber-physical systems, ranging from medical devices to transportation, to defense and avionics. The participating institutions seek to complement the 30-year-old conventional curriculum in systems science with one that admits computation as a primary concept. The curriculum changes will be aggressively promoted through a process of workshops and textbook preparation.	4,500,000.00

81	1035813	Center for Autonomous Transportation Systems	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1035813	The objective of this research is to study, develop and implement a comprehensive set of techniques that will eventually enable automobiles to be driven autonomously. The approach taken is to (a) address cyber-physical challenges of reliable, safe and timely operations inside the automobile, (b) tackle a range of physical conditions and uncertainties in the external environment, (c) enable real-time communications to and from the automobile to other vehicles and the infrastructure, and (d) study verification and validation technologies to ensure correct implementations. Intellectual Merits: The project seeks to make basic research contributions in the domains of safety-critical real-time fault-tolerant distributed cyber-physical platforms, end-to-end resource management, cooperative vehicular networks, cyber-physical system modeling and analysis tools, dynamic object detection/recognition, hybrid systems verification, safe dynamic behaviors under constantly changing operating conditions, and real-time perception and planning algorithms. Multiple intermediate capabilities in the form of active safety features will also be enabled. Broader Impacts: Automotive accidents result in about 40,000 fatalities and 3 million injuries every year in the USA. The global annual cost of road injuries is $518 billion. Many accidents are due to humans being distracted. Autonomous vehicles controlled by ever-vigilant cyber-physical systems can lead to significant declines in accidents, deaths and injuries. Autonomous vehicles can also offload driving chores from humans, and make time spent in automobiles more productive. Vehicular networks can help find the best possible routes to a destination in real-time. Broader impacts in this area are amplified by the project's partnerships with companies in the transportation and agricultural	2,160,000.00

								technology industries, and in information technology. Broader impacts are also sought through demonstrations and outreach to attract students into science and technology, and in particular to cyber-physical systems research.	
82	931843	ActionWebs	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=0931843	The objective of this research is to develop a theory of ActionWebs, that is, networked embedded sensor-rich systems, which are taskable for coordination of multiple decision-makers. The approach is to first identify models of ActionWebs using stochastic hybrid systems, an interlinking of continuous dynamical physical models with discrete state representations of interconnection and computation. Second, algorithms will be designed for tasking individual sensors, based on information objectives for the entire system. Third, algorithms for ActionWebs will be developed using multi-objective control methods for meeting safety and efficiency objectives. Two grand challenge applications for this research are in Intelligent Buildings for optimal heating, ventilation, air conditioning, and lighting based on occupant behavior and external environment; and Air Traffic Control for mobile vehicle platforms with sensor suites for environmental sensing to enable safe, convenient, and energy efficient routing. The intellectual merit of this research stems from a conceptual shift of ActionWebs away from passive information gathering to an action-orientation. This involves: modeling of ActionWebs using stochastic hybrid systems; taskable, multi-modal, and mobile sensor webs; and multi-scale action-perception hierarchies. The broader impact of the research is in two grand challenge national problems: energy efficient air transportation, and energy efficient, high productivity	4,500,000.00

								buildings, and will tackle social, privacy, economic, and usability issues. Integrated with the research is a program of coursework development in networked embedded systems, across stove pipes in EECS, Aero-Astro, Civil, and Mechanical Engineering departments. Outreach objectives include new course design at San Jose State University, and recruiting more women researchers.	
83	1035715	Assuring the Safety, Security and Reliability of Medical Device Cyber Physical Systems	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1035715	The objective of this research is to establish a new development paradigm that enables the effective design, implementation, and certification of medical device cyber-physical systems. The approach is to pursue the following research directions: 1) to support medical device interconnectivity and interoperability with network-enabled control; 2) to apply coordination between medical devices to support emerging clinical scenarios; 3) to ?close the loop? and enable feedback about the condition of the patient to the devices delivering therapy; and 4) to assure safety and effectiveness of interoperating medical devices. The intellectual merits of the project are 1) foundations for rigorous development, which include formalization of clinical scenarios, operational procedures, and architectures of medical device systems, as well as patient and caregiver modeling; 2) high-confidence software development for medical device systems that includes the safe and effective composition of clinical scenarios and devices into a dynamically assembled system; 3) validation and certification of medical device cyber-physical systems; and 4) education of the next-generation of medical device system developers who must be literate in both computational and physical aspects of devices. The broader impacts of the project will be achieved in three ways. Novel design methods and certification	4,500,000.00

								techniques will significantly improve patient safety. The introduction of closed-loop scenarios into clinical practice will reduce the burden that caregivers are currently facing and will have the potential of reducing the overall costs of health care. Finally, the educational efforts and outreach activities will increase awareness of careers in the area of medical device systems and help attract women and under-represented minorities to the field.	
84	1136027	Holistic Design Methodology for Automated Implementation of Human-in-the-Loop Cyber-Physical Systems	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1136027	This project develops a framework for design automation of cyber-physical systems to augment human interaction with complex systems that integrate across computational and physical environments. As a design driver, the project develops a Body/Brain Computer Interface (BBCI) for the population of functionally locked-in individuals, who are unable to interact with the physical world through movement and speech. The BBCI will enable communication with other humans through expressive language generation and interaction with the environment through robotic manipulators. Utilizing advances in system-level design, this project develops a holistic framework for design and implementation of heterogeneous human-in-the-loop cyber-physical systems composed of physically distributed, networked components. It will advance BBCI technology by incorporating context aware inference and learning of task-specific human intent estimation in applications involving semi-autonomous robotic actuators and an efficient wireless communication framework. The results of this project are expected to significantly speed up the design of complex cyber-physical systems. By accelerating the path from idea to prototype, this work shortens the time frame of and cost of	1,100,000.00

								development for assistive technology to improve the quality-of-life for functionally locked-in individuals. This project establishes an open prototyping platform and a design framework for rapid exploration of other novel human-in-the-loop applications. The open platform will foster undergraduate involvement in cyber-physical systems research, building confidence and expertise. In addition, new activities at the Museum of Science in Boston will engage visitors to experiment with systematic design principles in context of a brain computer interface application, while offering learning opportunities about basic brain functions.	
85	1239343	Cyber Enabled Manufacturing Systems (CeMs) for Small Lot Manufacture	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1239343	This grant provides funding for the development of Cyber Enabled Manufacturing (CeMs) process control for small lot manufacturing that incorporates a model of the process directly into the control algorithm. Such a model can be used to accommodate changes in the physical product and the manufacturing process and thus the manufacturing monitoring and control algorithm, so that changing conditions are easily accommodated without extensive additional experiments. A set of objectives of this physics and cyber-enabled manufacturing process control system are rational setting of manufacturing tolerances, real time prediction of manufacturing defects, real time control of process to eliminate defects, and real time monitoring and control for small lot manufacturing. The methodologies we propose to achieve these goals are high fidelity, physics based models including models of faults/defects, uncertainty quantification, reduced order models that run in real time, measurement, real time prediction, real time computer architecture, real time control with inverse solutions, and automating the CeMs process for generic manufacturing processes If successful, the results of this research will greatly	900,000.00

								reduce cycle time in producing new or modified products and improve the quality of manufacturing processes with accompanying reduction in waste, energy use, and cost. The development of such accurate control algorithms and their application to manufacturing processes can provide a competitive edge to US manufacturers. Perhaps more importantly, the education of engineers involved this research will supply US industry with employees who can apply this technology to many industrial processes.	
86	1135844	Information and Computation Hierarchy for Smart Grids	US	CPS	NSF	Division of Computer and Network Systems (CNS) - Cyber-physical Systems (CPS)		The electric grid in the United States has evolved over the past century from a series of small independent community-based systems to one of the largest and most complex cyber-physical systems today. However, the established conditions that made the electric grid an engineering marvel are being challenged by major changes, the most important being a worldwide effort to mitigate climate change by reducing carbon emissions. This research investigates key aspects of a computation and information foundation for future cyber-physical energy systems?the smart grids. The overall project objective is to support high penetrations of renewable energy sources, community based micro-grids, and the widespread use of electric cars and smart appliances. The research has three interconnected components that, collectively, address issues of computation architecture, information hierarchy, and experimental modeling and validation. On computation architecture, the framework based on cloud computing is investigated for the scalable, consistent, and secure operations of smart grids. The research aims to quantify fundamental design tradeoffs among scalability, data consistency, security, and trustworthiness for emerging applications of smart grids. On information hierarchy,	1,200,000.00

