Opportunity Report

"Towards Enhanced EU-US ICT Pre-competitive Collaboration"

Big Data

Vasilis Papanikolaou,
Nikos Sarris

iLAB
ATC SA, Greece

With support by:

Christian Sonntag,
Sebastian Engell

Process Dynamics and Operations Group
Dept. of Biochemical and Chemical Engineering
TU Dortmund University, Germany

Steffen Watzek, Yaning Zou
Lucas Scheuvens, Gerhard Fettweis

Mobile Communications Systems
Faculty of Electrical and Computer Engineering
TU Dresden University, Germany

Jonathan Cave,
Maarten Botterman

Department of Economics
The University of Warwick, UK
-
GNKS, IGF DC IoT, NLnet

Revised version V1.1

Please send any feedback to: V.Papanikolaou@atc.gr

ICT Policy, Research and Innovation
for a Smart Society

May 2018

www.picasso-project.eu

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

Executive Summary

This report describes the major results that were obtained by the PICASSO Expert Group on Big Data throughout the duration of the PICASSO project. The major contributions of this report are:

- **Technology themes** (chapter 3) and **collaboration opportunities and mechanisms** (section 4.3) that have been identified as being promising for EU-US collaboration in the Big Data sector, synthesized based on comprehensive analyses of:
- The **EU and US research and innovation priorities** in the technology sectors and related application domains (chapter 2),
- The **EU-US funding and collaboration landscape** (section 4.1), and
- **Barriers for EU-US collaboration** (section 4.2).

In chapter 3 of this report, the PICASSO Expert Group on Big Data has defined technology themes that are promising for EU-US collaboration:

- Interoperability and Standardization
- Adding a semantic layer to Big Data technology
- Integrating Linked Data and Big Data technology
- Enable discovery of deeper, fresher insights from all enterprise data resources
- Improve efficiency, effectiveness, and decision-making
- Facilitate more timely, agile response to business opportunities, threats, and challenges
- Provide a single view of diverse data resources throughout the business chain
- Support tighter security, protection, and governance of data throughout its lifecycle
- Improve the scale, efficiency, performance, and cost-effectiveness of data/analytics platforms

In addition to the technological topics presented in chapter 3, the Big Data Expert Group has identified opportunities in the areas of the Big Data ecosystem, standardization and regulation, and education and workforce.The contents of this report are based on in-depth discussions with a large network of international experts, analytical research by the Expert Groups, preliminary PICASSO results (i.e. the reports (1), (2), and (3)) and other feedback collection mechanisms such as a public consultation on the PICASSO website. Moreover, a dedicated workshop was organized (Transatlantic Workshop on Public Private Partnerships for Big Data Research and Innovation and Workforce Development[1]) in Versailles, France, on November 20, 2017 as a partnership between the **US National Science Foundation (NSF) Big Data Regional Innovation Hubs, the EU Big Data Value Association (BDVA), the PICASSO project, and INRIA**, with more than 50 EU and US participants from both academia and industry, to discuss and conclude on specific collaboration opportunities on Big Data between EU and US and potential collaboration mechanisms and initiatives. Adding to the above, the Big Data opportunity report was circulated for feedback collection and validation by members of various industrial and research-oriented associations and projects such as the *Big Data Value Association* (*BDVA*), *Big Data Europe*, and *NESSI ETP*. Moreover, members of other initiatives and US government agencies such as the *IEEE Big Data Initiative*, *South Big Data Regional Innovation Hub* and the *National Institute of Standards and Technology* (*NIST*), have also validated the contents of this report.

---

[1] http://www.picasso-project.eu/2017/11/27/trans-atlantic-workshop-on-public-private-partnerships-for-big-data/

The contents and the outcomes of this report are mainly addressed to the research community and policy makers willing to enhance collaboration between the two regions by defining common big data opportunities and challenges, both technological and societal, to be mutually tackled therefore maximise the impact of big data in a number of societal challenges.

The opportunity report provides a common view on priorities and future cooperation opportunities between the EU and the US and is a strong basis and guideline for concrete EU-US collaboration actions of the PICASSO project.

The PICASSO Project

The aim of the 30-months PICASSO project is (1) to reinforce EU-US collaboration in ICT research and innovation focusing on the pre-competitive research in key enabling technologies related to societal challenges - 5G Networks, Big Data, Internet of Things and Cyber Physical Systems, and (2) to support the EU-US ICT policy dialogue by contributions related to e.g. privacy, security, internet governance, interoperability, ethics.

PICASSO is oriented to industrial needs, provides a forum for ICT communities and involves 24 EU and US prominent specialists in the three technology-oriented ICT Expert Groups - 5G, Big Data, and IoT/CPS - and an ICT Policy Expert Group, working closely together to identify policy gaps in the technology domains and to take measures to stimulate the policy dialogue in these areas. A synergy between experts in ICT policies and in ICT technologies is a unique feature of PICASSO.

A number of analyses will be accomplished, as well as related publications, that will for a major part be made public and contribute to the project's outreach. Dedicated communication and dissemination material will be prepared that should support the operational work and widespread dissemination though different channels (website, social media, publications …). The outreach campaign will also include 30+ events, success stories, factsheets, info sessions, and webinars.

PICASSO Project Coordination:

Svetlana Klessova, Project Coordinator
inno TSD, France
+33 4 92 38 84 26
s.klessova@inno-group.com

About the PICASSO Project:

PICASSO is co-funded by the European Commission under the Horizon 2020 programme.
Start Date: 1st January 2016
Duration: 30 months
Total budget: 1,160,031 €, including a contribution from the European Commission of 999,719 €
Project Website: http://www.picasso-project.eu/

PICASSO Consortium Members:

inno	**inno TSD, France** – one of Europe's leading innovation management consultancy firms, specialised in helping major private and public stakeholders design and implement R&D and innovation projects. https://www.inno-tsd.fr/en
tu technische universität dortmund	**TECHNISCHE UNIVERSITÄT DORTMUND, Germany** – a leading German technically oriented research university with strong research groups in big data, communications, smart grids, e-mobility and cyber-physical systems. http://www.tu-dortmund.de
THHINK WIRELESS TECHNOLOGIES LTD	**THHINK WIRELESS TECHNOLOGIES LIMITED, United Kingdom** - an ICT company founded in 2009 after more than a decade of research and development in wireless and energy harvesting technologies. http://www.thhink.com/
ATC ATHENS TECHNOLOGY CENTER	**ATC SA, Greece** - an SME and Technology Centre in the field of ICT participating in 3 ICT European Technology Platforms: NESSI (Steering Committee member), NEM (member) and NETWORLD2020 (member), and founding member of European Big Data Value Association. http://www.atc.gr

APRE Agency for the Promotion of European Research	**AGENZIA PER LA PROMOZIONE DELLA RICERCA EUROPEA, Italy** – a non-profit research organisation, grouping together more than 100 members, including public and private research centres, industries, industrial associations, chambers of commerce, science parks and more than 50 universities, with the main objective to promote the participation in national and European RTD programmes. http://www.apre.it/
Honeywell	**HONEYWELL INTERNATIONAL INC, United States** – a multinational company and global leader that invents and manufactures technologies to address some of the world's toughest challenges initiated by revolutionary macrotrends in science, technology and society. The company's products and solutions are focused on energy and the environment, safety and security, and efficiency and productivity. http://honeywell.com/
GNKS Consult	**GNKS CONSULT BV, Netherlands** - conducting strategic and policy research and evaluation, building on excellence in understanding of the impact of the emerging Global Networked Knowledge Society http://www.gnksconsult.com/
TECHNISCHE UNIVERSITÄT DRESDEN	**TECHNISCHE UNIVERSITÄT DRESDEN, Germany** - a full-scale university with 14 faculties, covering a wide range of fields in science and engineering, humanities, social sciences and medicine. https://tu-dresden.de/
FIU FLORIDA INTERNATIONAL UNIVERSITY	**FLORIDA INTERNATIONAL UNIVERSITY, United States** - The Miami-Florida Jean Monnet Center of Excellence, (MFJMCE), a member of the global network of EU-sponsored Jean Monnet centers, has the mission to promote teaching, research and outreach activities relating to the EU. http://www.fiu.edu/; https://miamieuc.fiu.edu/
Technological Leadership Institute	**UNIVERSITY OF MINNESOTA, United States** – The Technological Leadership Institute bridges the gap between business and engineering. TLI's mission is to develop local and global leaders for technology enterprises. https://tli.umn.edu/

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Program
4G	4th Generation
5G	5th Generation
AI	Artificial Intelligence
AIOTI	Alliance of IoT Innovation
AV	Autonomous Vehicle
AWS	Amazon Web Services
B2B	Business-to-business
B2C	Business-to-customer
BBI	Bio-based Industries
BD	Big Data
BDVA	Big Data Value Association
BDVPPP	Big Data Value Public Private Partnership
CEDR	Conference of European Directors of Roads
CERN	Conseil Européen pour la Recherche Nucléaire
CPS	Cyber-physical System
CPSoS	Cyber-physical System of Systems
CPS-VO	CPS Virtual Organization
CPU	Central Processing Unit
CS	Clean Sky
CSAAC	Cyber Situational Awareness Analytical Capabilities
D2D	Device-to-Device
DARPA	Defense Advanced Research Projects Agency
DHS	Department of Homeland Security
DISA	Defense Information Systems Agency
DoC	Department of Commerce
DoD	Department of Defense
DoDIN	DoD Information Networks
DoE	Department of Energy
DoS	Department of State
DoT	Department of Transportation
DSL	Digital Subscriber Line
DSM	Digital Single Market
EC	European Commission
ECSEL	Electronic Components and Systems for European Leadership
EeB	Energy-efficient Buildings
EG	Expert Group
EPI	European Platform Initiative
ERA	European Research Area
EU	European Union
FBMC	Filter-Bank Multi-Carrier
FCC	Federal Communications Commission
FCH	Fuel Cells and Hydrogen
FET	Future and Emerging Technologies
FIRE	Future Internet Research & Experimentation
FoF	Factories of the Future
FP7	Framework Programme 7

FY	Financial Year
Gbps	Gigabit per second
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GENI	Global Environment for Networking Innovations
GFDM	Generalized Frequency-Division Multiplexing
GHz	Gigahertz
H2020	Horizon 2020
H2M	Human-to-machine
HD	High-definition
HMI	Human Machine Interface
HPC	High Performance Computing
HPUE	High Performance User Equipment
IA	Industry Association
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IERC	IoT European Research Cluster
IIC	Industrial Internet Consortium
IIoT	Industrial Internet of Things
IM	Innovative Medicine
IMS	Intelligent Manufacturing Systems
INCOSE	International Council on Systems Engineering
IoT	Internet of Things
IP	Intellectual Property
IPR	Intellectual Property Rights
ISM	Industrial, Scientific, Medical
ITER	International Thermonuclear Experimental Reactor
ITS	Intelligent Traffic System
ITU	International Telecommunication Union
JTI	Joint Technology Initiative
JU	Joint Undertaking
LTE	Long Term Evolution
M2M	Machine-to-Machine
M&S	Modeling and Simulation
MEC	Mobile Edge Computing
MHz	Megahertz
MIMO	Multiple Input Multiple Output
MoU	Memorandum of Understanding
ms	Millisecond
NACFAM	National Coalition for Advanced Manufacturing
NB-IoT	Narrowband IoT
NCP	National Contact Point
NCURA	National Council of University Research Administrators
NFV	Network Function Virtualization
NGI	Next Generation Internet
NGMN	Next Generation Mobile Networks
NIH	National Institutes of Health
NIPRNet	Nonsecure Internet Protocol Router Network
NISD	Network and Information Security Directive

NIST	National Institute of Standards and Technology
NIT	Networking and Information Technology
NITRD	Networking and Information Technology Research and Development
NSF	National Science Foundation
NTIA	National Telecommunications and Information Administration
OCF	Open Connectivity Foundation
OFDM	Orthogonal Frequency Division Multiplexing
OMG	Object Management Group
PAWR	Platforms for Advanced Wireless Research
PCAST	President's Council of Advisors on Science and Technology
PPP	Public Private Partnership
PWG	Public Working Group
QoE	Quality of Experience
R&D	Research and Development
R&I	Research and Innovation
RAT	Radio Access Technology
RDI	Research, Development, Innovation
RFC	Request for Comments
SAE	Society of Automotive Engineers
SDAV	Scalable Data Management, Analysis and Visualization
SDN	Software Defined Networking
SEED	Standard Energy Efficiency Data
SIPRNet	Secret Internet Protocol Router Network
SME	Small and Medium-sized Enterprises
SMLC	Smart Manufacturing Leadership Coalition
SoS	System of Systems
SOTA	State of the Art
SPIRE	Sustainable Process Industry
SRA	Strategic Research Agenda
SSG	Senior Steering Group
Tbit	Terabit
Tbps	Terabit per Second
TRL	Technology Readiness Level
TTIP	Transatlantic Trade and Investment Partnership
TV	Television
UE	User Equipment
UHD	Ultra High Definition
URLLC	Ultra-reliable Low-latency Communications
US	United States
USGS	US Geological Survey
V2I	Vehicle-to-infrastructure
V2V	Vehicle-to-vehicle
V2X	Vehicle-to-everything
V5GTF	Verizon 5G Technology Forum
VDA	Verband Der Automobilindustrie
VPN	Virtual Private Network
ZT-OFDM	Zero-tail OFDM

Table of Contents

1. Introduction

1.1. Purpose of the Report

This report provides a detailed overview of the current status of collaboration between EU and US in the area of Big Data, while presenting an in-depth analysis of technological themes-priorities and funding-supporting mechanisms available in each region, aiming on supporting big data research.

Findings of this report are based on various sources (i.e. from PICASSO project deliverables, online sources, reports, etc.) and inputs deriving from the Big Data Expert Group while roadmaps and reports produced by high level organisations and structures (i.e. European Commission, White House Science and Technology Office, NIST, etc.) have been extensively analysed and taken into account. Adding to the above, a database of more than 300 Big Data research projects has been created and analysed, including information both for EU and for US funded projects, funded by different initiatives, programmes and funding agencies.

The content and the outcomes of this report are mainly addressed to the research community and policy makers willing to enhance collaboration between the two regions by defining common big data opportunities and challenges, both technological and societal, to be mutually tackled therefore maximise the impact of big data in a number of societal challenges.

1.2. Technological Context

Data has become a key asset for the economy and our societies similar to the classic categories of human and financial resources. Whether it is geographical information, statistics, weather data, research data, transport data, energy consumption data, or health data, the need to make sense of "Big data" is leading to innovations in technology and the development of new tools and new skills.

