



COLUMBIA UNIVERSITY
Data Science Institute

Data Literacy: Workforce Essential, Workforce Right

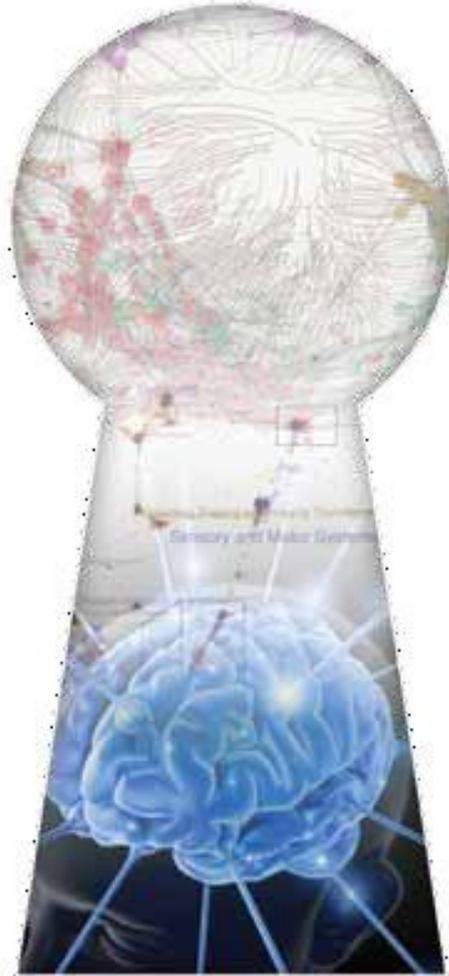
2nd Annual Transatlantic Symposium on ICT and Policy

Wilson Center

June 18, 2018

Catherine Cramer, Columbia University Data Science Institute

Data Driven Science



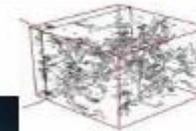
- The data revolution
- Now we are drowning in data
- eScience: “The 4th Paradigm”

Emergence of a Fourth Research Paradigm

1. Thousand years ago – **Experimental Science**
 - Description of natural phenomena
 2. Last few hundred years – **Theoretical Science**
 - Newton’s Laws, Maxwell’s Equations...
 3. Last few decades – **Computational Science**
 - Simulation of complex phenomena
 4. Today – **Data-Intensive Science**
 - Scientists overwhelmed with data sets from many different sources
 - Data captured by instruments
 - Data generated by simulations
 - Data generated by sensor networks
- **eScience is the set of tools and technologies to support data federation and collaboration**
- For analysis and data mining
 - For data visualization and exploration
 - For scholarly communication and dissemination



$$\left(\frac{a}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



(With thanks to Jim Gray)

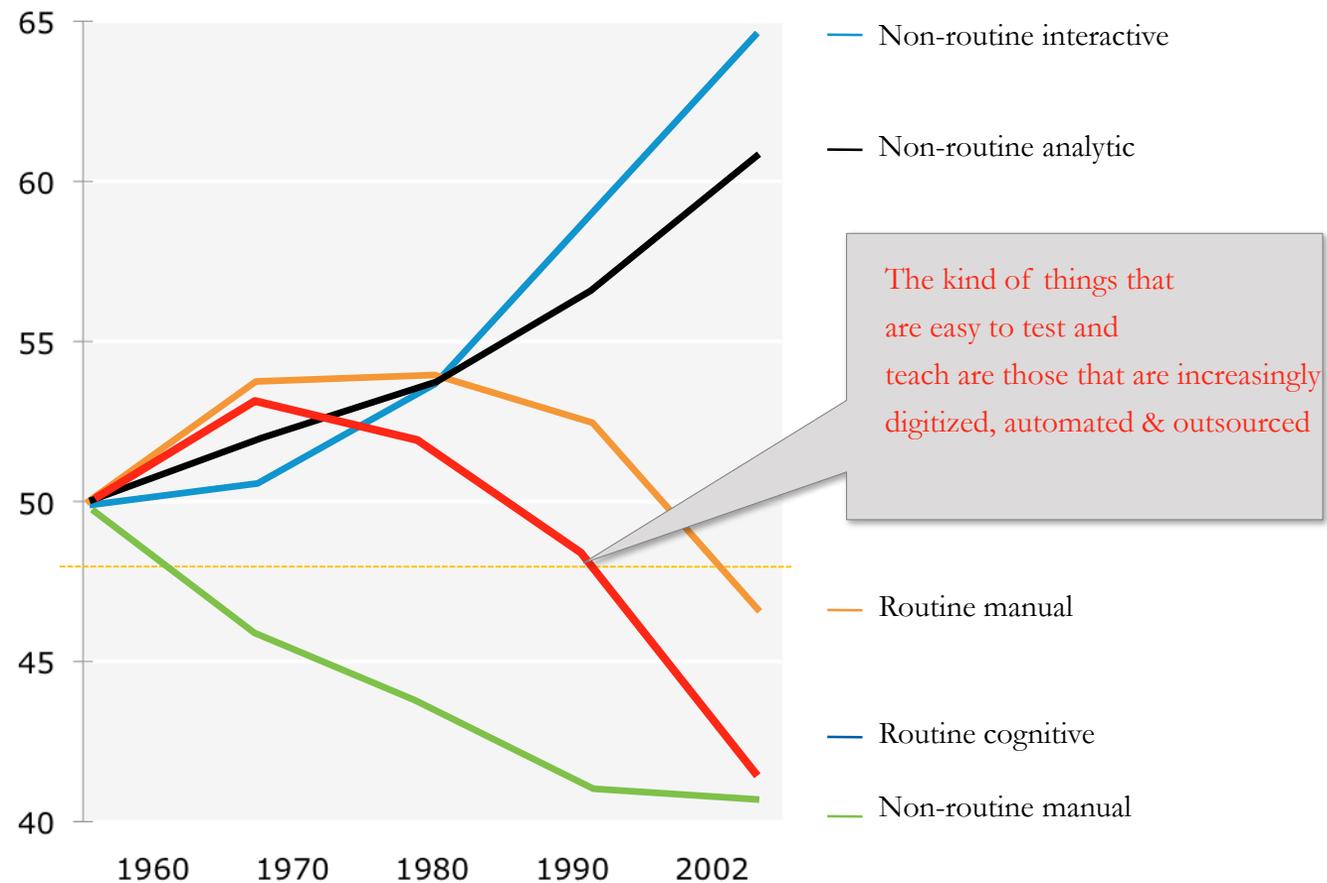
What's The Problem?