								temporal and spatial characteristics of information hierarchy are investigated with the goal of gaining a foundational understanding on how information should be partitioned, collected, distributed, compressed, and aggregated. The research also develops an open and scalable experimental platform (SmartGridLab) for empirical investigations and testing of algorithms and concepts developed in this project. SmartGridLab integrates the hardware testbed with a software simulator so that software virtual nodes can interact with physical nodes in the testbed. This research also includes a significant education component aimed at integrating frontier research with undergraduate and graduate curricula.	
87	1239478	CPS: Synergy: Achieving High-Resolution Situational Awareness in Ultra-Wide-Area Cyber-Physical Systems	US	CPS	NSF	Division of Information & Intelligent Systems (IIS) - Cyber-physical Systems (CPS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1239478	Energy infrastructure is a critical underpinning of modern society. To ensure its reliable operation, a nation-wide or continent-wide situational awareness system is essential to provide high-resolution understanding of the system dynamics such that proper actions can be taken in real-time in response to power system disturbances and to avoid cascading blackouts. The power grid represents a typical highly dynamic cyber-physical system (CPS). The ever-increasing complexity and scale in sensing and actuation, compounded by the limited knowledge of the accurate system state have resulted in major system failures, such as the massive power blackout of August 2003 and the most recent Arizona/California blackout of September 2011. Therefore, methods and tools for monitoring and control of these and other such dynamic systems at high resolution are vital to an emergent generation of tightly coupled, physically distributed CPS. This project employs the power grid as a target application and develops a high-resolution, ultra-wide-area situational awareness system that synergistically	

								integrates sensing, processing, and actuation. First, from the sensing perspective, high resolution is reflected in both measurement accuracy and potential for dense spatial coverage. Wide area, precise, synchronized, and affordable sensing in voltage angle and frequency measurements for large-scale observation is sorely needed to observe system disturbances and capture critical changes in the power grid. The crucial innovation of this work is to make accurate frequency measurement from low voltage distribution systems through the wide deployment of Frequency Disturbance Recorders (FDRs). Second, from a data processing perspective, high resolution is reflected in finer-scale data analysis to reveal hidden information. In practical CPS, events seldom occur in an isolated fashion; cascading events are more common and realistic. A new conceptual framework is presented in the study of event analysis, referred to as event unmixing, where real-world events are considered a mixture of more than one constituent root event. This concept is a key enabler for the analysis of events to go beyond what are immediately detectable in the system. The event formation process is interpreted from a linear mixing perspective and innovative sparsity-constrained unmixing algorithms are presented for multiple event separation and spatial-temporal localization. Third, to discover the high-level spatial-temporal correlation among root events in real time, a descriptive language is developed to discover patterns on the spatial and temporal information of root events. This descriptive language allows embedding pattern descriptions on the desirable and undesirable interactions between events in the system, which will then be compiled into distributed runtime constructs to be executed in deployed systems. Fourth, from the actuation	

								perspective, the system pushes the intelligence toward the lower level of the power grid allowing local devices to make decisions and to react quickly to contingencies based on the high-resolution understanding of the system state, enabling a more direct reconfiguration of the physical makeup of the grid. Finally, the methods and tools are implemented and validated on an existing wide-area power grid monitoring system, the North American frequency monitoring network (FNET). Escalating demands for electricity coupled with an outdated power transmission grid pose a serious threat to the US economy. The transformative nature of this research is to turn a large volume of real-time data into actionable information and help prevent potential outages from happening. The power grid is a typical example of dynamic cyber physical system. Providing high-resolution situational awareness for the power grid has a direct and immediate impact on this and other CPS. The research is coupled with a strong educational component including active recruitment of students from underrepresented groups supported by existing programs and broad dissemination of research findings.	
88	1564009	Handling a Trillion Unfixable Flaws on Billions of Internet-of-Things	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1564009	The Internet-of-Things (IoT) has quickly moved from concept to reality, with estimates that the number of deployed IoT devices will rise to 25 billion in 2020. However, studies show that many IoT devices have serious security vulnerabilities. Moreover, the limitations of IoT devices and scale of networks of IoT devices often make traditional IT security approaches impractical. This project takes a first-principles approach to rethinking network security and address these concerns for IoT networks: (1) scalable alternatives to simple perimeter defenses; (2) new methods to manage security in deployed networks of IoT devices; and (3)	1,100,000.00

								new security policies with sufficient generality to administer IoT devices and networks in many diverse use-cases. By providing a principled architecture to secure IoT deployments, the project will help realize the full potential societal benefits of IoT. The project will result in the development of novel open-source tools, modeling abstractions, algorithms, and prototype implementations that will reduce the time to deploy novel IoT security solutions, and make the results of the project available to the community. The project's PIs will engage in educational and outreach activities to train the next generation talent for the emerging area of IoT. In particular, the PIs plan to integrate the interdisciplinary research ideas into courses spanning security, networking, systems and cyber-physical systems. The project will also actively encourage participation from underrepresented groups. Finally, the tools and measurements generated by this project will inform and accelerate the industry transition to pervasive IoT deployments that are safe and secure.	
89	1407583	Networking Out of Thin Air	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1407583	This project proposes to lay the algorithmic foundations for ambient backscatter, a novel communication mechanism that transforms existing wireless signals (e.g., TV, cellular, and Wi-Fi transmissions) into both a source of power as well as a communication medium: thus, enabling communication without requiring additional bandwidth and/or power. Ambient backscatter networks have the potential to introduce a new practical model of communication; reduce world-wide battery usage and energy consumption of small computing devices; and reduce the human and economic costs of maintaining power. The proposed research, if successful, can open up ubiquitous communication applications in multiple	1,100,000.00

								domains including Internet-of-Things, wearable and mobile devices, as well as infrastructure-free localization. As small Internet-of-Things devices are increasingly embedded in objects and environments such as thermostats, books, furniture, and even implantable medical devices; a novel primitive like ambient backscatter would enable ubiquitous communication at unprecedented scales (without batteries or wires) and in location and times that were previously infeasible (day and night, indoors and outdoors). Further, since ambient backscatter consumes orders of magnitude lower power than traditional radio communication, it can enable always-on communication for battery-operated mobile devices such as smart phones and tablets as well as wearable devices such as head-on displays and smart watches. Finally, since with ambient backscatter, battery-free devices can communicate with each other, they are aware of their context and the devices around them; this enables missing-item localization in warehouses and medical facilities that are infeasible with existing RFID technologies, without an extensive deployment of power infrastructure. The proposed research will produce algorithms, designs, circuits, and system implementations that will enable communication and networking out of thin air. The key challenge in achieving this is that existing signals such as TV, cellular, and Wi-Fi transmissions already carry data and are fast-changing signals; transmitting additional information on top of these signals can be challenging. Further, since the proposed devices are powered using these ambient signals, circuit designs that can harvest energy from them efficiently are needed. This research will design algorithms for enabling communication without requiring additional bandwidth and/or power.	