Big data refers to large amounts of data produced very quickly by a high number of diverse sources. Data can either be created by people or generated by machines, such as sensors gathering climate information, satellite imagery, digital pictures and videos, purchase transaction records, GPS signals, etc. It covers many sectors, from healthcare to transport and energy. Generating value at the different stages of the data value chain will be at the centre of the future knowledge economy (4)*.*

Moreover, in a recent study 69% of corporate executives named greater data variety as the most important factor, followed by volume (25%), with velocity (6%) trailing – indicating that the big opportunity lies in integrating more sources of data, not bigger amounts (5)*.*

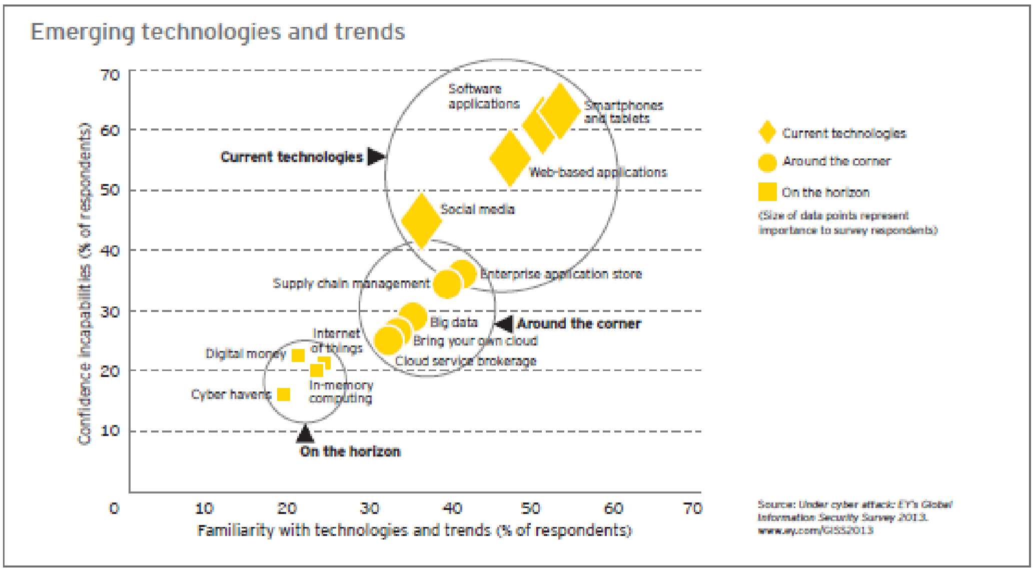

Figure 1: Big Data as an Emerging Technology **(6)**.

Big data is generating an intense amount of attention among businesses, media and even consumers, along with analytics, cloud-based technologies, digital channels and data visualization. These are all part of the current diverse ecosystem created by the technology megatrends. Some even herald the potential transformative power of the current trends as rivalling that of the internet. Yet, as in the early days of the internet, there is uncertainty about just what Big Data is, its potential benefits and the associated risks (6).

Both the EU and the US are funding numerous research and innovation activities related to Big Data in order to tackle a number of societal challenges, identified by policy makers and structures. However, performing joint research for Big Data is a challenge that cannot be overseen, as both regions face common challenges that can be jointly tackled. The Big Data sections of this report present an overview of research and innovation activities, action plans, funding opportunities, and challenges that both regions implement, in order to tackle a multi-angled topic such as Big Data. Moreover, we try to identify similarities and differences between these two regions, and any opportunities that need to be taken into account by policy makers for setting up joint funding schemes and initiatives.

The Big Data sections of this report were circulated for feedback collection and validation by members of various industrial and research-oriented associations and projects such as the *Big Data Value Association* (*BDVA*), *Big Data Europe*, and *NESSI ETP*. Moreover, members of other initiatives and US government agencies such as the *IEEE Big Data Initiative*, *South Big Data Regional Innovation Hub* and the *National Institute of Standards and Technology* (*NIST*), have also validated the contents of this report.

2. Research and Innovation Priorities in the EU and the US

This section presents an overview of the research and innovation priorities both in the EU and the US in the area of Big Data technologies. It provides an extensive analysis of the actions plans defined and implemented in both regions (EU and US) in order to support and boost growth of the Big Data sector, while it presents facts and information regarding the most critical application sectors, as these have been defined by top-tier structures and organizations. Adding to the above, the current report provides a summary of the most critical needs and drivers for some key application sectors such as smart cities, transportation and energy.

The findings of this section are based on various sources (i.e. from PICASSO project, deliverables, online sources, reports, etc.) and inputs deriving from the Big Data Expert Group. Roadmaps and reports produced by high level organisations and structures (i.e. European Commission, White House Science and Technology Office, NIST, etc.) have been extensively analysed and are presented in this section. Adding to the above, a database of more than 300 Big Data research projects has been created, including information both for EU and for US funded projects, funded by different initiatives, programmes and funding agencies

2.1. Big Data Technology Enablers

Big Data technologies are heavily dependent on various enabling technologies and fields that are being or will be applied to drive radical change in the Big Data field in general. Recent or future innovations in Data Storage technologies, IoT, Computation Capacity and other fields, are only some of the technology sectors that heavily influence the future of Big Data.

1) **Data Storage:** Data Storage can be considered as a critical key technology enabler for Big Data. As the cost for storing and maintaining complete data sets available for analytics (i.e. less than $600 to buy a disk drive with the capacity to store all of the world's music (7)), and new data technologies are on the way (i.e. Helium Drives, Shingled Magnetic Recording Drivers, etc.), companies and users will be able to store more and more data. The continued reduction of storage hardware costs, and improved data efficiency capabilities like de-duplication, and compression have become pervasive. In the most recent EMC Digital Universe research (8), IDC predicts the amount of digital data generated in each of the next two years will double, and will continue to double every two years for the rest of the decade.

2) **Computational Capacity:** Extremely large volumes of data have traditionally not been captured and processed for various reasons, most notably because the cost to do so was far greater than the value of insights companies could derive from its analysis. However, multiple factors and new technologies have lowered the cost and technology barrier for effective data processing, allowing companies of all sizes, to be able to unlock the value contained in different data sources. For instance, it is difficult for conventional relational databases to handle unstructured data, so software frameworks like Hadoop(R), for distributed storage and parallel processing of large datasets have been introduced to process non-structured data at high speed; making it easier to perform a more comprehensive analysis of big data (6). Big Computing at small prices, has given the opportunity to a number of organisations to look at, and deal with, data in ways not possible before. It's this computational capacity that has the real potential to transform data from a compliance burden into a business asset (9).

3) **Data Availability:** The third enabler is the increase in availability of data, especially of unstructured data types such as images, video, and audio. The EMC Digital Universe study predicts 30% of all digital data by the end of the decade will be security related with the majority being images, video, and audio. Traditional analytics cannot leverage these data types.

4) **Internet of Things (IoT):** The fourth enabler is the rapid growth of data from network-connected devices such as sensors. The EMC Digital Universe study predicts that 10% of all digital data will be

generated by network connected devices by the end of the decade. Today, network-connected devices generate about 2% of all digital data. Smart companies like GE, are leveraging this data to provide differentiated services. For example, a GE Wind Turbine contains about 20,000 sensors that generate 400 data points per second. The sensor data is analyzed in near real time to maximize the efficiency. This data is also stored and used for deeper analytics to improve maintenance and parts replacement. A GE Wind Turbine is more efficient and has higher availability enabled by next generation analytics. For more information, also refer to the IoT/CPS sections of this report.

5) **Data Analytics Tools:** The fifth enabler is a new set of analytics tools designed specifically to analyse large amounts of data, both structured (i.e. log data) and unstructured (images). Tools such as Hadoop, and Splunk were designed to analyse large data sets. These new analytics tools have been created through the Open Source community and have a low initial cost to at least get started. Next generation analytics has become affordable for big companies but smaller companies are also using these tools to find new revenue, and provide differentiated services.

2.2. EU Priorities & Landscape

In **July 2014** (10), the **European Commission** outlined a new strategy on Big Data, supporting and **accelerating the transition towards a data-driven economy in Europe**. The data-driven economy will stimulate research and innovation on data while leading to more business opportunities and an increased availability of knowledge and capital, in particular for SMEs, across Europe.

In the same strategic paper, the **European Commission admits** that **the European digital economy has been slow** in embracing the data revolution **compared to the USA** and also lacks comparable industrial capability. As a result, there are **fewer successful data companies in Europe than in the US** where large players have recognised the need to invest in tools, systems and new data-driven processes. However, significant new opportunities exist in a number of European industrial sectors (including health, smart factories (Industry 4.0), and agriculture) where the application of these methods is still in its infancy and global dominant players have not yet emerged. Moreover, since 2014, the figures have significantly improved for the EU but it still remains on 2nd position worldwide.

Monitoring Data Market, International Comparison, 2014, Units (000), EUR Million

N.	Name		EU	U.S.	Japan	Brazil
1.1	Number of Data Workers		6,102	10,457	3,344	1,031
2.1	Number of Data Companies		243,610	277,821	95,919	34,840
4.1	Value of the Data Market		€ 50,454	€ 103,935	€ 22,228	€ 5,289
4.2	Value of the Data Economy	Direct Impacts	€ 46,607	€ 99,398	€ 21,367	€ 5,289
		Backward Indirect impacts	€ 2,081	€ 4,536	€ 860	€ 217

Figure 2: Monitoring Data Market, International Comparison **(11)**.

2.2.1. EU Big Data Strategy

In order to tackle this, the European Commission has initiated a number of actions in order to support Big Data as a whole, and not only from the research oriented side. Moreover, to be able to seize these opportunities that Big Data presents and compete globally in the data economy, the EU (10):

- **Supports "lighthouse" data initiatives (in the shape of large-scale pilot actions)** capable of improving competitiveness, quality of public services and citizen's life
- **Develops enabling technologies**, underlying infrastructures and skills, particularly to the benefit of SMEs
- **Extensively shares, uses and develops its public data** resources and research data infrastructures
- Focuses **public R&I** on technological, legal and other bottlenecks
- Makes sure that the relevant **legal framework and policies** are data-friendly
- **Accelerates the digitisation of public administration** and services to increase their efficiency, and
- **Uses public procurement** to bring the results of data technologies to the market.

Adding to the above, the European Commission has designed and implements actions on various areas and topics, such as “soft” infrastructures, framework conditions, research and innovation topics, regulatory issues, etc.

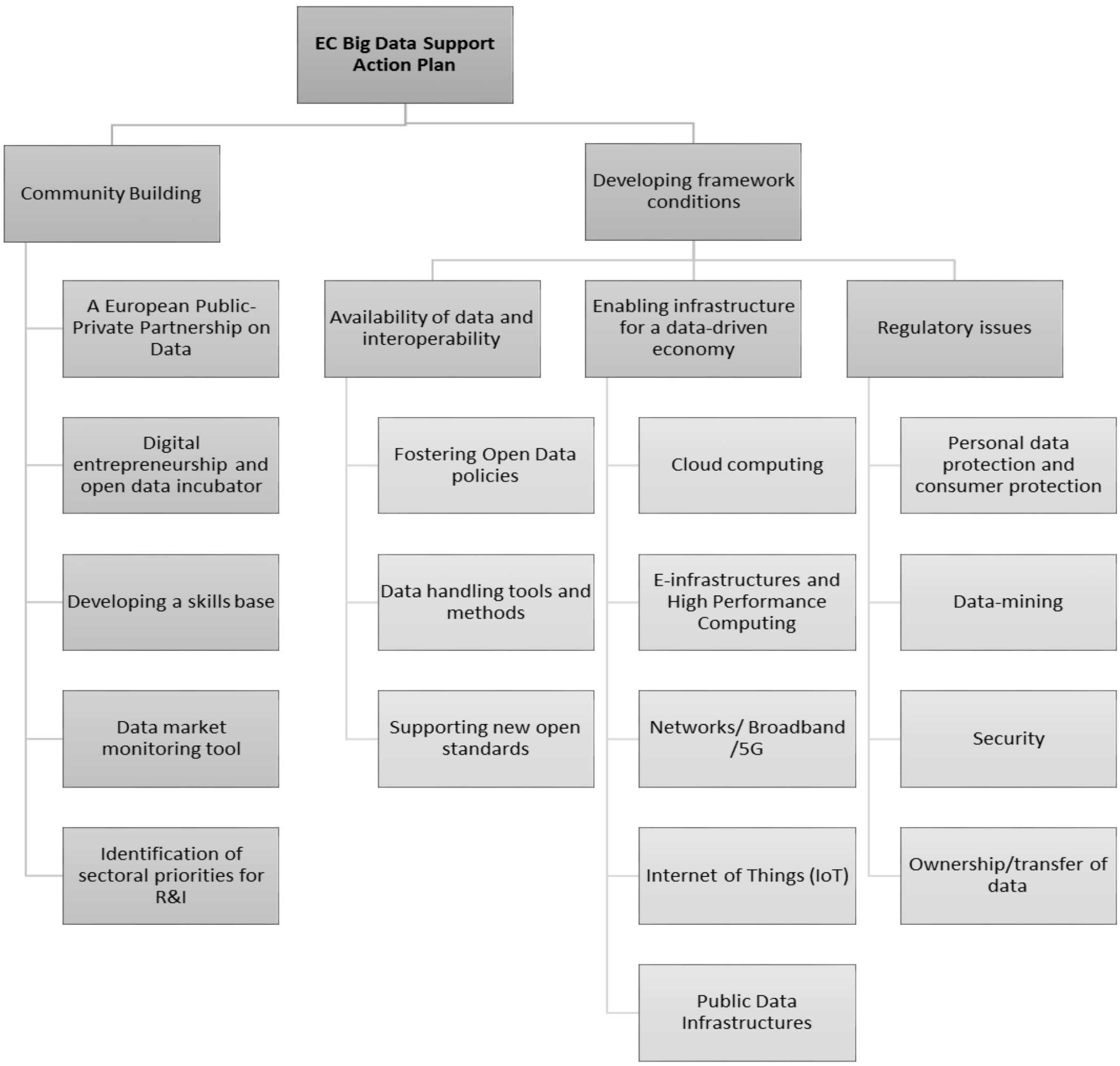

Figure 3: Big Data Support Action Plan.

2.2.2. Research & Innovation Priorities

Until today, the European Commission, through its main **Research & Innovation Funding Tool – H2020**, has issued a number of calls to support R&I projects in the Big Data area. More specifically, the **LEIT-ICT Work Programme** has issued the following calls and funded **55 projects** in total:

Table 1: Big Data related projects funded in H2020.