								This project will also design a complete network stack for this new communication primitive that will enable multiple such devices to co-exist with each other as well as with legacy receivers and leverage algorithms and techniques to enable new functionalities such as battery-free localization and demonstrate their applications in a number of domains including inventory, and healthcare.	
90	1345305	Deployment-Driven Evaluation and Evolution of the eXpressive Internet Architecture	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1345305	The eXpressive Internet Architecture (XIA) project aims to design, evaluate, and realize a future network architecture that provides inherent trustworthiness, supports long-term evolution of network use models and network technology, and addresses adoption through careful design of APIs and reasoned analysis of interactions (or tussle points) between multiple stakeholders. XIA offers intrinsic security by offering each entity the ability to unilaterally validate that it is communicating with the correct counterparties. XIA admits new usage models to evolve through the use of flexible addresses. This architectural flexibility also exposes new features, for example, enabling users to trade-off maintaining anonymity against maximizing efficiency in content retrieval. To date, the XIA team has developed an initial architecture; built prototype implementations of the XIA data plane, network protocol stack, and secure unicast communication using Scalability, Control and Isolation On next-generation Networks (SCION); engaged in a wide range of basic networking and security research; and investigated questions raised by future Internet architecture research. In the next phase of research on XIA, principal investigators (PIs) advance the original vision of XIA by deepening the ongoing research thrusts, driven in part by research questions exposed during the initial	4,500,000.00

								investigation. The research agenda also includes an increased emphasis on control protocols based on a unified control plane architecture. The research will be driven through two network deployments that challenge the architectural framework: a vehicular network deployment in the city of Pittsburgh, and a video delivery environment spanning the U.S. These deployments also leverage and deepen the work on secure network operations, including providing a highly available infrastructure and secure authentication mechanisms. Finally, these deployments necessitate further research on building a robust XIA network, and establish best practices for using the XIA architecture, including support for mobility and a rich session layer. The research will continue to be informed by issues of deployability and economic viability, and to consider XIA as perceived by end users. The plan therefore includes research on governance, economic implications of high availability, and user studies that investigate the interplay between privacy, transparency and user control. Finally, the PIs will develop methods for performing a comparative evaluation of future Internet architectures. This evaluation methodology includes a definition of evaluation criteria, a discussion of the importance of direct and indirect evaluation, and evaluation using the XIA network environments and other use cases. Intellectual Merit: The intellectual merit of the proposed work is an ambitious study of the effectiveness of a new architectural framework, evaluated in the context of two challenging network environments. The cross-cutting expertise of the team enables broad and deep investigation, starting from specification and implementation of many key network and security building blocks, up to application-specific methods	

								delivering innovative solutions in the vehicular and video delivery environments. The evaluation spans multiple disciplines, ranging from low-level performance evaluation to reasoning about incentive-compatibility and economic benefits of the PIs' approach to a quantification of user experience. Broader Impacts: The project will establish a framework for evaluating network architectures that will be broadly applicable to FIA research. The PIs propose a variety of education and outreach activities to accompany the proposed research. Educational activities include new courses on clean-slate network design, and teaching and internship activities for under-represented and high school students. Broader outreach activity also includes a plan to involve and inform policymakers on the health of the current and future Internet led by a PI who formerly served at the FCC.	
91	1540066	Facilitating Free and Open Access to Information on the Internet	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1540066	This project develops methods to provide citizens information about technologies that obstruct, restrict, or tamper with their access to information. Internet users need an objective, independent, third-party service that helps them determine whether their Internet service provider or government is restricting access to content, specific protocols, or otherwise degrading service. Towards this goal, we are (1) monitoring attempts to block or manipulate Internet content and communications; and (2) evaluating various censorship circumvention mechanisms in real-world deployments}. The project develops a large-scale measurement and monitoring service that measures network reachability and performance from a variety of access networks to various Internet services; infers whether ISPs or governments are restricting or otherwise throttling access to various applications and services; and detects attempts to tamper with	1,100,000.00

								information presented to users. The project also studyies the policy ramifications of making information about censorship and information tampering available to Internet users. It will provide up-to-date information about both the extent of censorship and information tampering in countries around the world and technologies countries are using to implement censorship and thwart censorship circumvention tools. Discoveries are disseminated through real-time portals and through regular written reports and academic publications.	
92	1345300	The Next-Phase MobilityFirst Project - From Architecture and Protocol Design to Advanced Services and Trial Deployments	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1345300	The Next-Phase MobilityFirst (MF) project aims to have a major impact on the architecture of the future Internet by re-architecting it to address the needs of emerging mobile platforms and applications. Adoption of technologies arising from this project may be expected to provide improved efficiency, security and robustness that would benefit both network operators and end-users of the Internet. This project, originally funded as a collaborative research effort under the NSF Future Internet Architecture (FIA) program (2010-13) in which the MF architecture was designed over the past 3 years, is centered on a new name-based service layer which serves as the narrow-waist of the protocol; this name-based services layer makes it possible to build advanced mobility-centric services in a flexible manner while also improving security and privacy properties. The architecture incorporates novel storage-aware routing techniques which provide significant improvements in mobile network capacity and functionality. The next phase of the MobilityFirst project is aimed at making the transition from early-stage architecture and prototyping to advanced real-world services and trial network deployments. The research and experimental trials agenda is aimed at validating	1,575,000.00

								and refining the core name service, routing, security and management components of the MF architecture, while also responding to emerging trends in network technology and services such as the cellular mobile data explosion, the growth of content, the emergence of cloud computing, and software-defined network (SDN) technology. Intellectual Merit: This project includes several research thrusts aimed at transitioning the MobilityFirst architecture to advanced services and field deployable technology. These include: (1) advanced name-based network services and development of enhanced global name service (GNS) technology; (2) network security and privacy designs and enhancements; (3) design of advanced content services; (4) application of MobilityFirst protocols to next-generation mobile cloud computing; (5) design of advanced context-aware services; (6) technical and economic study of cellular-Internet convergence; (7) software-defined network (SDN) ready protocol design; and (8) technology platforms, router implementation and deployment strategies. These research thrusts will be informed by three distinct real-world network environment trials: a "mobile data services" trial with a wireless ISP (5Nines) in Madison; WI; a "content production and delivery network" trial involving several public broadcasting stations in Pennsylvania connected by a greenfield optical network called PennREN; and a "context-aware public service" weather emergency notification system (CASA) with end-users in the Dallas/Fort Worth area. These network environment trials are the centerpiece of the proposed project, and are expected to provide a firm basis for validation of the MobilityFirst protocol stack and its usefulness for developing advanced mobile, content, context and cloud applications, while	

								also advancing the technology to the field-deployment stage. Expected outcomes from the project include research results on security, privacy, content/context/cloud services and SDN; MobilityFirst protocol stack software revisions; router technology implementations; multiple real-world trial deployments of the technology; and experimentally supported evaluations of the architecture. This project is a collaborative effort involving Rutgers, UMass, MIT, Duke, U Michigan, U Wisconsin, and U Nebraska with the participation of several industrial research and network environment trial partners. Broader Impacts: The MobilityFirst project will have impact as a new approach to a future Internet that by design addresses mobility and mobile platforms, and as an enabler of new mobile Internet applications of social value such as context-aware emergency notification services. The release of open source protocol software may be expected to help to stimulate further experimental research on future Internet architectures across the networking community. The project also contributes to education and training in the key areas of Internet and mobile network technology.	
93	1228994	Detection and Analysis of Large-Scale Internet Infrastructure Outages	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1228994	Despite the Internet's status as critical infrastructure, there is little scientific instrumentation dedicated to monitoring global Internet behavior. In particular, we have no rigorous framework for measuring, analyzing, or quantifying the impact of network outages, filtering, or other abnormal connectivity dynamics on a global scale. This project applies successful research results in analyzing recent macroscopic Internet connectivity disruptions to the development, testing, and experimental deployment of an operational capability to detect, monitor, and characterize such large-scale infrastructure outages. The investigators are seeking to	1,350,000.00

								validate and extend a methodology for identifying not only which networks have been affected by an outage, but also which mechanisms have been used to effect a deliberate disruption. The two intellectual themes of the research are: (1) extracting signal from malware-induced background radiation in Internet traffic (IBR); and (2) combining multiple types of data (active probing, passive IBR, routing data, geolocation, and registry databases) to delineate the scope and progression of the outage. The project will also develop quantitative indicators to gauge the impact of geophysical disasters on Internet infrastructure, including the dynamics of loss and restoration of service. A transition of these research outcomes into practice will yield a system specification, implementation, and experimental operational deployment to detect and monitor global connectivity failures on a planetary scale. In addition to improving our understanding of how measurements yield insights into network behavior, and strengthening our ability to model large scale complex networks, such a system will illuminate infrastructure vulnerabilities that derive from architectural, topological, or economic constraints, suggesting how to mitigate or eliminate these weaknesses in future Internet architecture and measurement research.	
94	1414177	Mapping Interconnection in the Internet: Colocation, Connectivity and Congestio	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1414177	As the global Internet expands to satisfy the demands and expectations of an ever-increasing fraction of the world's population, profound changes are occurring in its interconnection structure, traffic dynamics, and the economic and political power of different players in the ecosystem. These changes not only impact network engineering and operations, but also present broader challenges for technology investment, future network design, public policy, and scientific study of the Internet itself. And yet, from both scientific and policy	1,100,000.00

		n						perspectives, the evolving ecosystem is largely uncharted territory. Two related transformations of the Internet ecosystem motivate this project: the emergence of Internet exchanges (IXes) as anchor points in the mesh of interconnection, and content providers and Content Delivery Networks (CDNs) as major sources of traffic flowing into the Internet. By some accounts over half the traffic volume in North America now comes from just two content distributors (Youtube and Netflix). This shift constitutes the rise of a new kind of hierarchy in the ecosystem, bringing new constraints on existing players who need to manage traffic on their networks to minimize congestion, resulting in tussles among commercial players as well as between the private sector and regulatory bodies, at the expense of users suffering degraded quality of experience. The goal of this project is to characterize the changing nature of the Internet's topology and traffic dynamics, and to investigate the implications of these changes on network science, architecture, operations, and public policy. The first task is to construct a new type of semantically rich Internet map, which will elucidate the role of IXes in facilitating robust and geographically diverse but complex interdomain connectivity. This map will guide the second task: a measurement study of traffic congestion dynamics induced by evolving peering and traffic management practices of CDNs and ISPs, including methods to detect and localize the congestion to specific points in the network. Data used will include measurements gathered using CAIDA's global active measurement infrastructure Archipelago (Ark), as well as data contributed by industry players for use in validation and refinement of methods. This project will advance the scientific community's understanding of the evolving Internet by developing and applying a more	

								sophisticated model of the physical Internet topology. The inquiry is structured to pursue new methods of analysis and insights into the interaction between CDN dynamics, interconnection at IXes, and network performance. Results, including the sharing of collected and curated data sets, will also inform related fields, including traffic modeling, infrastructure resiliency and protection, network economics, industrial organization, and future network architecture design. Results will be disseminated via publications, conferences, industry meetings, online lectures, web pages. CAIDA and CSAIL will jointly host annual international interdisciplinary workshops to close the loop between empirical research results, theoretical modeling activities, and policymaking. By providing unbiased data about the evolving Internet ecosystem, the proposed research has the potential to be of great value to Internet users, network operators, and governments.	
95	1534138	Smart CROwdsourced Urban Delivery (CROUD) System	US	IoT	NSF	Division of Industrial Innovation and Partnership (IIP)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1534138	This Partnerships for Innovation:Building Innovation Capacity (PFI:BIC) aims to develop a crowd-sourced urban delivery (CROUD) system that promises an efficient, greener, urban delivery service system. The rise of e-commerce is rapidly changing the landscape of retail business. By 2017, online sales will account for more than 10% of this $4.5 trillion industry, according to a 2014 U.S. Census Bureau report. To stay competitive in this growing market, retailers are under enormous pressure to quickly deliver the goods to their consumers at low prices. Large asset-based carriers (e.g., UPS and FedEx) are not particularly cost-efficient for express local deliveries in urban areas, essentially because their distribution networks are designed to transport through hubs rather than directly between customers and retailers. This inherent inefficiency is aggravated by mounting demand for same-day delivery--widely	900,000.00

								considered the Holy Grail for e-commerce at present--that requires more frequent dispatch and in turn increases transportation cost. Recent arrival of tech giants such as Amazon, Google and Uber in this battlefield attests to the tremendous opportunities and challenges in the urban delivery industry. Ironically, there exists a large volume of unused capacity in the transport system. For one thing, because the urban delivery industry is highly fragmented, consolidating its capacity is often difficult. Another underutilized capacity exists in millions of private passenger vehicles on the road with empty trunks that could be used for delivery. Addressing this lack of adequate coordination between demand and supply holds the key to solving the express urban delivery challenge. A promising solution, enabled by recent advances in wireless communication and mobile computing, is crowd-sourcing technology that has been successfully applied in passenger transport (e.g., Uber and Lyft). The overarching goal of this project is to develop and evaluate a CROwd-sourced Urban Delivery (CROUD) system. With CROUD, consumers will enjoy faster, cheaper and more reliable delivery service. The retail industry will see a stronger consumer demand. The delivery industry will improve its profitability and reduce its environmental impact, while employing a highly mobile and efficient workforce. The project will also educate next generation entrepreneurs by providing training opportunities and materials for upgrading curricula in multiple disciplines. The scientific contributions of the project span across several disciplines, and they are closely interconnected in supporting a human-centered smart CROUD system. First, the project will lead to pricing mechanisms that focus on matching consumers with couriers under varying market conditions. The use of matching models	

								is novel in pricing analysis for two-sided markets, such as thouse found in CROUD systems, and it promises insights in other markets. Second, the project will help understand and predict behaviors and choices of humans in a CROUD system, a field completely new to behavioral econometrists. New econometric models, calibrated with behavioral data collected in this project, will be built into management strategies to enhance the consumer-courier-technology interaction in the system. Third, the proposed research will develop smartphone-based motion detection methods that are uniquely suited to track and interpret the activities of couriers. Finally, the project will create computational tools to facilitate collaborative delivery among couriers and to optimize routing plans based on real-time information. In doing so, new formulations, algorithms and insights will be generated for challenging problems in operations research and computational economics such as network design with relays, real-time vehicle routing considering transfer and environmental impacts. This project will be led by Northwestern University through its Transportation Center (NUTC), with three researchers from two departments: Department of Civil and Environmental Engineering and Department of Managerial Economics and Decision Science. The primary industry partner is Zipments Inc. (small business, New York City, New York) and the other primary academic partner is the University of Illinois at Chicago (UIC) (non-profit research university, Chicago, Illinois). Two broader context partners will also join the team. They are Center for Commercialization of Innovative Transportation Technology (non-profit research institution operated within NUTC, Evanston, Illinois) and Center for Supply Chain and Logistics management (non-profit research institution affiliated	

								with UIC, Chicago, Illinois).	
96	1345318	Named Data Networking Next Phase (NDN-NP)	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1345318	Named Data Networking (NDN) is a Future Internet Architecture inspired by years of empirical research into network usage and a growing awareness of unsolved problems in IP. Its premise is that the Internet is primarily used as an information distribution network, which is not a good match for IP, and that the future Internet's "thin waist" should be based on named data rather than numerically addressed hosts. This proposal continues research started in 2010 under NSF's FIA program. It applies the principal investigators' (PIs) increasing understanding of NDN opportunities and challenges to two national priorities--Health IT and Cyber-physical Systems--to further evolve the architecture in the experimental manner that has proven successful. The research agenda is organized to translate key results in architecture and security into library code that guides application development towards native NDN designs. It simultaneously continues fundamental research into the challenges of global scalability and opportunities for innovation created by "simply" routing and forwarding data based on names. The NDN research agenda includes: 1) Applications--exploring naming and application design patterns; rendezvous, discovery and bootstrapping; the design of in-network storage; and use of new synchronization primitives; 2) Security & trustworthiness--providing basic building blocks of key management, trust management, and encryption-based access control, and anticipating future security challenges; 3) Routing and forwarding strategy--developing and evaluating path-vector, link-state, and hyperbolic options for inter-domain routing, addressing routing security and trust, and designing fast forwarding and mobility support; 4)	2,000,000.00