ICT-15-2014 - Big data and Open Data Innovation and take-up	13 Projects
ICT-22-2014 - Multimodal and Natural computer interaction	11 Projects
ICT-16-2015 - Big data - research	10 Projects
ICT-14-2016-2017 - Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation	7 Projects
ICT-15-2016-2017 - Big Data PPP: Large Scale Pilot actions in sectors best benefitting from data-driven innovation	2 Projects
ICT-17-2016-2017 - Big data PPP: Support, industrial skills, benchmarking and evaluation	1 Projects
ICT-18-2016 - Big data PPP: privacy-preserving big data technologies	4 Projects
ICT-35-2016 - Enabling responsible ICT-related research and innovation	7 Projects

The 2016 calls have been designed according to the **Strategic Research and Innovation Agenda**, issued by **Big Data Value Association**. The Big Data Value Association (**BDVA**) is the industry-led private counterpart to the EU Commission to implement the **Big Data Value Public Private Partnership programme (BDV PPP)**. BDVA has over **170 members** all over Europe with a well-balanced composition of large and small and medium-sized industries as well as research organizations[2]. BDVA, has identified a number of challenges and outcomes that need to be tackled through research and innovation activities, thus shaping research and innovation priorities for the European Commission to serve (12).

Table 2: Big Data Value Association Research Priorities.

Data Management	
Challenges	**Outcomes**
1. Semantic annotation of unstructured and semi-structured data **2.** Semantic interoperability **3.** Data quality **4.** Data management lifecycle **5.** Data provenance **6.** Integration of data and business processes **7.** Data-as-a-service	**1.** Techniques and tools for handling unstructured and semi-structured data. **2.** Languages and techniques for semantic interoperability such as standardized data models and interoperable architectures for different sectors enriched through semantic terminologies **3.** Languages, techniques and tools for measuring and assuring data quality, **4.** Methods and tools for a complete data management lifecycle **5.** Languages and tools for data provenance **6.** Methods and Tools for the sound integration of analytics results from data and business processes **7.** Data-as-a-service model and paradigm
Data Processing Architectures	
Challenges	**Outcomes**

[2] http://www.bdva.eu/

1. Processing of data-in-motion and data-at-rest 2. Decentralization 3. Heterogeneity 4. Scalability 5. Performance	1. Real-time architectures for data-in-motion 2. Decentralized architectures 3. Techniques and tools for processing real-time heterogeneous data sources 4. Scalable and dynamical data approaches 5. Efficient mechanisms for storage and processing
Data Analytics	
Challenges	**Outcomes**
1. Semantic and knowledge-based analysis 2. Content validation 3. Analytics frameworks & processing 4. Advanced business analytics and intelligence 5. Predictive and prescriptive analytics	1. Improved models and simulations 2. Semantic analysis 3. Event and pattern discovery 4. Multimedia (unstructured) data mining 5. Deep learning techniques for business intelligence
Data Protection	
Challenges	**Outcomes**
1. Robust data anonymity 2. Generic and easy to use data protection approach 3. Risk based approaches	1. Robust anonymisation algorithms 2. Protection against reversibility 3. Pattern hiding 4. Multiparty mining
Data Visualisation and User Interaction	
Challenges	**Outcomes**
1. Visual data discovery 2. Interactive visual analytics of multiple scale data 3. Collaborative, intuitive, and interactive visual interfaces 4. Interactive visual data exploration and querying in a multi-device context	1. Scalable data interactive visualization approaches and tools 2. Cross-platform data visualization frameworks 3. New paradigms for interactive visual data exploration, discovery, and querying 4. 3D visualization techniques and tools 5. Personalized end-user centric data interactive visualization mechanisms 6. Domain-specific data interactive visualization approaches 7. Techniques and tools for visualization of interrelated/linked data 8. Plug-and-play reusable components for data visualization

Throughout the project analysis that we undertook, we have been able to **categorise** most **EU funded projects**, from **Big Data H2020 related calls**, according to their technical or non-technical orientation. The following figure shows the resulting distribution:

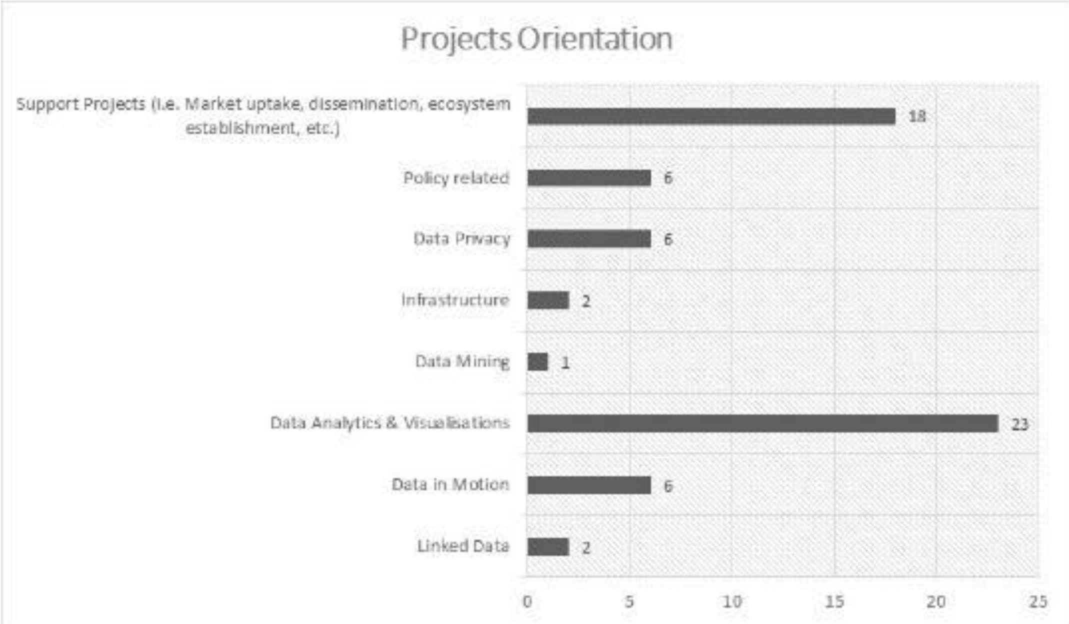

Figure 4: Big Data EU projects orientation.

An overall outcome that can be derived from the above analysis is that, so far, the **European Commission has funded** a large amount of research projects towards **Data Analytics and Visualisation**, which do not target a specific application sector (with some exceptions such as the two ICT-15-2016 lighthouses are a major invest; the two projects TT and DataBio which have a joint budget of 35 MEUR. Two more of such lighthouses are envisioned for the 2017 call). Moreover, the European Commission has allocated a large amount of resources to **fund projects aiming to create an ecosystem around Big Data** and therefore **enhance the creation of such an ecosystem** in order to commercially explore the value of Big Data for the benefit of the people.

2.3. US Priorities & Landscape

The US has made a first critical and important step towards **supporting the Big Data sector**, in 2012, by launching a $200 Million investment plan for R&D, for initiating projects to solve some of the Nation's most pressing challenges. The **"Big Data Research and Development Initiative"** (13) was supported by **six Federal departments and agencies** in order to improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data. The following departments and agencies were supporting this initiative:

1. National Science Foundation (NSF)
2. National Institutes of Health (HHS/NIH)
3. Department of Energy (DOE)
4. Department of Defense (DOD)
5. Defense Advanced Research Projects Agency (DARPA)
6. US Geological Survey – Big Data for Earth System Science (USGS)

Each of these agencies/departments have issued calls for funding Big Data projects in their area of interest such as **analytics, energy, security**, etc.

2.3.1. US Big Data Strategy

Following to the 2012 **"Big Data Research and Development Initiative"**, the US has launched a strategic plan in order to support that initiative by guiding Federal agencies as they develop and expand their individual mission-driven programs and investments related to Big Data. The **"Federal Big Data Research and Development Strategic Plan"** (14) was published on **May 2016**, by the **Networking and Information Technology Research and Development (NITRD)** Program, in order to guide Big Data research towards National priorities such as science, medicine, and security; ensuring the Nation's continued leadership in research and development; and enhancing the Nation's ability to address pressing societal and environmental issues facing the Nation and the world through research and development.

The plan is based on inputs from a series of Federal agency and public activities, and a shared vision:

"We envision a Big Data innovation ecosystem in which the ability to analyze, extract information from, and make decisions and discoveries based upon large, diverse, and real-time datasets enables new capabilities for Federal agencies and the Nation at large; accelerates the process of scientific discovery and innovation; leads to new fields of research and new areas of inquiry that would otherwise be impossible; educates the next generation of 21st century scientists and engineers; and promotes new economic growth."

The Plan is built around seven strategies that represent key areas of importance for Big Data research and development (R&D). Priorities listed within each strategy highlight the intended outcomes that can be addressed by the missions and research funding of NITRD agencies.

- → **Strategy 1:** Create next-generation capabilities by leveraging emerging Big Data foundations, techniques, and technologies
- → **Strategy 2:** Support R&D to explore and understand trustworthiness of data and resulting knowledge, to make better decisions, enable breakthrough discoveries, and take confident action
- → **Strategy 3:** Build and enhance research cyberinfrastructure that enables Big Data innovation in support of agency missions
- → **Strategy 4:** Increase the value of data through policies that promote sharing and management of data
- → **Strategy 5:** Understand Big Data collection, sharing, and use with regard to privacy, security, and ethics
- → **Strategy 6:** Improve the national landscape for Big Data education and training to fulfil increasing demand for both deep analytical talent and analytical capacity for the broader workforce
- → **Strategy 7:** Create and enhance connections in the national Big Data innovation ecosystem

2.3.2. Research & Innovation Priorities

Higher-level Priorities

The **"Federal Big Data Research and Development Strategic Plan"** sets a number of Research and Supportive priorities related to Big Data, which all Funding and Research agencies have to take into account before allocating resources to specific projects. Figure 5 illustrates these priorities:

Figure 5: US Big Data research priorities.

Moreover, **NSF** has identified a number of research and technological priorities (15) that need to be addressed through funding projects. The following table summarises all priorities:

Table 3: NSF Big Data research priorities.

Collection, Storage, and Management of "Big Data"	Data Analytics	Research in Data Sharing and Collaboration
Data representation, storage, and retrieval	Computational, mathematical, statistical, and algorithmic techniques for modelling high-dimensional data	Tools for distant data sharing, real-time visualization, and software reuse of complex data sets
New parallel data architectures, including clouds	Learning, inference, prediction, and knowledge discovery for large volumes of dynamic data sets	Cross-disciplinary model, information and knowledge sharing
Data management policies, including privacy and access	Data mining to enable automated hypothesis generation, event correlation, and anomaly detection	Remote operation and real time access to distant data sources and instruments
Communication and storage devices with extreme capacities	Information infusion of multiple data sources	
Sustainable economic models for access and preservation		

Last but not least, the **Executive Office of the President, President's Council of Advisors on Science and Technology (PCAST)**, released a report on Big Data Privacy, **"Big Data: A Technological Perspective"** (16) in which they define some key research themes that need to be addressed and supported through policy over the following years.

Table 4: Big Data research priorities according to PCAST.

Big Data Analytics	Big Data Infrastructure	Privacy Protection
Data mining	Cloud	Cryptography and encryption
Data fusion and information integration	Big Data Centres	Privacy Mechanisms
Image and speech recognition		
Social-network analysis		

R&I Project Analysis

By taking into account the above and by analyzing the available information from **130** Research and Innovation projects funded by **NSF, NIH** and **DARPA**, we have managed to identify what are the most critical technical subjects funded, in a project level. The table below, represents an "umbrella" taxonomy of these projects:

Table 5: Technological Domains for Big Data Projects.

Technological Domain(s)	Funded Projects
Data Analytics	**45**
Data Analytics, Privacy	**2**
Data Analytics, Visualization	**1**
Data Analytics, Processing	**1**
Big Data Curriculum	**1**
Biology	**1**
Chemistry, Analytics	**1**
Cloud Computing	**1**
Clustering	**1**
Collaboration, Interactions	**1**
Crowdsourcing, Visual Analytics	**1**
Data Mining	**1**
Datasets, Linguistics, Behavior	**1**
Decision Making	**2**
Economy	**2**
Eco-Routing	**1**
Healthcare	**1**
Image Data	**1**
Image Data, Cyber Security, Digital Entertainment	**1**
Infrastructure	**33**
Medicine, Analytics	**1**
Microbial Organisms, Genome	**1**
New Media, Social Networks, Training	**1**
Space	**1**
STEM (Data Analytics, E-Learning)	**5**
Video, Images, Multimedia Forensics	**1**
Visualization	**21**
Water Supply, Smart Cities	**1**
Grand Total	**130**

Big Data Regional Innovation Hubs

What is also important to mention is that in 2015, NSF launched **the Big Data Regional Innovation Hubs program (BD Hubs)** (17) to foster regional, cross-sector collaborations and multi-sector projects to foster innovation with Big Data. As a complement to the institutional gateways, the regional hubs provide the ability to engage with local or regional stakeholders, e.g., city, county, and state governments, as well as permit a focus on regional issues. These collaborative activities and partnerships play a critical role in building and sustaining a successful national Big Data innovation ecosystem. The four hubs, financed by the call are:

- **Northeast BD Hub**
- **South BD Hub**
- **Midwest BD Hub**
- **West BD Hub**

The figure below[3] shows the geographical representation of the four Hubs:

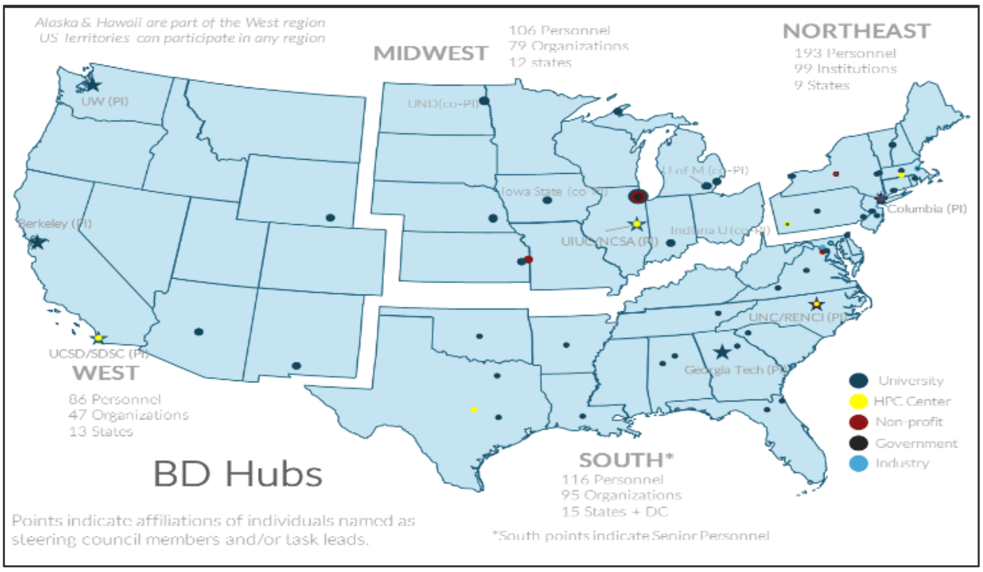

Figure 6: Big Data US Hubs[4].