								Scalable forwarding--aiming to support real-world deployment, evaluation and adoption via an operational, scalable forwarding platform; 5) Libraries & tools--developing reference implementations based on the team's fundamental results; 6) Social & economic impacts--considering specific questions of the target environments and broader ones arising in a "World on NDN." The PIs choose Mobile Health and Enterprise Building Automation & Management Systems as specific environments to validate the architecture and drive new research. Domain experts will be 1) Open mHealth, a non-profit patient-centric ecosystem led by Deborah Estrin (Cornell) and Ida Sim (UCSF); and 2) UCLA Facilities Management, operators of the second largest Siemens building monitoring system on the West Coast. To guide research on the security dimensions of these environments and NDN more generally, the NDN team has convened a Security Advisory Council to complement its own effort. Intellectual Merit: NDN builds on lessons learned from the success of IP, preserving the thin waist, hierarchical naming, and end-to-end principles. The design recognizes the major shift in the applications communication model, from the "where" (i.e., the host/location) to the "what" (i.e., the content). Architecting a communications infrastructure around this shift can radically simplify application designs, enabling them to communicate directly using content names they desire and leaving it to the network to figure out how and from where to retrieve it. NDN also recognizes that the biggest weakness in the current Internet architecture is lack of security, and incorporates a fundamental building block to improve security by requiring that all content be signed. Broader Impacts: The success of new architectures	

								requires community involvement and uptake. NDN has built momentum through a commitment to an open source model that has spurred substantial research activity in both architecture and current implementation. Project members are often invited to present at "future Internet" meetings around the world, and the PIs have performed high-visibility demos of NDN's ability to handle large scale distribution. Industry is also showing increasing participation. Finally, NDN has significantly impacted students, generating several Ph.D. theses, related industry internships, and both graduate and undergraduate classes that can now present a comprehensive alternative to IP to stimulate discussion of what network architecture design really means.	
97	1531046	Sustainable Ecosystem of Smart Applications	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1531046	The economic and competitive value of widely available advanced networking has become well known, with availabity of next generation networking impacting factors as basic as homeowner property values and housing rents. However, deployment of next generation networking in U.S. cities and regions has lagged that of a number of other nations. This project assembles and jump-starts a growing and sustainable ecosystem of smart gigabit city testbeds which support applications delivering important new advances in healthcare, education, public safety, and other national priority areas. These advanced ecosystems leverage advanced Internet concepts developed by prior NSF research programs that are not yet available on today's commercial Internet. By leveraging existing multi-gigabit links to interconnect the cities, and by providing city-based interoperable application infrastructure the project allows entrepreneurs and academics in one city to write and trial visually-compelling and ultra-responsive new Internet applications that can be replicated into other cities and regions. An important	1,850,000.00

								key will be the involvement of citizens and community organizations in building and experimenting with advanced networking applications addressing national priorities. The project will: 1) Mobilize and interconnect 12 to 15 cities and regions as smart gigabit city testbeds and connect to a small number of other volunteers from the U.S. UCAN community anchor community; 2) leverage GENI technology and deploy a new community-focused computing infrastructure; support the develop and adapt national priority and smart city applications that uniquely leverage this infrastructure and can be replicated in the other cities and regions; 4) Identify novel unsolved problems that can help define areas for future Internet research and 5) share information among ecosystem participants. To maximize the network effects, the project will utilize interoperable advanced wired and wireless applications platforms, interoperable federation, authentication and authorization, and built-in security measures and measurement/logging. This project builds the initial infrastructure upon the distributed infrastructure pioneered and deployed by the GENI and prior US Ignite projects in the academic or civic networking infrastructures of 56 cities in the United States.	
98	1012798	Towards Trustworthy Interactions in the Cloud	US	IoT	NSF	Division of Computer and Network Systems (CNS)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1012798	As one of the most promising emerging concepts in Information Technology, outsourced computation (also known as cloud computing) is transforming our perception of how IT is consumed and managed, yielding improved cost efficiencies and delivering flexible, on-demand scalability. Cloud computing reduces IT resources and services to commodities acquired and paid-for on-demand through a fast-growing set of infrastructure, platform, and service providers.	1,350.00

								Despite the relatively fast growth and increased adoption of clouds, our understanding of aspects related to their security, privacy, and economic value proposition -- and hence our ability to trust them -- is lacking. This project addresses this challenge by (a) extending cloud service-level agreements to govern aspects such as integrity of outsourced services, information leakage control, and fair market pricing; (b) developing mechanisms that allow providers to guarantee and users to verify that such trust-enhancing service-level agreements are being followed; and (c) exposing trustworthiness guarantees and tradeoffs to cloud customers and system integrators in ways that are both practical and usable. The research work pursued in this project is timely as it addresses the issues of cloud trustworthiness early enough to avoid having the conflicts among its various stakeholders develop unchecked -- as was the case with the Internet decades ago. Doing so has the potential of improving the utility and hardness of our cyber-infrastructure, with significant benefit to our economy and society. The project will ultimately lead to a better marketplace for computing resources, in which users are assured that the services they acquire satisfy their performance, security, and privacy expectations.	
99	964639	Security and Privacy Preserving Data Mining and Management for	US	IoT	NSF	Division of Industrial Innovation and Partnership (IIP)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=0964639	A fundamental but challenging issue in information security is secure sharing and management of sensitive data and information among numerous organizations that form large-scale e-enterprises. Today, an increasing number of enterprises are using the Internet for managing and sharing users? and enterprise information through online databases. However, security and privacy of data is an overriding concern currently limiting the proliferation of information technology.	1,100,000.00

		Distributed Domains						Among others, the two key security issues are; (a) interoperability challenge of diverse security policies exercised by collaborating organizations for sharing sensitive information and; (b) using collaborative knowledge for detecting and responding to any emerging security or emergency threats. The objective of this project to develop a scalable multi-domain information security framework which: (a) facilitates large-scale integration, mining, and analysis of multimedia data residing in multiple domains for identifying sensitive information that require controlled dissemination (b) allows distributed real-time content analysis of multimedia data streams across multiple-domains for correlating detectable events, (c)l provides integration and evolution of inter-domain access control policies for distributed multimedia applications, and (d) ensures protection of user?s information. The project uses several innovative approaches to meet these objectives. First, novel real-time data mining and stream-correlation methods is used for extracting useful content and detecting events embedded in data streams by using multiple sliding window operators. Second, an efficient privacy/security based multimedia data clustering mechanism allows a prevailing security policy to be adaptable to the changing context and the datasets. The uniqueness of this approach is the adaptive nature of security policies. In addition, a novel XML-based integrated mechanism is used for encoding of multimedia data in conjunction with multi-domain security policies for ensuring technical viability and portability of the system. The proposed research is expected to have direct and long-term impact on developing secure information infrastructures, such as e-commerce, digital libraries, and healthcare systems, which are projected to have an	

								important role at a global scale for the next several decades. The research results can help in alleviating the heavy financial risks associated with information and system security and in general, meeting the national information infrastructure security needs and threat management. The research will be pursued in collaboration with Cyber Center, the Center for Education and Research in Information Assurance and Security (CERIAS) and PURVAC facilities at Purdue University with which the PI?s have strong affiliation and on-going collaboration. Results and tools developed will be made available over the Web to research community. In The team has a track record of providing their research results and tools to various research communities in the areas of multimedia databases and information security management. For further information see the project web site at the URL: http://cobweb.ecn.purdue.edu/~dmultlab/	
100	1547272	Center for Trustworthy Scientific Cyberinfrastructure	US	IoT	NSF	Division of Advanced Cyberinfrastructure (ACI)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1547272	The National Science Foundation funds over seven billion dollars of research annually, nearly all of which relies heavily on information technology. The digital data produced and computing systems used by that research are subject to the same risks as other data and computing systems on the Internet. Appropriate cybersecurity is necessary both to make today's scientific discoveries possible and to ensure that the science is trustworthy. However, NSF science is often necessarily performed in open, collaborative environments that span organizational and national boundaries. These constraints limit the use of traditional cybersecurity paradigms and technologies. Moreover, maintaining the usability of computer systems while providing cybersecurity is critical for many scientists who are not information technology experts. Different science domains also have varying requirements for	4,500,000.00

								data confidentiality and integrity. As the NSF Cybersecurity Center of Excellence, the Center for Trustworthy Scientific Cyberinfrastructure (CTSC) brings together experts in cybersecurity, knowledgeable and experienced in the scientific endeavor, who will provide the NSF community with leadership and support necessary to tackle the unique cybersecurity challenges in maximizing the production of trustworthy NSF science. CTSC will directly support individual NSF cyberinfrastructure projects and Large Facilities through collaborative engagements that address specific project needs. CTSC engagement activities include (but are not limited to) security reviews, security architecture design, identity and access management, and software assurance. CTSC will provide cybersecurity situational awareness to the NSF cyberinfrastructure community through timely advisories and notices. In collaboration with the Department of Energy's Energy Science Network, CTSC will develop and publish an information security threat model scoped to the particular assets and interests of the open science community. CTSC will continue to organize the annual NSF Cybersecurity Summit for Large Facilities and Cyberinfrastructure, providing the community with the opportunity to share best practices, attend practical training sessions, and collaborate on solving common challenges. CTSC will perform outreach and dissemination of best practices via the CTSC website (http://trustedci.org), blog posts, email lists, and online chats, as well as providing cybersecurity training in person and via online courses.	

101	1213052	Architecting the Next Generation Memory Hierarchy - A Holistic Approach	US	IoT	NSF	Division of Computing and Communication Foundations (CCF)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1213052	The memory system continues to be a major performance and power bottleneck in nearly all computing systems. And, it is becoming increasingly more so with major application, architecture, and technology trends. Embedded applications that acquire and process real-time data, Internet and cloud applications that have to analyze large databases, and the exa-scale era HPC applications that need to crunch voluminous data sets are just a few examples of increasingly data-intensive applications that require high memory capacity, performance, and energy efficiency. Thus, the well-known memory wall problem has become even more difficult to surmount and needs a fundamental rethinking of the memory hierarchy design for future computing platforms. The goal of this proposal is to fundamentally and holistically rethink the design of the entire memory hierarchy taking into consideration the emerging device/memory technologies and to exploit the design trade-offs at different layers of the system stack -- from devices to micro-architecture, compilers and runtime systems. The solution will cover innovations in architecting and optimizing the entire memory path consisting of the caches, on-chip networks, memory controller and main memory. The objective is to enable 100X improvement in memory capacity over the next decade, while providing 5X improvement in performance and 10X improvement in energy efficiency. The proposed research has the potential to transform the design of next-generation memory systems for the multi-core era, which is expected to be a ubiquitous part of the entire IT sector. The cross-cutting nature of this research can foster new research directions in several areas, spanning technology/energy-aware design, computer architecture, compilers, and system/application	1,225,000.00

								software. With the memory system forming the backbone of nearly every envisioned future application domain, the broader impact of this research can accelerate the design and deployment of future applications. This project will enable transfer of research results to industry, enhance undergraduate and graduate student training including under-represented students, and contribute to the development of new research and teaching tools.	
102	1329509	SecureVault Cloud Platform	US	IoT	NSF	Division of Industrial Innovation and Partnership (IIP)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1329509	This Small Business Innovation Research (SBIR) Phase II project will design and build a technology platform that allows cloud infrastructures to run clients' tasks with full computation privacy. The cloud provider itself is unable to access to customers' workloads even in the presence of malicious-intent direct physical access to the hardware itself. The technology significantly increases cloud defense capabilities, and provides strong computation privacy, and regulatory compliance guarantees to any regulatory- and security-sensitive customers. This level of security is achieved by (i) endowing a traditional infrastructure with a collaborative security layer of active transparent defense and control, a cloud ?immune system?, and (ii) new mechanisms that allow full computation privacy for outsourced tasks. The immune system has its own control network and can instantly share critical information and react independently for meaningful operation and fast recovery even under sustained attack and severe compromise. It also transparently monitors and accesses the server runtimes and can strongly react and immediately present a united defense across the entire cloud. Moreover, inherent privacy and confidentiality assurances allow safe computation execution and data hosting with full security, even in the case of a compromised curious service provider with full	1,000,000.00

								physical access to the infrastructure. The broader impact/commercial potential of this project will manifest itself in all dimensions of our increasingly internet and cloud-reliant society and technology. Because clouds introduce significant risks, technology-backed comprehensive security and privacy assurances are essential to establish cloud computing as a truly viable alternative to in-house IT. Unfortunately, these guarantees are not achievable with any of today's offerings or research efforts, despite being direly needed by all major potential cloud adopters, including financial, governmental and healthcare markets, heavily governed by security, regulatory compliance and intellectual property constraints. As a result, rather than risk regulatory compliance breaches, and unauthorized access to proprietary business logic, these major markets have simply not adopted the cloud. Countless surveys outline security and privacy consistently as the top concern preventing adoption, and markets have reacted with strong reservations. This project develops a disruptive technology that bridges this gap and makes cloud adoption secure and attractive in all these markets. As a result it will lead to exponential cross-market growth. Finally, enabling cloud deployment with full computation privacy, regulatory compliance and data confidentiality will significantly increase enterprise agility and competitiveness.	
103	1541106	Interdependent Electric and Cloud Servies for Sustainable,	US	IoT	NSF	Division of Advanced Cyberinfrastructure (ACI)	http://www.nsf.gov/awardsearch/showAward?AWD_ID=1541106	Electric energy networks are the cornerstone of the civil infrastructure of our society. These networks provide the energy essential to carrying out daily operations in education, health care, commerce, entertainment, defense, and government. However, electric energy markets, due to their vertical integration, often exclude customers from the processes associated with energy production, pricing, transmission and distribution. Smart	1,350,000.00

		Reliable, and Open Smart Grids						grids and distributed generation schemes have been proposed as mechanisms to modernize energy grids and balance the current power structures in electric markets. In a smart grid, computers and communications networks are attached to the power generation, transmission, distribution and load elements, establishing a mechanism to gather information, control generation, control demand, diagnose problems, bid for prices in energy markets, and forecast energy consumption. However, a smart grid creates interdependencies between the energy network and the computer network since the energy network powers the computers that in turn control the operation of the energy grid. In this project, a team from the University of Puerto Rico, Mayaguez (UPRM) will study smart grids and the interdependency between the energy grid and the IT infrastructure that is setup to manage it. This project champions a transformation of the electric grid, moving it away from being centered on centralized utilities that supply most, if not all, power services. Instead, the grid becomes a marketplace of third-party power-service suppliers, who compete to sell their electric services over the Internet. These services include energy block purchases, storage, billing, weather forecasting, energy demand forecasting, and other ancillary services. This brings in an important societal element - it empowers common citizens, whose homes are now renewable energy generation systems, to become suppliers and key actors in the energy market. This project is thus aimed at designing and developing the basic science and technology for an Open Access Smart Grid in order to create truly sustainable energy markets. In this project, the smart grid is modelled as a collection of interdependent electric and cloud services, whose	

								collaborative interactions help manage the smart grid. All the electric services (e.g., energy, storage, billing) are exposed to users as REST-based cloud services, enabling the development of algorithms and applications for customers, power producers, and other users to consume or subscribe to these electric services, collect operational data and customer feedback, and support analytics to predict electric energy demands. Microgrids and renewable energy systems will be important components in this framework, as they enable modularization of the grid into autonomous or semi-autonomous subsystems. The research team will develop methods to map reliable power microgrids into electric services that can be rapidly brought online to compensate for lost generation capacity or to obtain more affordable energy. A major challenge with microgrid systems is activating them without introduction major power disturbances in the system. Another challenge is forecasting the availability of renewable energy, which will be addressed this by developing rain-cell tracking frameworks for solar and wind output estimation services, and the determination of local sensors requirements to improve short-term forecasts services. Finally, the team will apply the social acceptance model to the development, implementation, management and assessment of the Open Access Smart Grid with the purpose of identifying the institutional change necessary for the integration of all stakeholders and the effective democratization of electric services.	
104	DE-FOA-0001493	Addressing Risk and Uncertainty in the Future Power	US	CPS	Department of Energy	National Energy Technology Laboratory	http://www.grants.gov/search-grants.html?fundingCategories%3D	The nation's wholesale electricity markets and transmission planning are in a state of transition. The Department of Energy's (DOE's) Office of Electricity Delivery and Energy Reliability (OE) is interested in both operational and planning modeling and computation methodologies/techniques needed to support the	1,600,000.00