More specifically, all four hubs have identified their technological and application priorities related to their exact needs, as they have been identified by the partners. The table below shows a detailed analysis of all Rings (Parallel priorities) and Spokes (Vertical priorities) for each Hub.

Table 6: Big Data Hubs Priorities.

Hubs	Rings (Horizontal Priorities/Technical)	Spokes (Vertical priorities/Application Sectors)
Northeast BD Hub	1.Data Literacy 2.Data Sharing 3.Ethics 4.Privacy & Security	Education Finance Cities & Regions Health Energy Discovering Science
South BD Hub	1.Sharing & Infrastructure	Health Disparities & Analytics

[3] BIG DATA REGIONAL INNOVATION HUBS &SPOKES: Accelerating the Big Data Innovation Ecosystem, Fen Zhao, Staff Associate, Strategic Innovation, CISE Directorate, Office of the Assistant Director , National Science Foundation

[4] Graphic taken from: BIG DATA REGIONAL INNOVATION HUBS &SPOKES: Accelerating the Big Data Innovation Ecosystem, Fen Zhao, Staff Associate, Strategic Innovation, CISE Directorate, Office of the Assistant Director , National Science Foundation

	Development 2.Economic Modelling, Security & Policy	Coastal Hazards Industrial Big Data Materials & Manufacturing Habitat Planning
Midwest BD Hub	1.Data Science 2.Resources: Data, Tools & Services 3.Education Training & Workforce Development	Business Analytics Metropolitan Sciences Food-Energy-Water Digital Agriculture Transportation Advanced Manufacturing Health & Biomedical Sciences Network Sciences
West BD Hub	1. Accelerating Innovation for National Priorities 2.Strengthening Our Data Science Community 3.Fostering Cross-Cutting Teams	Metro Data Science Precision Medicine Managing Natural Resources and Hazards Big Data Technology Data-enabled Scientific Discovery and Learning

What can be seen from this table is that some application sectors are critical for all Big Data Hubs. More specific, **Health, Smart Cities, Manufacturing** and **Hazard prevention** are the most common ones.

2.4. Postgraduate Education on Big Data

There are an average of 1,894 Big Data jobs posted on Dice (18) on any given day, Dice spokeswoman Rachel Ceccarelli said. That's up 41 percent year on year; two years ago, only 438 such jobs were listed (18). How those jobs would be filled, however, wasn't entirely clear. Until recently, big-data education programs were few and far between. In the last few years, graduate, undergraduate, and professional-education programs have begun popping up to address this gaping need. Now that they've begun to emerge, the demand is considerable. In order to capture the overall essence of education on Big Data related themes, we have gone through an extensive desktop research of a large volume of Master courses related to Big Data, mainly from EU and US universities. In addition we also documented some Master degrees coming from universities located in Asia and Oceania, in order to create a "rough" comparison among all regions, although this is a task that is not related to this current study.

Table 7: No of Big Data related Masters per region.

Continent	No of Big Data related Masters
Asia	10
Europe	114
North America	53
Oceania	3
Grand Total	180

From our desktop research, we discovered that there are at least 167 Masters courses offered by EU and US universities. From these 167 degrees, 114 are offered by EU universities and 53 by US universities. What needs to be mentioned at this stage is that these Master courses are directly related to Big Data according to each universities' statements about their programme and have been categorized as Big Data programmes in the StudyPortals international database (19).

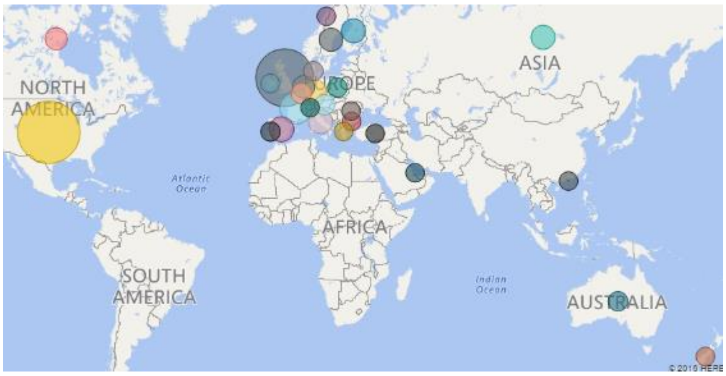

Figure 7: Big Data related Masters geographical distribution.

An impressive finding from this analysis is that there is a clear diversification of titles among all these courses. The most common title among courses is Data Science (28) and Data Analytics (13), while only 17 of them contain the term “Big Data” in their titles.

Table 8: No of Common Big Data Master Titles.

Title	No of titles
Data Science	28
Data Analytics	13
Information Technology	6
Business Analytics	4
Computer Science	3
Big Data Analytics	2
Analytics	2
Data Mining & Knowledge Management	2
Data Science for Management	2
Computational Science	2
Data Studies	2
Data Telecommunications and Networks	2
Database Systems	2
……………	…….

A valuable outcome of this exercise is that, as with any new field, colleges and universities are catching up. There are more than a few different names, including Analytics, Data Science, Business Analytics, and every possible combination of those words, and they’re offered by all kinds of departments, from engineering and computer science to business and marketing. What matters, more than the name, is that the programs find the right balance between technical computer skills, business and marketing knowledge, and statistical analysis. Most programs are interdisciplinary, because it takes the right combination of experts to teach so many different areas.

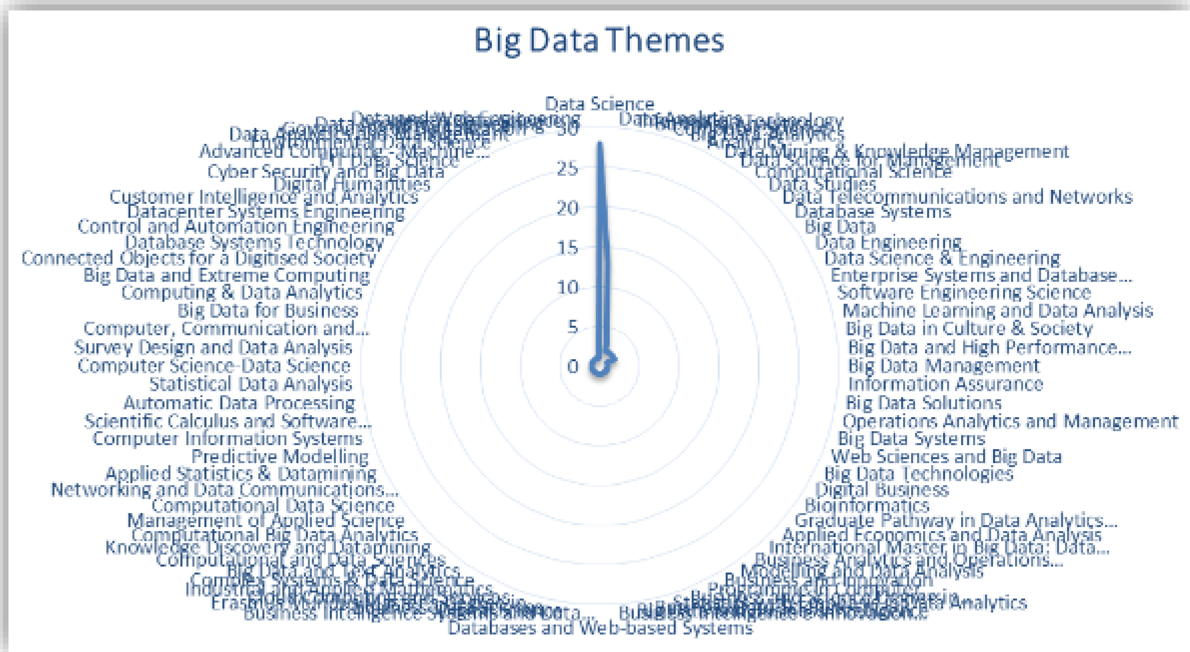

Figure 8: Big Data themes.

2.5. Application Sectors

2.5.1. EU Application Sectors

The European Commission has already acknowledged (4) and embraced that generating value at the different stages of the data value chain will be at the centre of the future knowledge economy. Good use of data can bring opportunities also to more traditional sectors such as transport, health and manufacturing. Improved analytics and processing of data, especially Big Data, will make it possible to:

- transform Europe's service industries by generating a wide range of innovative information products and services;
- increase the productivity of all sectors of the economy through improved business intelligence;
- better address many of the challenges that face our societies;
- improve research and speed up innovation;
- achieve cost reductions through more personalised services
- increase efficiency in the public sector.

Following from the above, the European Commission has already defined which **domains are critical for EU** and is supporting Big Data Research and Innovation projects in all of the following domains, known as **Societal Challenges** (20):

Health, demographic change and wellbeing;

Food security, sustainable agriculture and forestry, marine and maritime and inland water research, and the Bioeconomy;

Secure, clean and efficient energy;

Smart, green and integrated transport;

Climate action, environment, resource efficiency and raw materials;

Europe in a changing world - inclusive, innovative & reflective societies;

Secure societies - protecting freedom and security of Europe and its citizens.

**All images above are courtesy of the Big Data Europe Project (https://www.big-data-europe.eu)*

Furthermore, the European Commission, has indicated a number or priority sectors in the **ICT-15-2016-2017** (21)**.** These priority sectors are **health, energy, environment, earth observation, geospatial, transport, manufacturing, finance and media**.

What is also critical to mention at this stage is that through our analysis, we have identified which and how many projects the European Commission has funded under each societal challenge. Until December 2016, the Commission has funded **14** projects in **Health, demographic change and wellbeing**, followed by **9** projects in **Smart, green and integrated transport** and **7** in **Europe in a changing world - inclusive, innovative and reflective societies** and **7** in **Secure societies - protecting freedom and security of Europe and its citizens** respectively. Also, it is important to point out that **6** projects have been funded in the **Media & Social Media Industry**, showing the importance of Social Media Data from a business and sociological perspective. Adding to the above, **13** projects have been funded dealing with **Horizontal technological and non-technological themes**, which affect and can be applied in in more than one sectors. More information can be found in Figure 9 below.

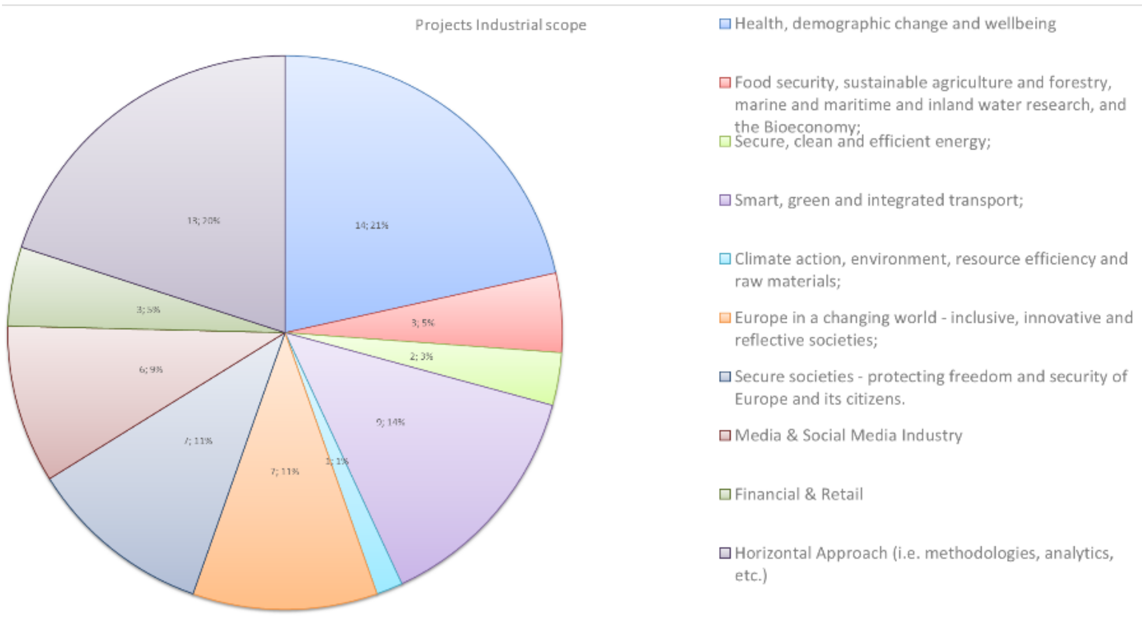

Figure 9: Big Data Projects Scope.

2.5.2. US Application Sectors

Higher-level Application Sectors

A number of strategic documents have been analysed in order to capture what are the application sectors (industrial sector) which US are mainly targeting. The following documents have been used for this analysis: NIST Big Data Interoperability Framework (22), NSF Big Data Research priorities (15), Dr. John P. Holdren, (Assistant to the President for Science and Technology and Director of the Office of Science and Technology

Policy) statements (13) and the "Big Data: A Technological Perspective" (16) , released by the Executive Office of the President, President's Council of Advisors on Science and Technology.

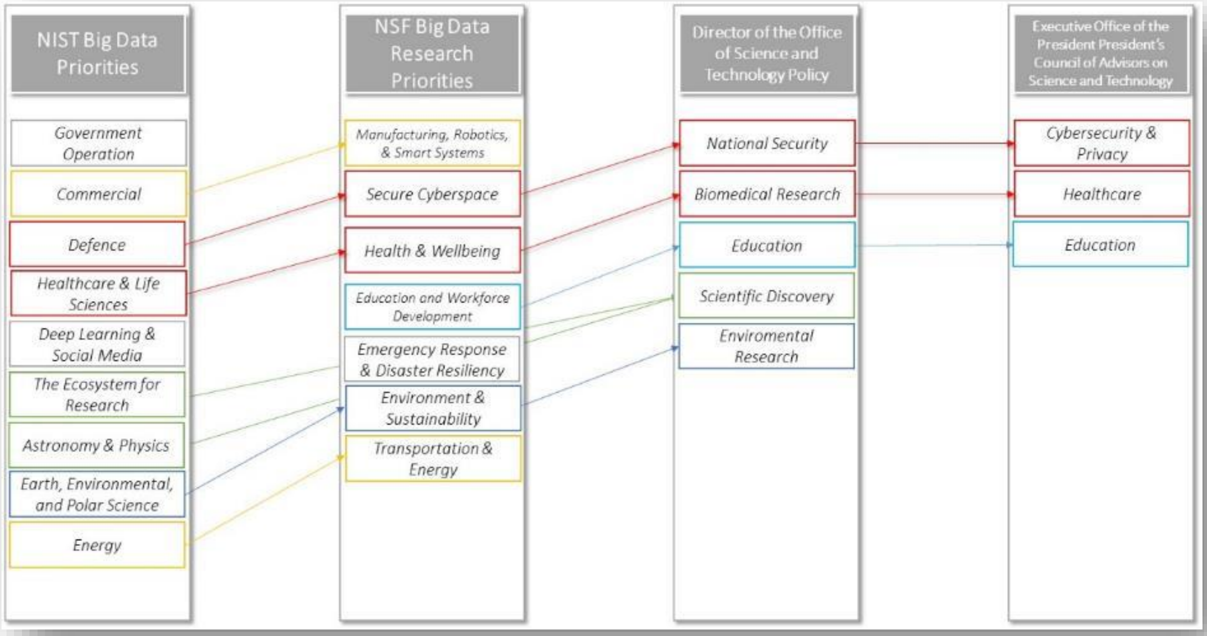

Figure 10: Big Data relevant Application Sectors (US).