		System					ST%7CScience%20and%20Technology%20and%20other%20Research%20and%20Development	future engineering and market functions required by these systems. Addressing risk and uncertainty is central to meeting the needs and ensuring reliability is a fundamental requirement of the system. There are three research areas for this Funding Opportunity Announcement (FOA): a) wholesale market operations, b) transmission planning, and c) demand-side participation. In accordance, with Section III-Eligibility Requirements, this FOA is being restricted to United States (US) Colleges, Universities, and University-affiliated Research Institutions with accredited undergraduate and graduate programs.	
105		Cyber-Physical Systems for Global Cities	US	CPS / IoT	NIST		http://www.nist.gov/el/smartgrid/cpsforglobalcities.cfm	Cyber-Physical Systems (CPS) provide cities with a pathway to enhance and integrate key infrastructures and systems to dramatically improve delivery of government and other services to citizens. In the past, cities have used isolated and labor-intensive methods to manage key infrastructures such as energy, water, transportation, emergency response and others in one-by-one solutions that are not interoperable. This approach is costly for regular operations and presents significant problems for resilience planning and disaster response in which multiple infrastructures must interact. To increase resilience, improve the quality of life, and reduce budgets, cities are increasingly looking to advanced cyber-physical systems concepts in which hybrid engineered and IT systems can be integrated for smart city management. Many Cyber-Physical Systems (CPS) innovators already have technologies (i.e., building blocks) and their impact can be maximized by fostering collaboration among the innovators to create interconnected solutions to provide tangible benefits to end users. Many smart community efforts are one-off projects with heavy emphasis on customization and inadequate consideration for future upgradability and	

								extensibility, which end up causing increased cost and inefficiency. As a result, many Smart Cities/Communities deployments are isolated and do not enjoy the economy of scale. This program addresses a key need for progress in smart cities efforts: consensus standards for interoperability and replicability across systems and measurement science for performance comparisons and evaluation, validation, verification, and management. Central to the program's strategy is the development of measurement science and standards in the context of real deployments at scale to ensure the outcomes are relevant and useful in the real world.	
106		Cyber-Physical Systems Testbed Design Concepts	US	CPS	NIST		http://www.nist.gov/el/smartgrid/cpstestbedconcepts.cfm	A key challenge to progress in cyber-physical systems (CPS) is the lack of robust platforms for experiment and testing, which NIST is addressing through development of a modular, composible multi-domain CPS testbed. Today, many CPS experiments are done either in operational systems or in domain-specific testbeds. The former are limited by the severe constraints required to ensure that experiments and testing do not affect reliability and safety for systems that are providing critical, often life- and health-safety, functions in real time. The latter are limited by the inability to test in constrained environments the general applicability of CPS concepts and technologies intended for implementation across multiple domains and in varied applications. The Program is addressing this need through the development of design principles for modular, composable testbeds that are interoperable with facilities across the nation and around the world for varying scale and readily reconfigurable for work across domains and applications, and through development of a cross-sector CPS testbed.	

107		Reference Architecture for Cyber-Physical Systems	US	CPS	NIST		http://www.nist.gov/el/smartgrid/cpsarchitecture.cfm	Cyber-Physical Systems (CPS) are smart systems that include engineered interacting networks of physical and computational components. CPS and related systems (including the Internet of Things, Industrial Internet, and more) are widely recognized as having great potential to enable innovative applications and impact multiple economic sectors in the world-wide economy. These application areas include energy infrastructures, advanced manufacturing, building control, transportation, health care, and others. Current design and management approaches for these systems are domain-specific, resulting in redundant efforts, limited sharing across domains, and a lack of robust, formal methods for design, evaluation, verification, and validation. This project addresses these limitations through the development of a CPS Framework with common vocabulary, analysis methodology, reference architecture concepts and use cases to serve as the basis for shared development, information exchange, and new formal methods applicable across domains. This work is conducted in partnership with public and private sector stakeholders through the NIST CPS public working group (CPS PWG) comprising technical experts from industry, academia, and government.	
108		Industrial Internet Consortium	US	IoT	-	Object Management Group	http://www.iiconsortium.org/	Imagine a highway where cars are able to safely navigate to their destinations without a driver. Imagine a home where an elderly patient's health is closely monitored by her hospital physician. Imagine a city that significantly reduces waste through sensor-embedded water pipes, buildings, parking meters and more. These are no longer a part of the distant future. These scenarios are starting to happen now, through the convergence of machines and intelligent data in what is called the Industrial Internet. The Industrial Internet will transform industry through intelligent, interconnected	

								objects that dramatically improve performance, lower operating costs and increase reliability. The Industrial Internet Consortium is a global not-for-profit, open membership organization formed to accelerate the development, adoption, and wide-spread use of interconnected machines and devices, intelligent analytics, and people at work. Founded by AT&T, Cisco, General Electric, IBM, and Intel in March 2014, the Industrial Internet Consortium catalyzes and coordinates the priorities and enabling technologies of the Industrial Internet.	
109		Open Connectivity Foundation	US	IoT	-		https://openconnectivity.org/	Billions of connected devices (devices, phones, computers and sensors) should be able to communicate with one another regardless of manufacturer, operating system, chipset or physical transport. The Open Connectivity Foundation (OCF) is creating a specification and sponsoring an open source project to make this possible. OCF will unlock the massive opportunity in the IoT market, accelerate industry innovation and help developers and companies create solutions that map to a single open specification. OCF will help ensure secure interoperability for consumers, business, and industry. WHY ARE WE DOING THIS? We want to connect the next 25 billion devices for the Internet of Things. We want to provide secure and reliable device discovery and connectivity across multiple OSs and platforms. There are multiple proposals and forums driving different approaches... but no single solution addresses the majority of key requirements. To get industry consolidation around a common,	

								interoperable approach. To build a broad industry consortium of companies to create a scalable solution. WHAT ARE WE DOING? Defining the specification, certification & branding to deliver reliable interoperability -- a connectivity framework that abstracts complexity. Our open specification allows anyone to implement and it is easy for developers to use It includes predictable IP protection & branding for certified devices (via compliance testing) and service-level interoperability. There is also an Open Source implementation of our specification - IoTivity Project This Open Source implementation is designed to enable application developers and device manufacturers to deliver interoperable products across Android, iOS, Windows, Linux, Tizen, and more.	
110		Internet of Things (IoT) Initiative	US	IoT	-	IEEE	http://iot.ieee.org/	The IEEE Internet of Things is one of IEEE's important, multi-disciplinary, cross-platform Initiatives. The Internet of Things (IoT) is one of the most exciting technological developments in the world today and the global technical community is coalescing around the thought-leading content, resources, and collaborative opportunities provided by the IEEE IoT Initiative. More information is revealed daily about the Internet of Things and its potential to transform how we communicate with machines and each other. To bring clarity to and disseminate information globally, IEEE Future Directions launched the IEEE IoT Initiative in 2014. It serves as a home for the global community of engineering and technology professionals in industry, academia, and government working in related technologies. Here, professionals learn, share knowledge, and collaborate on this sweeping	

								convergence of technologies, markets, applications, and the Internet. Participants in the community have access to the most trusted resources developed including publications, videos, articles, and interviews, as well as webinars, Hangouts, presentations, workshops, and conferences, this web portal, and much more. You are invited to join, participate, and enjoy the benefits of becoming part of this cutting-edge Initiative. The IEEE IoT Technical Community is comprised of individuals involved in research, implementation, application, and usage of this Internet-enabled vision of our future. As an active member of the community, you will stay abreast of developments in this multi-disciplinary area and remain at the forefront of Internet of Things research, development, and planning. Members of the Technical Community have access to the most trusted resources being developed, such as the IEEE IoT eNewsletter, a bi-monthly, online publication that features practical and timely technical information and forward-looking commentary on developments and deployments around the world. Members of the Technical Community are invited to participate in progressing the definitions within the IoT through its collaborative document, Towards a Definition of the Internet of Things. Your contribution of knowledge and perceptions to this "living document" can facilitate a better understanding of the subject, lead to further research, and advance our understanding of this emerging concept. To really understand the Internet of Things requires application, implementation, and execution in the real world. The Initiative launched the IoT Scenarios program to provide the community an interactive platform on which to engage with use cases, service descriptions, business models, and reference implementations that will be key to developing a vibrant	