What can be shown by the above figure is that sectors such as **Defence, Healthcare** and **Education** are the most important application sectors, as defined by the policy**.**

R&I Project Analysis

By taking into account the above and by analysing **153** Research and Innovation projects funded by NSF, NIH and DARPA, we have identified which application domains have been funded the most. From the table below, one can see that **Computer Science** (including Data Analytics, Data Processing, Data Visualisation) domain and is the most funded area. This happens mainly because a number of projects do not focus on only one application or industrial area but can be applied to a number of different sectors.

Table 9: US Big Data R&I projects analyzed by application sectors.

Technological Area	Funded Projects
Bio	1
Computer Science	**100**
Crowd Sourcing	1
Education	6
Electronics	1
Facilities Management	1
Finance	1
Healthcare	**26**
Human Networks	1
Infrastructure	2
Life Sciences	1
Linguistics	1

Materials	1
Media	3
Networks	4
Relied Logistics	1
Space	1
STEM	1
Grand Total	**153**

2.6. Conclusions

Both regions (EU & US) are already implementing solid policies related to Big Data customised specifically to their needs and challenges that they face and urgent to tackle. However, despite several similarities, critical differences exits making each policy unique for each region

2.6.1. Similarities & Differences at Design & Implementation Level

Similarities between policies in both regions exist and are evident from the policy reports generated by the federal government or agencies in charge of these policies.

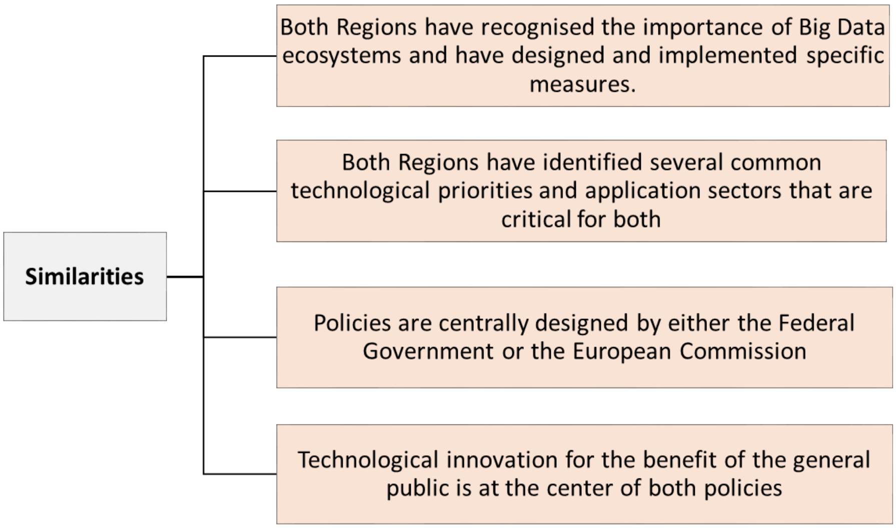

Figure 11: Significant Big Data Similarities.

Moreover, a number of significant differences exist, when policies from the two regions are being compared. Most differences are related to the implementation model between the two regions.

Figure 12: Significant Big Data Differences.

Specific information on similarities and differences in technological and application domains can be found in the below section.

2.6.2. Similarities & Differences in Technology and Application Domains

Technology Domains

From our analysis, throughout the whole document and by taking into account input from several sources such as Expert Group Members, other PICASSO deliverables and Policy papers, common research and innovation topics have been identified. **Data Analytics**, **Data Protection & Privacy**, **Data Processing Architecture** and **Data Visualisation and User Interaction** are the technological priorities that are in the heart of policies and strategies for both regions. However, each region also gives emphasis to additional technological domains such Big Data Infrastructures and Ethics (for the US) and Data Management (for the EU).

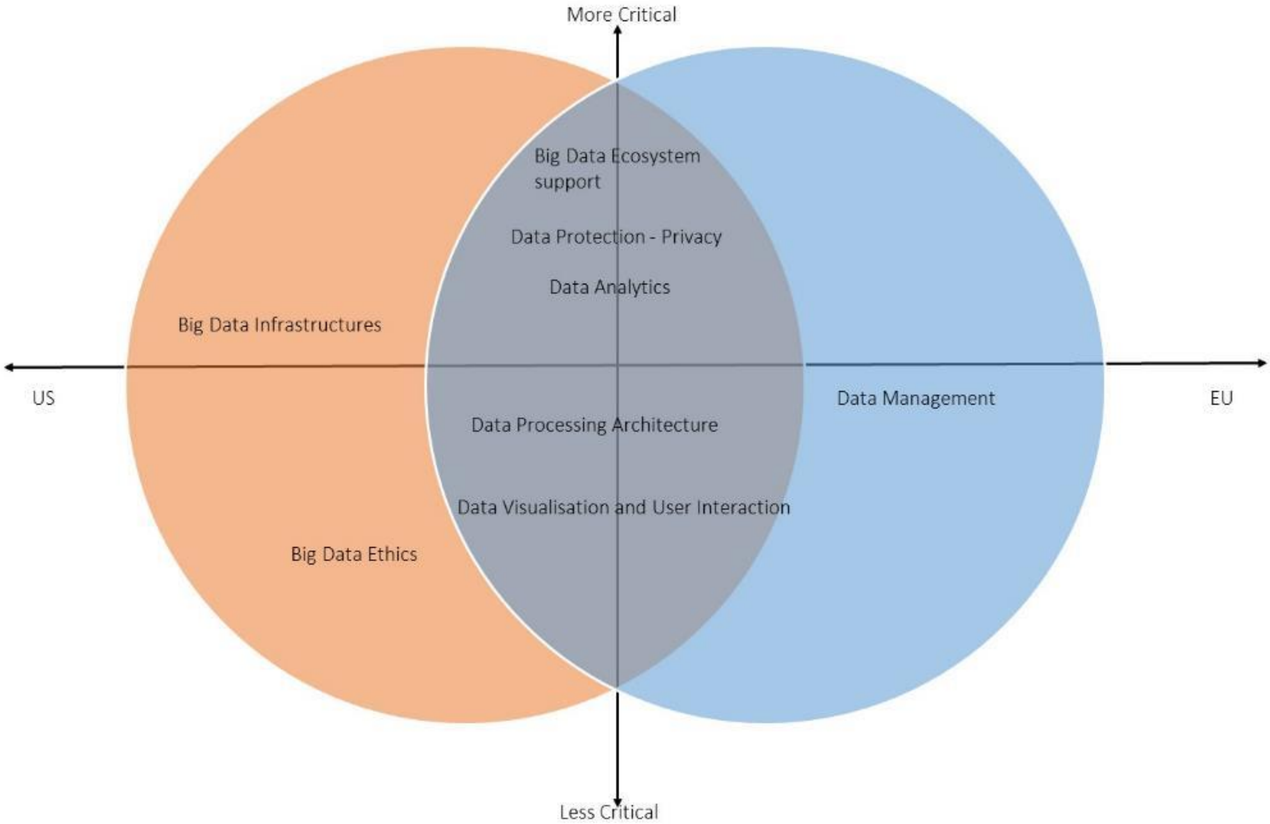

Figure 13: EU-US Common Big Data Priorities.

Application Domains

From our analysis, it can be seen that some application sectors are of extreme importance, both for the US and for the EU. **Health**, **Security**, **Smart Energy** and **Environment** are the most critical sectors for both regions. Moreover, sectors such as **Smart Transportation**, **Government**, **Manufacturing (Smart Production)**, **Finance** and **Agriculture** are also under a common microscope, although it seems that these sectors receive less support. Last but not least, what can also be seen is that the US has identified the domain of **Natural Resources and Hazards** as extremely important while the EU has given a significant attention and support to the **Media** domain.

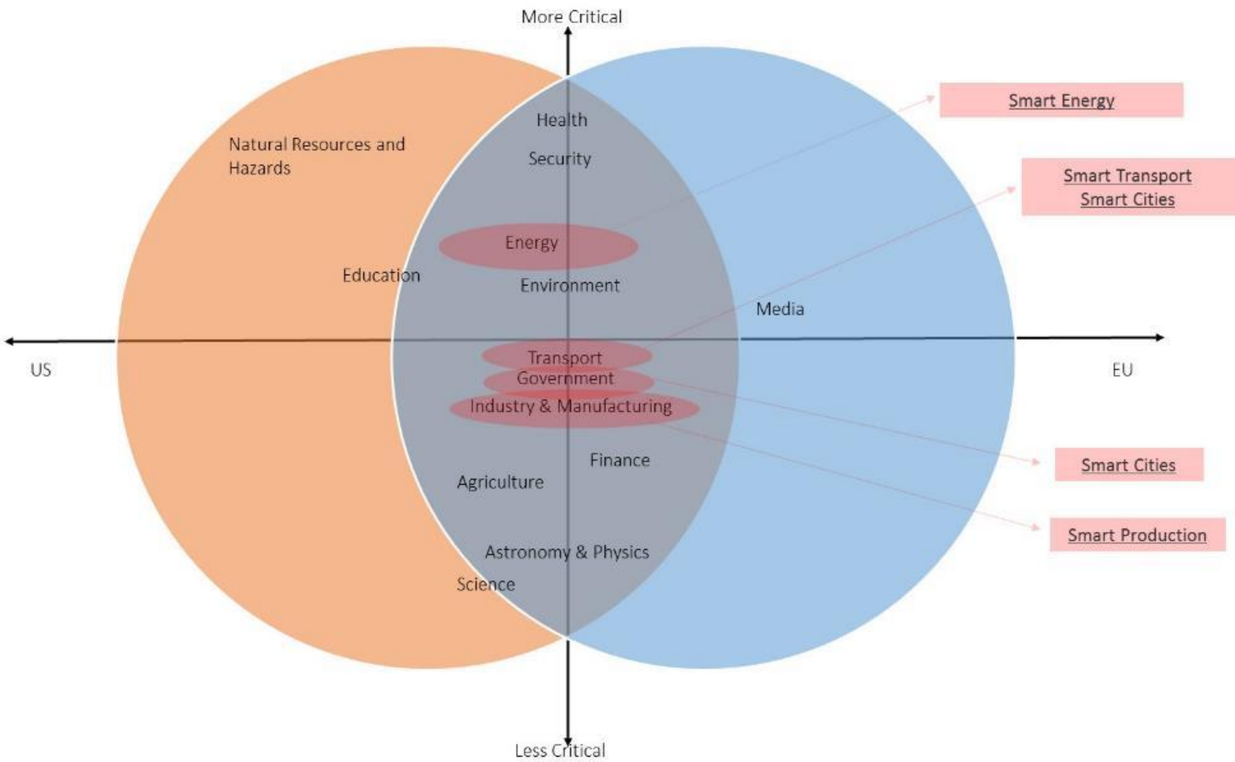

Figure 14: US & EU Common Big Data Application Sectors.

3. Technology Themes for EU-US Collaboration

Throughout the process of generating the Big Data opportunity report, a number of policy and technological reports have been taken into account in order to shape a clear set of technological opportunities for collaboration between the EU and US. In addition feedback from the Big Data Expert Group has been also taken into account for formulating a concrete picture. From the elaboration of this feedback, we have been able to identify a number of technological opportunities that are of highest priority of both regions. What needs to be mentioned is that set of opportunities will be updated in the forthcoming months taking into account additional feedback and input from external experts and policy makers willing to contribute. The most important technological opportunities are:

1. Interoperability and Standardization
2. Adding a semantic layer to Big Data technology
3. Integrating Linked Data and Big Data technology
4. Enable discovery of deeper, fresher insights from all enterprise data resources
5. Improve efficiency, effectiveness, and decision-making
6. Facilitate more timely, agile response to business opportunities, threats, and challenges
7. Provide a single view of diverse data resources throughout the business chain
8. Support tighter security, protection, and governance of data throughout its lifecycle
9. Improve the scale, efficiency, performance, and cost-effectiveness of data/analytics platforms

4. Opportunities and Barriers for EU-US Collaboration in Technology Sectors

This chapter gives a brief overview of the EU-US funding and collaboration environments in section 4.1 and summarizes barriers that may hamper EU-US collaboration in section 4.2. Section 4.3 provides recommendations of concrete opportunities that were identified as the most promising mechanisms for technological collaborations on the R&I themes presented in chapter 3, and the recommendations will be refined, validated, and promoted during the remainder of the PICASSO project.

The contents of sections 4.1 and 4.2 were created by the IoT/CPS Expert Group (with inputs from the Big Data and 5G Expert Groups), and the contents of 4.3 are based on these sections. Additional sources include inputs and pointers from numerous external experts from EU and US funding agencies, industry associations, and academia that were interviewed by the IoT/CPS Expert Group, the analyses presented in section 2, the PICASSO reports (2) and (1), materials and feedback by the EU projects *DISCOVERY* (23), *BILAT USA 2.0*, *BILAT USA 4.0, CPS Summit*, and *TAMS4CPS*, and the interactive PICASSO IoT/CPS webinar that was held on February 2, 2017.

4.1. The EU-US Funding and Collaboration Environment

4.1.1. EU and US Funding

The US R&I funding landscape is structurally very different to the EU landscape. EU-level funding is mostly centralized and is realized via major funding programmes such as *H2020*, the *ECSEL Joint Undertaking*, and *ERA-NET* (which focuses on pooling and coordinating funding of EU member states for EU joint calls) that provide EU-wide frameworks for R&I funding activities, covering all levels from fundamental over translational and applications-oriented research to knowledge transfer, innovation, and commercial deployment. In the US, the funding landscape is much more fragmented. Research and innovation is mostly funded by federal research programs that are set up by different federal agencies and that reflect directly the government's priorities and interests (3). Research funding is also available at the state level, but state funding normally focuses on specific local needs and is not usable for international collaboration.

Applications-oriented R&I funding is often provided directly by companies or industry-led associations to partnering research institutions in the form of grants, with a focus on short-term returns. Initiatives such as *H2020* or dedicated programs by US agencies usually focus on funding relatively large R&I projects, for which it usually takes months between the funding application and the start of work. On the other hand, direct funding by industry often focuses on a smaller scope and a relatively quick (e.g. within a few weeks) start of work after initial funding talks.

A major contact point in the federal US funding landscape in the areas of IT, computing, networking, and software is the *Networking and Information Technology Research and Development* (*NITRD*) Program, a multi-agency program that coordinates the funding of all federal agencies in this area. It has specific contact points that coordinate research across all agencies, such as CPS research and wireless communications incl. 5G.

The *National Science Foundation* (*NSF*) exclusively funds basic research and has a major CPS research program with more than 350 funded projects, plus funding for IoT research. The NSF has explored collaborations with the EU in the past, most successfully in the areas of environmental health and safety technology. In addition, there are several bilateral cooperation agreements with EU member states, such as the US-German IoT/CPS program, and interview partners have indicated significant interest in future programs for EU-US collaboration in the areas of IoT and CPS. The NSF will not cover EU costs, but it may cover costs for EU researchers visiting the US and vice versa. The NSF has already shown interest on collaborations on low-TRL research and is a good fit because it has a major initiative in CPS, in which energy aspects are of particular interest.

The NSF is a leader in supporting Big Data research efforts as well. These efforts are part of a larger portfolio of Data Science activities. NSF initiatives in Big Data and Data Science encompass research, cyber-infrastructure, education and training, and community building. In addition to funding the Big Data solicitation, and keeping with its focus on basic research, NSF is implementing a comprehensive, long-term strategy that includes new methods to derive knowledge from data; infrastructure to manage, curate, and serve data to communities; and new approaches to education and workforce development. "Big Data" is a new joint solicitation supported by the National Science Foundation (NSF) and the National Institutes of Health (NIH) that will advance the core scientific and technological means of managing, analyzing, visualizing, and extracting useful information from large and diverse data sets. This will accelerate scientific discovery and lead to new fields of inquiry that would otherwise not be possible. NIH is particularly interested in imaging, molecular, cellular, electrophysiological, chemical, behavioural, epidemiological, clinical, and other data sets related to health and disease.

In the 5G area, the NSF coordinated the $400 million Advanced Wireless Research Initiative launched in 2016. As a first step, a Project Office for establishing the *Platforms for Advanced Wireless Research* (*PAWR*) has been created. The NSF has explored collaborations with the EU in the past, most successfully in the areas of health and safety technology. In addition, there are several bilateral cooperation agreements with EU member states, e.g. with Finland and Ireland. Potential collaboration mechanisms involving the NSF are e.g. joint workshops and mirrored calls.

The *National Institute of Standards and Technology* (*NIST*) is an important, more applications-oriented player in ICT funding (with a focus on supporting their own labs, not academia in general) and is active in a variety of technological areas and application sectors. In particular, it has a *Cyber Physical Systems Program* and a *CPS Public Working Group* that is currently developing a CPS framework (24), and its wireless networks division has a *5G & Beyond Program* and coordinates the *5G Millimeter Wave Channel Model Alliance* as well as working group developing the *Future Generation Communications R&D Roadmap*. NIST has already shown significant interest in the PICASSO work.

The parent organization of NIST, the *Department of Commerce* (*DoC*), also promotes other activities in the IoT/CPS domain. In 2016, the DoC has set as a policy priority to engage with the EU Digital Single Market initiative in the area of the free and open internet, and it also promotes activities in the telecommunications domain. The *National Telecommunications and Information Administration* (*NTIA*) of the DoC focuses on expanding broadband internet access and expanding the efficient use of spectrum, and it has recently published a "green paper" that reviews the current technological and policy landscape for the IoT and that highlights potential benefits and challenges, and possible roles for the federal government in fostering the advancement of IoT technologies in partnership with the private sector (25). In this paper, the NTIA promotes a globally connected, open, and interoperable IoT environment and recommends governmental support for US industry initiatives, greater collaboration between (private) standards organizations, the crafting of balanced policy and building coalitions, the enabling of infrastructure availability and access, and the promotion of technological advancement and market encouragement. The NTIA sees the role of government in the promotion of robust interagency coordination, public-private collaboration, and international engagement, while avoiding over-regulation that could stifle IoT innovation. International collaboration is encouraged across a range of activities and topics, including a consistent common policy approach for the IoT, cross-border data flows, privacy, and cyber-security, based on formal dialogues with top international partners on digital economy issues.

Other agencies that are potentially of interest as US partners for PICASSO collaboration mechanisms are the *Department of Energy* (*DoE*) that supports more applications-oriented research and development in areas such as clean energy, environmental cleanup, climate change, and other areas, has a strong track record in collaborations with European universities and research centers, and has shown interest in topics such as grid modernization and integrating renewables, the *Department of State* (*DoS*), the *Department of Homeland Security* (*DHS*), *Department of Defense* (*DoD*) agencies such as *DARPA*, the *Air Force Office of Scientific Research*, the *Army Research Office*, and the *Office of Naval Research*, and US foundations such as *Gordon and*

Betty Moore Foundation and the *Blavatnik Family Foundation*. In addition, the *TAMS4CPS* project found that US national labs (such as Sandia) may be suitable contacts regarding funding for collaborations on more applications-oriented research.

The DoD is "placing a big bet on big data" investing approximately $250 million annually (with $60 million available for new research projects) across the military departments in a series of programs that will:

- Harness and utilize massive data in new ways and bring together sensing, perception and decision support to make truly autonomous systems that can maneuver and make decisions on their own.
- Improve situational awareness to help warfighters and analysts and provide increased support to operations. The Department is seeking a 100-fold increase in the ability of analysts to extract information from texts in any language, and a similar increase in the number of objects, activities, and events that an analyst can observe.

The *Defense Information Systems Agency* (*DISA*) offers a cloud-based set of solutions that enables the collection of large amounts of data from across the DoD Information Networks (DODIN) and provides the analytics and visualization tools to make sense of the data. The set of solutions is called *Cyber Situational Awareness Analytical Capabilities* (*CSAAC*) and is available on both *the Nonsecure Internet Protocol Router Network* (*NIPRNet*) and *Secret Internet Protocol Router Network* (*SIPRNet*). By using CSAAC, DoD network analysts and operators have a broader and more comprehensive view of DODIN activity than ever before. CSAAC enables informed decision making and enhances the overall security posture of DoD networks.

According to Deltek Principle Research Analyst Alex Rossino's new calculations, the *Defense Advanced Research Projects Agency*'s (*DARPA*'s) budget requests for big data research and development programs will grow by 39 percent in fiscal year 2016. In the past two years, DARPA's big data investments - which focus on advanced algorithms, analytics and data fusion, among other things - have spiked 69 percent, growing from just under $97 million in FY 2014 to more than $164 million in FY 2016. In addition, in 2012, DARPA initiated the 3-year $100M XDATA program to develop computational techniques and software tools for processing and analyzing massive amounts of mission-oriented information for Defence activities. Furthermore, to encourage future collaboration and innovation across the mathematic, computer science and visualization communities, DARPA open sourced the solutions for the general public.

The DoD and DARPA also support for example a spectrum collaboration challenge, where competitors are reimagining spectrum access strategies and developing new paradigms of collaborative decision-making where radio networks will autonomously collaborate and reason about how to share radio spectrum.

The *Department of Energy* will provide $25 million in funding to establish the *Scalable Data Management, Analysis and Visualization* (*SDAV*) Institute. Led by the Energy Department's Lawrence Berkeley National Laboratory, the SDAV Institute will bring together the expertise of six national laboratories and seven universities to develop new tools to help scientists manage and visualize data on the Department's supercomputers, which will further streamline the processes that lead to discoveries made by scientists using the Department's research facilities. The need for these new tools has grown as the simulations running on the Department's supercomputers have increased in size and complexity. Moreover, the DoE, with the support of partners and allies, has created the SEED Platform Collaborative to help put big data to work on one of the biggest problems in the global effort against the negative effects of climate change - the waste of energy in big buildings. The new *Standard Energy Efficiency Data* (*SEED*) Platform Collaborative creates a remarkable three-year partnership with regional and local governments to help them collect and manage data that tracks energy use in buildings, set aggressive goals for energy efficiency in them, and transform cities and regions into energy-saving leaders.

Other governmental agencies that support Big Data R&I are the *National Institutes of Health* (*NIH*) and the *US Geological Survey* (*USGS*). The NIH has announces that the world's largest set of data on human genetic variation – produced by the international 1000 Genomes Project – is now freely available on the *Amazon Web*

Services (*AWS*) cloud. At 200 terabytes – the equivalent of 16 million file cabinets filled with text, or more than 30,000 standard DVDs – the current 1000 Genomes Project data set is a prime example of big data, where data sets become so massive that few researchers have the computing power to make best use of them. AWS is storing the 1000 Genomes Project as a publically available data set for free and researchers only will pay for the computing services that they use. The USGS has financed, through its John Wesley Powell Center for Analysis and Synthesis, a number of projects on Big Data in order to improve its understanding of issues such as species response to climate change, earthquake recurrence rates, and the next generation of ecological indicators. Funding was providing scientists a place and time for in-depth analysis, state-of-the-art computing capabilities, and collaborative tools invaluable for making sense of huge data sets.

Non-governmental actors play a major role in translational and application-oriented R&I, collaboration, and funding in the US and the EU, and are the main drivers in for applications-oriented ICT advancement. Non-governmental actors include multi-national companies (which have an inherently international point of view and are particularly dominant in the IoT sector), and industry-led associations and standardization bodies such as the *Industrial Internet Consortium* (*IIC*), the *International Council on Systems Engineering* (*INCOSE*), the *Smart Manufacturing Leadership Coalition* (*SMLC*), the *Object Management Group* (*OMG*), the *National Coalition for Advanced Manufacturing* (*NACFAM*), the *Conference of European Directors of Roads* (*CEDR*), and others. Our discussions with representatives from industry-led associations have shown that companies and associations are promising potential partners for future EU-US collaborations, also because they are less affected by governmental policy than federal agencies.

4.1.2. EU-US Collaboration

To our knowledge, no specific calls are currently published for foreigners' participation within H2020 (3). According to research conducted by the *BILAT USA 2.0* project, "nearly one-quarter of individual organisations' policy measures provide funds to other countries as long as the leading organisation is a U.S.-based university or other research institution. About 40% of the measures do not provide funding to non-U.S. institutions. The remaining 40% have specific pre-requisites for allowing receipt of U.S. funds by third countries".

In a recent study, the *DISCOVERY* project (23) analyzed the participation rate of US partners in H2020 projects and found that out of 52 running H2020 projects with US participation (with starting dates before June 2016), only three projects focus on IoT topics, and none on CPS topics, while the majority of projects are in the scope of personal healthcare (due to an existing bilateral agreement on health R&I between the EU and the US). Two of the three IoT projects are within the scope of the *Future Internet Research & Experimentation* (*FIRE*) European initiative, which previously participated in a successful EU-US collaboration with its US counterpart, the NSF-funded *Global Environment for Networking Innovations* (*GENI*) program. The collaboration focused on the organization of joint thematic workshops and the exchange of personnel between the EU and the US.

On the EU side, there are several examples where specific programmes opened project participation, and even funding in some cases, to US partners. The *Conference of European Directors of Roads* (*CEDR*), a consortium of public national road authorities or equivalents of European countries that focuses on applications-oriented research on road transportation topics, opened a recent call for projects to US participants[5], including the possibility of receiving funding from *CEDR*. The goal of this collaboration effort was to gain access to leading research experience from the US. The ERA-NET instrument that supports public-public partnerships for joint, transnational activities between EU member states (possibly with EU-level funding contributions) recently organized a workshop with the goal of making US and Brazilian funding agencies aware of the ERA-NET work and to discuss collaboration opportunities[6]. Follow-up activities are planned. In addition, selected ERA-NET

---

[5] http://www.cedr.eu/download/other_public_files/research_programme/call_2016/CEDR-Call-2016-Information-Dec-2016.pdf
[6] https://www.b2match.eu/jpisgoglobal2016

programmes complement EU member state funding with external initiatives, including US-based funding. An example is *the Infrastructure Innovation Programme* (*Infravation*) for road infrastructure innovation[7].

Many multi-national companies (which by definition have subsidiaries in different countries that often collaborate) and industry-led associations have a strong track record of international collaboration and are open to participating in EU-US collaboration efforts. As an example, the *Industrial Internet Consortium* (*IIC*) is a global initiative that promotes the growth of the industrial IoT by bringing together partners from around the world, coordinating ecosystem initiatives, and bridging between regional initiatives (such as *Industrie 4.0* in Germany). Particular focus is currently placed on the 27 joint testbed initiatives[8], involving partners from many different countries. These joint testbeds provide realistic industrial environments for joint pre-competitive R&I projects so that new technologies, applications, products, services, and processes from different partners can be initiated, developed, and tested. As an example, the first of these testbeds, *Track&Trace*, was established appr. 2 years ago, is located in Germany, involves partners from the EU, the US, and India, and focuses on the development and testing of future smart, hand-held tools in manufacturing, maintenance, and industrial environments.

While collaboration initiatives between governmental agencies (such as the NSF and the EC) involve only few large organizations and are usually coordinated and set up internally, establishing collaborations between many different actors (such as government agencies on one side and industry-led associations, or even single large enterprises and SMEs on the other side) may require significant coordination and support activities. An example of a non-profit organization that specializes on this kind of match-making is the *Intelligent Manufacturing Systems* (*IMS*) Global Research and Business Innovation Program[9], which is partly funded by the EC. The program aims to integrate and connect US manufacturing industries and associations with EC programmes (where EC-foreign partners must provide their own funding). They focus on two services, direct matchmaking to set up R&I projects with partners from the member states, and thematic project clustering programmes for existing projects that provide collaboration support, such as the organization of workshops for international exchange.

4.2. Barriers

This section summarizes major barriers that must be overcome to implement successful EU-US collaborations. Most of these barriers have been identified in discussions within the IoT/CPS Expert Group and personal interviews done by the IoT/CPS Expert Group with external experts. Additions were provided by the Big Data and 5G Expert Groups.

4.2.1. Structural Differences in Funding Environments

As described in section 4.1, the US R&I funding landscape is structurally very different to the EU landscape along several dimensions.

First, EU-level funding builds on centralized framework programmes that do not have a counterpart in the fragmented US landscape. There are no overarching US or EU programmes currently that focus on closing the gap between the funding structures the gap between centralized EU and decentralized US funding, although programs such as *Intelligent Manufacturing Systems* (*IMS*, see previous section) provide bridging services for specific sectors. It seems unlikely that such overarching programmes are viable due differences in policy and

---

[7] http://www.infravation.net

[8] http://www.iiconsortium.org/test-beds.htm

[9] http://www.ims.org

due to the large administrative overhead that comes with the coordination of many different agencies and companies.