								industry.	
111		Allseen Alliance	US	IoT	-		https://allseenalliance.org/	The AllSeen Alliance is a cross-industry consortium dedicated to enabling the interoperability of billions of devices, services and apps that comprise the Internet of Things. It has 200+ companies as members.	
112	MultiPARTES	Multi-cores Partitioning for Trusted Embedded Systems	EU	CPS, mixed criticality	FP7		http://www.multipartes.eu/	Growing complexity of applications makes the integration of security and dependability an issue in many domains (e.g. energy supply, transportation, industrial control, aerospace, etc). The engineering of embedded systems needs to take these aspects into account. However, guaranteeing security and dependability in a situation of increasing system complexity is leading to unacceptable development cost and time to market, especially for SMEs, due to the price of tools. The main challenge of this project is supporting mixed criticality embedded systems on multicore open source virtualized platforms in such a way that the development, validation and certification efforts can be lower than the corresponding effort required on independent hardware platforms when using an appropriate methodology. An approach to increasing maintainability and to avoid the growing validation and certification effort is to incorporate mechanisms that establish multiple partitions on the same hardware platform with strict temporal and spatial separation between the individual partitions. In this approach, applications with different levels of dependability can be placed in different partitions and can be validated (and certified if required) in isolation, the MultiPARTES approach. This allows the user to manage complexity while keeping down an escalation of the development effort, but this concept needs to be adapted and applied to multicore and heterogeneous multicore systems. This project aims at developing tools and solutions	2,850,000.00

								based on mixed criticality virtualization systems for multicore platforms. The starting point for the virtualization support is XtratuM, a cost-effective open source hypervisor developed specifically for real-time embedded systems by one of the project participants (UPVLC) – a hypervisor that is being increasingly used by the aerospace industry. Based on this approach, MultiPARTES will offer a rapid and cost-effective development support of dependable real-time embedded systems enabling critical and non critical applications to run on the same hardware platform. To achieve this goal we will develop an innovative multicore-platform virtualization layer based on XtratuM. We will devise a methodology permitting the partitioning of multicore systems, thereby speeding up development and production of mixed-criticality applications based on the partitioning. We will demonstrate these open virtualization solutions on COTS hardware platforms and on enhanced heterogeneous multicore hardware platforms to show the increase of time and space isolation, overcoming some of the COTS hardware limitations. The results will be evaluated in case studies in three application sectors: wind power, video surveillance and aerospace.	
113	CONTREX	Design of embedded mixed-criticality CONTRol systems under consideration of EXtra-functional	EU	CPS, mixed criticality	FP7	European Mixed-Criticality Cluster	https://contrex.offis.de	Up to now mission & safety critical services of SoS (Systems of Systems) have been running on dedicated and often custom designed HW/SW platforms. In the near future such systems will be accessible, connected with or executed on devices comprising off-the-shelf HW/SW components. Significant improvements have been achieved supporting the design of mixed-critical systems by developing predictable computing platforms and mechanisms for segregation between applications of different criticalities sharing computing resources. Such platforms enable techniques for the compositional	6,050,000.00

		properties						certification of applications' correctness, run-time properties and reliability. CONTREX will complement these important activities with an analysis and segregation along the extra-functional properties real-time, power, temperature and reliability. These properties will be a major cost roadblocks when scaling up the number of applications per platform and the number of cores per chip, in battery powered devices or switching to smaller technology nodes. CONTREX will enable energy efficient and cost aware design through analysis and optimisation of real-time, power, temperature and reliability with regard to application demands at different criticality levels. To reinforce European leadership and industrial competiveness the CONTREX approach will be integrated into existing model-based design methods that can be customized for different application domains and target platforms. CONTREX will focus on the requirements derived from the automotive, aeronautics and telecommunications domain and evaluate its effectiveness and drive integration into existing standards for the design and certification based on three industrial demonstrators. Valuable feed-back to the industrial design practice, standards, and certification procedures is pursued. Our economic goal is to improve energy efficiency by 20 % and to reduce cost per system by 30 % due to a more efficient use of the computing platform.	
114	DREAMS	Distributed REal-time Architecture for Mixed Criticality	EU	CPS, mixed criticality	FP7	European Mixed-Criticality Cluster	https://www.uni-siegen.de/dreams/home	The objective of DREAMS is to develop a cross-domain architecture and design tools for networked complex systems where application subsystems of different criticality, executing on networked multi-core chips, are supported. DREAMS will deliver architectural concepts, meta-models, virtualization technologies, model-driven development methods, tools, adaptation strategies and	10,999,950.00

		Systems						validation, verification and certification methods for the seamless integration of mixed-criticality to establish security, safety, real-time performance as well as data, energy and system integrity. The ambitious goals of the project are supported by 16 project partners distributed across six European countries. The project partners include major European companies encompassing large enterprises (Alstom, STMicroelectronics, Thales, TÜV Rheinland) and SMEs (FENTISS, RealTime-at-Work, TTTech, Virtual Open Systems,). Furthermore leading universities and research institutes (FORTISS, IKERLAN, ONERA, Polytechnic University of Valencia, SINTEF, Technological Educational Institute of Crete, Technical University of Kaiserslautern and University of Siegen) contribute to DREAMS. The four-years DREAMS project is coordinated by University of Siegen and is one of three projects funded by the European Commission in the area of mixed-criticality systems from the FP7 ICT call 10. The EC financial contribution amounts to 11 million euros.	
115	PROXIMA	Probabilistic real-time control of mixed-criticality multicore and manycore systems	EU	CPS, mixed criticality	FP7	European Mixed-Criticality Cluster	www.proxima-project.eu	In the next decade, EU industries developing Critical Real-Time Embedded Systems (CRTES) (safety, mission or business critical) will face a once-in-a-life-time disruptive challenge caused by the transition to multicore processors and the advent of manycores, tantamount to complex networked systems. This challenge brings the opportunity to integrate multiple applications onto the same hardware platform bringing significant advantages in performance, production costs, and reliability. It also brings a severe threat relating to a key problem of CRTES; the need to prove that all temporal constraints will be satisfied during operation. Current CRTES, based on relatively simple singlecore processors, are already extremely difficult to analyse for	4,650,000.00

								temporal behaviour, resulting in errors in operation costing EU industry billions each year. The advent of multicore and manycore platforms exacerbates this problem, rendering traditional temporal analysis techniques ineffectual. A new approach is needed.The PROXIMA thesis is that the temporal behaviour of mixed-criticality CRTES executing on multicore and manycore platforms can be analysed effectively via innovative probabilistic techniques. PROXIMA defines new hardware and software architectural paradigms based on the concept of randomisation. It extends this approach across the hardware and software stack ensuring that the risks of temporal pathological cases are reduced to quantifiably small levels. On top of this, PROXIMA builds a comprehensive suite of probabilistic analysis methods integrated into commercial design, development, and verification tools, complemented by appropriate arguments for certification. PROXIMA provides a complete infrastructure; harnessing the full potential of new processor resources, demonstrating and supporting effective temporal analysis, bringing the probabilistic approach to a state of technological readiness, and priming multiple EU industry sectors in its use via a number of case studies.	
116		Internet Architecture Board (IAB)	US	IoT	-	IETF	https://www.iab.org/	The Internet Architecture Board provides long-range technical direction for Internet development, ensuring the Internet continues to grow and evolve as a platform for global communication and innovation. In its work, the IAB strives to: Ensure that the Internet is a trusted medium of communication that provides a solid technical foundation for privacy and security, especially in light of pervasive surveillance, Establish the technical direction for an Internet that will enable billions more people to connect, support the	

								vision for an Internet of Things, and allow mobile networks to flourish, while keeping the core capabilities that have been a foundation of the Internet’s success, and Promote the technical evolution of an open Internet without special controls, especially those which hinder trust in the network.	