Second, different US funding agencies target specific technology readiness levels. The NSF focuses solely on basic research while other agencies (such as NIST, the DoE, national labs) focus on more applications-oriented translational research, and companies directly fund applications-oriented R&I. On the other hand, EU projects usually target several levels at the same time, and a single project may include basic research work, applications to realistic use cases, and even commercial deployment of novel technologies. Thus, high-level collaboration mechanisms, such as joint funding programmes or calls, are difficult to set up in a way that takes these differences into account. However, lower-level mechanisms that e.g. focus on the integration of US companies or industry-led associations for specific tasks within an EU project will be easier to accomplish.

Finally, there may be differences in the time spans between the application and the start of funding. EU projects are complex constructs that involve large consortia of partners from both, academia and industry, and it usually takes several months from the submission of an application to the start of funding. On the other hand, companies often have very specific R&I needs that can be achieved with relatively small effort, and they require a short-term return and a quick start of funding (e.g. within a few weeks) after application. However, EU projects are interesting for US companies for longer-term, more visionary R&I despite these timing differences, because these projects often run for several years, which provides planning security.

4.2.2. Administrative Overhead and Legal Barriers

International collaboration efforts always incur an administrative and bureaucratic overhead that can be a major barrier, as determined by the IoT/CPS expert group. There are many different potential mechanisms for EU-US collaboration, several of which have been successfully implemented before. The EU project *TAMS4CPS* has published proposals for such mechanisms (26), which can be separated into three different groups.

High-level, *top-down*, *heavyweight mechanisms* provide comprehensive frameworks for international collaboration. These include e.g. the **high-level multilateral agreements** between different countries (such as the 2016 Implementing Arrangement that was recently signed between the EU and the US[10]), large **thematic, targeted funding programmes** (such as the joint EC-NIH programme that supports EU-US collaboration in the health sector), and **joint calls** for R&I projects that pool funding all involved countries. High-level mechanisms usually require strong political support, and it often takes many years (estimated in interviews until 2020 when starting now) and a very large amount of work of all involved partners to set up such mechanisms.

Lower-level, *bottom-up*, *lightweight* mechanisms focus on specific collaboration aspects with smaller, targeted actions that can be set up relatively easily and quickly, and that occur a much smaller overhead than top-down programmes. These range from the **organization of joint workshops, conferences, and series of seminars** over support for the **mobility of researchers, staff exchange, fellowships to students, and training and education** and the trans-Atlantic provision of **access to research infrastructure, testbeds, and demonstrators** to (at the upper end in terms of complexity) relatively loose connections between calls for R&I projects, such as **coordinated calls** (for which both sides execute calls on a specific thematic topic that are temporally synchronized and that may support the involvement of external partners from both sides of the Atlantic, but where evaluation and funding is organized separately by each side) and **project twinning** (e.g. by implementing lightweight collaboration actions between existing R&I projects and consortia). The EC is currently planning to include coordinated calls and twinning into future work programmes as an instrument of a focused international strategy. It is e.g. planned to launch coordinated calls with Brazil, Japan and South Korea in the future (27).

[10] http://ec.europa.eu/research/iscp/index.cfm?pg=usa

Finally, *collaboration support mechanisms* do not directly implement collaboration actions but provide support that facilitates the set-up of such actions. These include e.g. the facilitation of US participation in mainstream H2020 projects, the enhancement of framework conditions for trans-Atlantic collaboration, and the promotion of the visibility of EU/US programmes, as e.g. done in the *BILAT USA 4.0*, *PICASSO*, and *DISCOVERY* projects.

Our analysis and the interviews have conclusively shown that heavyweight mechanisms do currently not have a good chance of being successfully implemented in the IoT/CPS sector, particularly in the current political climate and if they require pooling of EU and US funding (see also below)[11]. The major reasons are the large overhead in the face of a lack of clearly visible benefits of such programmes and the fast evolution of the ICT field (and in particular of the IoT) that cannot be suitably reflected over the long time frames that are needed to set up high-level programmes.

Legal requirements are seen as major barriers for EU-US collaboration as well. In fact, many companies, for which the availability of external funding is often not an important requirement in joint R&I projects, see legal requirements as the major barrier for international collaboration. Companies are not interested in signing complex, restrictive legal documents, and initiatives that facilitate collaborations involving companies (such as the *Intelligent Manufacturing Systems* (*IMS*) program) restrict the legal requirements for partners by providing lightweight agreements and MoUs (memoranda of understanding).

It was noted by several interview partners that the need for US partners (in particular companies) to sign H2020 grant and consortium agreements has made it virtually impossible to involve commercial partners in H2020 aspects. However, this requirement has recently been removed under a new "Implementing Arrangement"[12] that was signed in October 2016 by the EU and the US. Under this new agreement, US organizations that do not receive any funding under H2020 are allowed to partake in research efforts and other relevant activities in the scope of EU projects without having to sign grant and consortium agreements, thus providing a new basis for EU-US R&I collaboration.

4.2.3. Lack of Clarity of the Benefits of EU-US Collaboration

The IoT/CPS expert group found that a major barrier to international collaboration is a lack of awareness and clarity about the benefits of EU-US collaboration activities for the participants, and a key requirement is the identification of these benefits and their communication to funding agencies, industry, and academia. Obviously, the more administrative and bureaucratic overhead a collaboration measure creates, the larger and more convincing the benefits must be. Questions that must be answered include e.g. "Is there a skill gap which can be complemented by collaboration?", "Is there mutual economic benefit?", "What will be missed if there is no collaboration?", or "What are the common interests?" (see section 2).

Generally, collaborations within the research community are easier to justify than academic-commercial or pure commercial collaboration. The research community is inherently global and universal, and often significant advances in key areas are only possible in international collaboration efforts, e.g. by leveraging what EC academia can contribute, and vice versa. Major success stories of successful international collaboration efforts are e.g. CERN and the nuclear fusion reactor ITER. Another major benefit of EU-US research collaboration is that the expansion of the horizons of scientific human capital (e.g. of students, graduates, post-docs) is a prerequisite for successful scientific research.

The identification of benefits for the inclusion of companies into collaboration efforts is more involved. There must be immediate incentives that justify the effort and the release of internal information and IP. Short-term

[11] Note that bilateral agreements between the US and a single EU member state are easier to implement than multilateral agreements between the US and the EU. Successful programs have e.g. been implemented between the US and Germany, the US and the UK, and the US and Ireland.

[12] http://ec.europa.eu/research/iscp/index.cfm?pg=usa

benefits must be identified for concrete commercial and application scenarios within a restricted thematic area (such as additive manufacturing or specific scenarios involving the industrial IoT). Some general benefits for the involvement of companies in EU-US collaboration efforts are that in the globalized age, the merging of technologies from different parts of the world is an important competitive advantage that can lead to economic growth, that collaborations increase global visibility of a company, that different regions possess different strengths that can complement each other, and that collaboration may mitigate risks. For example, the US is strong in software and computing while the EU has unique strengths in smart production and cyber-physical systems development and deployment. In such a case, complementarity can create more than the sum of the parts when bringing different sectors together (provided the collaboration is not too close to commercial interests of the participants).

The advancement of international standardization and the sharing of infrastructure, testbeds, and demonstrators are other key benefits of EU-US collaboration (where again CERN and ITER are good examples of successful shared infrastructure). Infrastructure and testbeds are expensive to build, thus sharing will benefit both sides, and EU-US collaborations on standardization will set the standard for the rest of the world, in particular for the IoT sector in which all player are aware that trying to build a region-specific IoT is doomed to fail. Global efforts are seen as the only way forward.

In its recent survey (23), the *DISCOVERY* project asked respondents to identify the benefits that are most important for EU-US ICT collaboration. Gaining competitive advantages by an extended view of future challenges was identified as the most important benefit, followed by creating overseas relationships, sharing and gaining insights into research activities, and gaining international visibility.

4.2.4. Restrictions due to Intellectual Property Protection

Collaboration may be difficult on topics of **high near-term commercial importance**, i.e. innovation efforts that focus on products and services that may lead to large profitable businesses in the near term. Different regions are in competition, and industrial policy focuses on measures that reinforce own industry. This barrier is seen as important in all analyzed application sectors, and this is also a conclusion by the *BILAT USA 4.0* project that has found a lack of bilateral funding agreements between the EU and the US in areas with immediate economic outcomes. They state that “one reason for the lack of joint funding agreements may be that there are immediate economic outcomes where the US has a competitive advantage compared to the EU in the areas of technology levels, entrepreneurship, supporting start-ups, and venture capital.” (28).

It is thus arguably easier to collaborate on basic research than on applied research. An example is the *FET* (*Future and Emerging Technologies*) EC programme that focuses on basic research. Here, it is much easier to involve US partners (even including trans-Atlantic funding) than in other, more applications-oriented programmes, such as the ECSEL Joint Undertaking. One exception is the joint work on international standards and interoperability. While this is of commercial importance, it usually does not require companies to disclose information and technology that affects stand-out features of their products.

The Big Data expert group found that industrial competition between US and EU has a long tradition: It is widely accepted that EU and US are two competing regions, especially on technologically driven industries. Especially in the area of Big Data, Europe has been slow to adopt compared to the United States. More than half of worldwide revenue from big data is expected to come from the USA, and only one in twenty top big data companies is European (29). Thus, it can be very challenging for funding agencies and organisations from these regions, to collaboratively tackle research of high TRL (Technology Readiness Level) or applied research topics. However, tackling basic research subjects and topics can be an alternative.

The 5G expert group has identified this barrier as important for research topics that are already considered as study or work items in global standardization bodies, like 3GPP and IEEE. Hence, it will be easier to collaborate on fundamental research than on applied research.

4.2.5. Lack of Joint EU-US Funding Mechanisms and Policies

Generally, most of the EU funding will be used to fund EU companies and research institutes, and US funding will focus on the support of US organizations and companies. Thus, EU-US collaboration will always be a complement, or even an exception, to local and regional funding. This is not expected to change in the near future and is one of the reasons why high-level mechanisms such as joint calls or thematic, targeted funding programmes are difficult to implement (see above).

The Big Data expert group has also found that joint funding is a challenging task: As already known, US structures (both private and public) who are based in the US, have limited access to EU funding. US structures are eligible for participation in EU projects, but financial support is only available for calls where this is specified, e.g. International Cooperation calls targeting collaboration with the USA or the "Health" programme in general. Potential US participants are therefore encouraged to contact research and innovation funding organisations in the US to seek support for their participation in Horizon 2020. No jointly agreed mechanism is currently in place for co-funding Horizon 2020 research and innovation projects. On the other side, EU organisations willing to participate in US research programmes, face similar challenges, as it is almost impossible to receive funding from US agencies. Results from the newly signed EU-US agreement (signed in October 2016), which offers new opportunities for research cooperation, remain to be seen.

4.2.6. Export Control and Privacy Restrictions

Topics touching **export control issues**, **sensitive or classified data / information**, or **privacy issues** should be avoided. The EU and US national priorities, rules, and regulations are very different and will be difficult to harmonize, and generally legal and policy differences will be difficult to overcome in these areas. In particular export control issues have been identified in interviews as major blocking factors of international collaborations. Such issues must be dealt with appropriately before starting any collaboration actions.

The Big Data expert group found that data privacy is a complicated issue: The collection and manipulation of Big Data, as its proponents have been saying for several years now, can result in real-world benefits. However, it can also lead to big privacy problems (30). Both the EU and the US, have established a number of laws, policies and directives dictating the use of personal data by organisations and institutions willing to benefit from them. There are many differences between the laws regarding data privacy in the European Union and the United States, with the E.U. generally allowing more rights to the individual. With no single law providing comprehensive treatment to the issue, America takes a more ad-hoc approach to data protection, often relying on a combination of public regulation, private self-regulation, and legislation (31). Even after the US and the EU signed the EU-US Privacy Shield Framework (32), open issues remain, making it very challenging and complicated for organisations coming from these different regions to collaborate on research topics related to personal data. Moreover, the situation in EU is no homogenous across member states; e.g., Directive on Protection of Personal Data needs to be ratified and implemented by the member states, which may lead to inconsistencies.

4.2.7. Lack of Awareness and Knowledge

A lack of awareness and knowledge of EU and US actors of the other side is detrimental to collaboration. E.g., *BILAT USA 4.0* found that interested US actors may be unaware of how EU funding schemes operate (including misconceptions on how US partners can participate in H2020), and are not aware of the R&I priorities of the other side. In addition, it is often straightforward to connect to other initiatives within the US, but the EC landscape is fragmented, and the responsibilities may not be clear to US agencies.

This barrier is confirmed by an investigation of the *DISCOVERY* project (23) that identified as main barriers the lack of information on funding opportunities and programmes, the lack of knowledge about specific research areas and topics that are open to international cooperation, difficulties to understand the rules of participation in other countries, and a lack of partner search tools and methods.

Currently, several EC projects are working on solutions for these issues, including *PICASSO*, *TAMS4CPS*, *DISCOVERY*, and *BILAT USA 4.0.*

4.2.8. Lack of Interoperability and Standards

A **lack of interoperability and (device) standards** can be a barrier to collaboration. This is true for several of the application sectors and, in more detail, in (1). In addition, IoT/CPS systems were noted by our interview contacts as sometimes being highly regulated, which can stifle innovation.

4.3. Collaboration Opportunities

This section provides an overview of potential opportunities for collaboration between EU and US in the area of Big Data (note that technological opportunities are given in chapter 3). Chapter 3 together with the sections below provide a holistic picture for collaboration in Research & Innovation topics, Education and Additional Collaboration Opportunities between the two regions, in order to jointly tackle challenges that have been identified by the policy makers. The current section has been updated with findings and proceedings from the **Transatlantic Workshop on Public Private Partnerships for Big Data Research and Innovation and Workforce Development**[13], which was held in Versailles, France, on November 20, 2017 as a partnership between the **US National Science Foundation (NSF) Big Data Regional Innovation Hubs, the EU Big Data Value Association (BDVA), the PICASSO project, and INRIA.** The workshop was **conceived as a kick-off for an ongoing collaboration** aimed at helping both the **EU and the US cultivate partnerships and develop a workforce** poised to advance and apply data science now and in the future. The workshop had three stated goals:

- To **determine best practices for supporting collaborative research programs** in data science, with a special focus on public-private partnerships and innovation in the areas of smart cities, transportation, health, and the nexus of environment, food, energy, and water;
- To **identify promising areas for bilateral EU-US research and data sharing**, especially among the US National Science Foundation's Big Data Regional Innovation Hubs and Spokes program and the EU Big Data Value Public-Private Partnership program; and
- To **examine ways to develop the burgeoning discipline of data science** in order to train a skilled workforce capable of keeping up with the rapid growth in opportunities to collect, analyze, and apply data.

4.3.1. Big Data Ecosystem Opportunities

Moreover, from our analysis, it is obvious that establishing and supporting a **Big Data Ecosystem** for creating value and getting the most out of Big Data, is the highest priority for both regions. Both the US and the EU have indicated the importance of such an ecosystem (in numerus policy briefs) and are implementing specific activities for supporting such an initiative.

More specific, the **US Big Data SSG** (Big Data Senior Steering Group) has identified the need for **"development testbeds" or "sandboxes" to** enable conversion of agency-funded R&D results into innovative production capabilities, as well as for engaging in proofs of concept with both open source and proprietary commercial off-the-shelf solutions. On the other hand, the **EU commission** has launched a number of **"Lighthouse Projects"** in

[13] http://www.picasso-project.eu/2017/11/27/trans-atlantic-workshop-on-public-private-partnerships-for-big-data/

order to take existing technologies and apply the, to innovative use cases (with possibly slight adaptation and enhancement of the technologies).

The implementation of a joint programme or set of projects for establishing international partnerships to jointly tackle specific challenges will give a huge boost to Big Data industry. Such a Joint program/project would enable the sharing of experiences, results, and capabilities among agencies and organisation, shorten the development phase of a project, and allow agencies and organisation to assimilate and integrate new results and solutions quickly. Industry engagement in the program would demonstrate broader utility, foster better interoperability, and potentially provide long-term sustainability of solutions. Pilots and testbed infrastructure could be shared among agencies and organisation, thereby helping to maximize investments and share the benefits of projects and technologies that would otherwise remain isolated.

4.3.2. Standardisation & Regulation

Adding to the technological priorities, additional collaboration opportunities exist in non-technological areas such as standardisation and regulation.

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. However, a joint EU-US standardisation board working on this subject, could fill in an existing large gap and bring both regions to the technological forefront.

Moreover, regulation is also a key enabler in the field for global adoption of services and this is already well recognised with activities such as Safe Harbour and Privacy Shield. However, a Joint initiative between the two regions will create a fertile and fruitful environment in which the industry could operate without having to individually overcome the burden of different policies and regulations for each region.

4.3.3. Opportunities in Education & Workforce

What can also be extracted from this exercise is that **there is a great potential for EU and US universities to collaborate in order to fulfil the huge demand of Big Data graduates**, and cooperate in order to learn from each other, mutually sharing experiences. Moreover, the cooperation of educational institutions and businesses, coming from both regions will benefit both sides in order to better understand the needs and, possibly, define new ways and curriculums for tackling them. Adding to the above, at a skills level it was noted that it is difficult to recruit for smart jobs. There are also issues of transferring engineers between the EU and US. If an EU engineer wishes to work in the US there is a need to learn and get US qualifications even though they may have very good European qualifications. This makes the transfer of people and skills difficult. There is also a need to retrain on US standards if engineers are engaged in sectors where different standards apply. Here, the role of education is critical, as harmonization of skills, standards and the process of accreditation would all be beneficial.

4.3.4. Big Data for Smart Cities[14]

Transatlantic collaborations will be essential to identifying smart city initiatives and examining their successes and failures. Improved data sharing, standardization, and interoperability, especially for publicly-generated or crowdsourced data, could be undertaken by specialized working groups jointly funded by the EU and US. These

[14] Update from the Transatlantic Workshop on Public Private Partnerships for Big Data Research and Innovation and Workforce Development

collaborations would also uncover and potentially improve synergies for areas where public perceptions and priorities differ between the US and the EU, especially in regard to data collection and use. Different attitudes toward crime, for example, would impact what data is collected, how it is shared, and where it is applied.

4.3.5. Big Data and the Environment-Food-Energy-Water Nexus[15]

EU Lighthouse projects partners and US BD Hubs and Spokes-affiliated researchers should examine their projects for matchmaking possibilities that can lead to improved data sharing, interoperability, and international standards. For example, JTC-1 is an ISO in use across several projects and could serve as a model for others. Transatlantic webinars or conferences should be convened to answer questions about sensors, including the strengths and weaknesses of various products, how they can be used, and how their data can be interpreted and applied. Studying US and EU citizen science projects could also yield valuable lessons. Additional data processing and improved metadata would also lead to a better understanding of how data is collected, accessed, used, and shared, and inspire improvements.

4.3.6. Big Data for Better Health[16]

Big Data are already considered as critical for the Health sector. With a growing need for efficient and accessible healthcare, companies and healthcare organizations are starting to invest in applications and analytical tools that help healthcare stakeholders identify value and opportunities, in fields such as[17]:

- **Build sustainable healthcare systems:** The healthcare industry is constantly faced with competitive and legislative pressure and must determine ways to reduce the cost of care, while efficiently managing resources. Healthcare organizations should focus on understanding the patient and improving patient care by promoting effective resource utilization.
- **Collaborate to improve care and outcomes:** Healthcare organizations should improve patient engagement and personalize healthcare initiatives that improve the quality and efficiency of care. Understanding a patient individually is important when designing tailored yet effective healthcare programs.
- **Increase access to healthcare:** A major issue with healthcare is access. In order for the population to thrive, healthcare must be available and accessible. Educating consumers on preventive care can improve health and reduce the demand and waste of healthcare resources.

EU and US should synchronise their efforts mainly on reinventing electronic health records by incorporating machine learning advancements to automate clinical documentation and on performing meta-analyses on EHR data quality, which would inform data model standardization. These activities could achieve much-needed interoperability and standardization, thus enabling more sharing of data sets with strong privacy protections. As Health data are directly related to nearly all aspects of one's life, including the physical environment, living conditions, education, lifestyle, economic stability, and social support systems, by integrating these fields with traditional medical data into mHealth initiatives could generate more advanced models and insights.

---

[15] Update from the Transatlantic Workshop on Public Private Partnerships for Big Data Research and Innovation and Workforce Development

[16] Update from the Transatlantic Workshop on Public Private Partnerships for Big Data Research and Innovation and Workforce Development

[17] https://www.business2community.com/big-data/why-is-big-data-important-in-healthcare-0576823

Moreover, adoption of the Blue Button Standard[18] which allows patients to view and download their personal health records, is also a critical joint theme both for EU and for the US.

4.3.7. Potential Collaboration Mechanisms

The Big Data expert group has found that the following mechanisms, which are derived from discussions with EG members, interview results, outcomes from the the Transatlantic Workshop on Public Private Partnerships for Big Data Research and Innovation and Workforce Development and propositions made by TAMS4CPS EU Project (26), are suitable to promote collaboration, targeting both the participation of US organisations to EU initiatives and vice versa.

1) **Provide matching funds to EU or US organisations** for participating to international programmes
2) **Enhance the visibility of existing research tools**, such as Marie Skłodowska-Curie actions, ERC, etc.
3) **Provide funding to supportive activities**, such as joint workshops, seminars of conferences.
4) **Provide funding for US organisations in H2020** and vice versa.
5) **Highlight and Upgrade the role of existing structures**, such as the **TABC – TransAtlantic Business Council**
6) **Establish Bilateral Thematic Structures**, for example a joint structure with BDVA and US Big Data US Hubs (Big Data EU-US Task Force for Enhancing Collaboration)
7) **Establish joint calls, twinning of research projects, co-fund schemes**
8) **Active support of the mobility of researchers, staff exchange, fellowships to students, trans-Atlantic training and education**

[18] https://en.wikipedia.org/wiki/Blue_Button

5. Conclusions and Outlook

This report outlines new **technology themes and priorities** and **collaboration and cooperation opportunities and mechanisms** that have been identified as being promising for EU-US collaboration in Big Data. The themes and opportunities were synthesized based on comprehensive analyses of the EU and US research and innovation priorities in the technology sectors and related application domains, the current EU and US policy environment and priorities, the EU-US funding and collaboration landscape, and technological and policy barriers for EU-US collaboration. The contents of this report have been validated and refined extensively, e.g. based on in-depth discussions and online distribution and feedback actions with a large network of international experts, analytical research by the Expert Groups, preliminary PICASSO results, and other feedback collection mechanisms such as a public consultation on the PICASSO website.

6. References

1. **Thompson, H. and Ramos-Hernandez, D.** *Analysis of Industrial Drivers and Societal Needs.* 2016. PICASSO report D1.5. http://www.picasso-project.eu/wp-content/uploads/2016/11/PICASSO-Analysis-of-Industrial-Drivers-and-Societal-Needs-Public-Version.pdf.

2. —. *Panorama of ICT landscape in EU and US: ICT, Policies, Regulations, Programmes, and Networks in the EU and US.* 2016. PICASSO Report D1.3.

3. **Bezzi, M. und Calderaro, M.** *Overview on ICT-related Access Opportunities in EU and US.* 2016. PICASSO Report D1.2.

4. **European Commission.** Digital Single Market: Big Data. *European Commission.* [Online] 09 14, 2016. [Cited: 11 28, 2016.] https://ec.europa.eu/digital-single-market/en/big-data.

5. **Bean, Randy.** Variety, Not Volume, Is Driving Big Data Initiatives. *MIT Sloan.* [Online] 3 28, 2016. [Cited: 12 12, 2016.] http://sloanreview.mit.edu/article/variety-not-volume-is-driving-big-data-initiatives/?utm_source=twitter&utm_medium=social&utm_campaign=sm-direct.

6. **Ernst & Young.** *Big data, Changing the way businesses, compete and operate, Insights on governance, risk and compliance.* 2014.

7. **Freeland, Chrystia.** In Big Data, Potential for Big Division. *New York Times.* [Online] 01 12, 2012. [Cited: 01 03, 2017.] http://www.nytimes.com/2012/01/13/us/13iht-letter13.html.

8. **DELL EMC.** DELL EMC. *DELL.* [Online] [Cited: 01 03, 2017.] https://www.emc.com/leadership/digital-universe/index.htm.

9. **Mitchell, Robert L.** 8 big trends in big data analytics. *ComputerWorld.* [Online] 10 23, 2014. [Cited: 01 03, 2017.] http://www.computerworld.com/article/2690856/big-data/8-big-trends-in-big-data-analytics.html.

10. **European Commision.** Towards a thriving data-driven economy. *EC Official Website.* [Online] 02 24, 2016. [Cited: 12 12, 2016.] https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy.

11. **Lupiáñez-Villanueva, Francisco.** *Big data in Europe: new environment, new opportunities.*

12. **Big Data Value Association.** *Strategic Research & Innovation Agenda Version 2.0.* 2016.

13. **Office of Science and Technology Policy.** White House. *OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE.* [Online] 03 29, 2012. [Cited: 10 10, 2016.] https://www.google.gr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjZ8YnR46XRAhXJN1AKHQf1AUQQFggYMAA&url=https%3A%2F%2Fwww.whitehouse.gov%2Fsites%2Fdefault%2Ffiles%2Fmicrosites%2Fostp%2Fbig_data_press_release_final_2.pdf&usg=AFQjCNGDfomg7zyDTyUq77n.

14. **NITRD.** *The Federal Big Data Research and Development Strategic Plan.* 2016.

15. **Wactlar, Howard.** CISE Directorate, National Science Foundation, NIST Big Data Meeting, June, 2012:. *NIST.* [Online] 06 2012. [Cited: 12 15, 2016.] https://www.nist.gov/document-6748.

16. **PCAST.** *Big Data & Privacy: A Technological Perspective.* s.l. : Executive Office of the President, 2014.

17. **National Science Foundation.** Big Data Regional Innovation Hubs (BD Hubs). *NSF.* [Online] 2015. [Cited: 12 12, 2016.]

18. **Noyes, Katherine.** Educating the 'big data' generation. *Fortune.* [Online] 06 02, 2014. [Cited: 12 19, 2016.] http://fortune.com/2014/05/27/educating-the-big-data-generation/.

19. **Masters Portal.** Masters in Data Science & Big Data. *Masters Portal.* [Online] 12 15, 2016. [Cited: 12 15, 2016.] http://www.mastersportal.eu/disciplines/282/data-science-big-data.html.

20. **European Commission.** Societal Challenges. [Online] [Cited: 12 01, 2016.] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges.

21. **H2020, ICT Work Programme 2015-2016.** ICT-15-2016-2017. [Online] [Cited: 12 15, 2016.] http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/ict-15-2016-2017.html.

22. **NIST Big Data Public Working Group.** NIST Big Data Interoperability Framework V1.0. *NIST.* [Online] [Cited: 11 15, 2016.] https://www.nist.gov/el/cyber-physical-systems/big-data-pwg.

23. **Ursa, Y., Zeng, X. und Cabrera, M.F.** *Report on ICT Research and Innovation Priorities for EU-North America Cooperation.* s.l. : DISCOVERY project, 2016. Deliverable D2.1.

24. **Cyber Physical Systems Public Working Group.** *Framework for Cyber-Physical Systems.* s.l. : NIST, US, 2016. https://s3.amazonaws.com/nist-sgcps/cpspwg/files/pwgglobal/CPS_PWG_Framework_for_Cyber_Physical_Systems_Release_1_0Final.pdf.

25. **Department of Commerce (DoC), NTIA.** *Fostering the Advancement of the Internet of Things.* 2017. Green Paper. Available at: https://www.ntia.doc.gov/other-publication/2017/green-paper-fostering-advancement-internet-things.

26. **TAMS4CPS EU Project.** *Strategic Research Agenda for Collaboration - Modelling and Simulation for Cyber-Physical-Systems.* Nov. 30, 2016. Draft document.

27. **European Commission.** *Scoping Papers for the Horizon 2020 Work Programme 2018-2020.* 2016. Available at: ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/17_thematic_papers.zip.

28. **Consortium, BILAT USA 4.0 Project.** *Analyzing Report on Consultation Process with Funders and Policy Makers.* 2016.

29. **Davies, Ron.** *Briefing for the European Parliament.* s.l. : EPRS European Parliamentary Research Service, September 2016. PE 589.801.

30. **Armerding, Taylor.** The 5 worst Big Data privacy risks (and how to guard against them). *CSO.* [Online] 12 08, 2014. [Cited: 01 01, 2016.] http://www.csoonline.com/article/2855641/big-data-security/the-5-worst-big-data-privacy-risks-and-how-to-guard-against-them.html.

31. **Politics & Policy.** The European Union and Internet Data Privacy. *Politcs & Policy.* [Online] [Cited: 01 05, 2017.] http://politicsandpolicy.org/article/european-union-and-internet-data-privacy.

32. **DG Justice.** The EU-U.S. Privacy Shield. *DG Justice.* [Online] 11 24, 2016. [Cited: 01 05, 2017.] http://ec.europa.eu/justice/data-protection/international-transfers/eu-us-privacy-shield/index_en.htm.