How the demand for skills has changed

Economy-wide measures of routine and non-routine task input (US)

Mean task input as percentiles of the 1960 task distribution

Data science has transformed many sectors of society



What are the Needed Skills?



- Emphasis on **design, construction, visualization** of large-scale data;
- Federation of **multivariate data streams**;
- **Exploratory, stochastic, pattern-seeking and interdisciplinary** approaches;
- **Machine learning; semantic databases and ontologies**;
- More available and **interoperable large data structures**.

What's The Problem?



No equivalent transformation of science teaching.





All citizens in the 21st century should
be data literate by the time they
graduate from high school.
(future workforce)





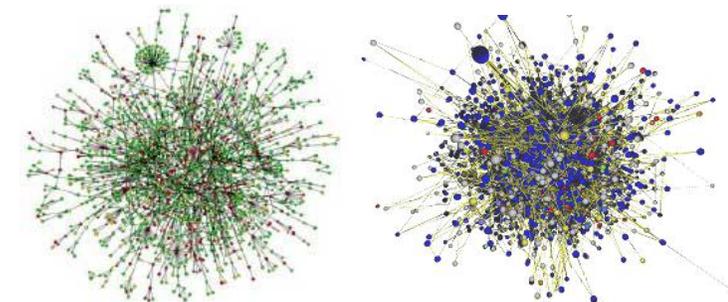
All citizens in the 21st century should be data literate by the time they graduate from high school.

(future workforce)



....as should those in the workforce

NOW *(current workforce)*



Planning Grant



Big Data Literacy: Building Capacity for Regional Collaboration in Closing the Big Data Divide

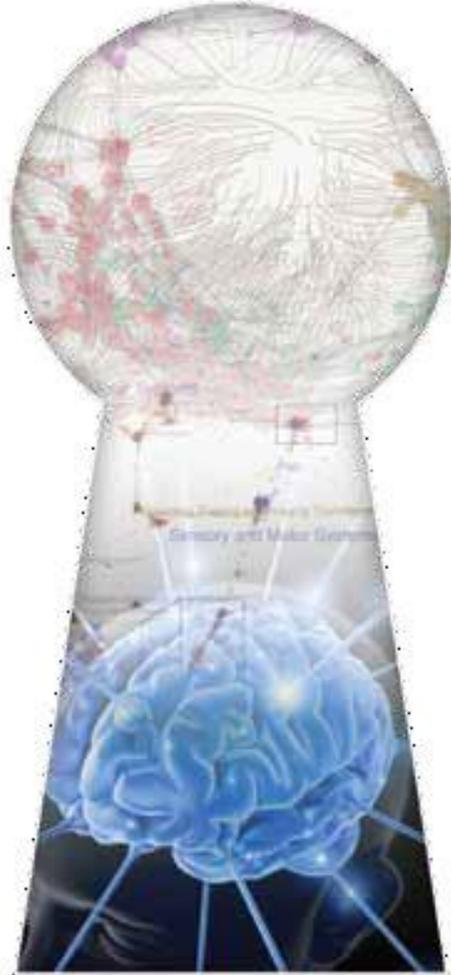
- galvanize **big data communities of practice** around learning and education
- identify and articulate the **nature and quality of big data education resources**
- draft a **set of big data literacy essential principles** and a planning document featuring strategies to advance lifelong and life-wide big data literacy for grades P-20 across learning settings.

Planning Grant



- make **recommendations for its trajectory** over the next five years
- propose a set of **priority education strategies** for subsequent implementation
- subsequent initiative, to **devise, implement, and evaluate programs, curricula and educational and career development interventions** which will address the evolving needs of big data education in **all sectors of society**.

Collaborative Inquiry

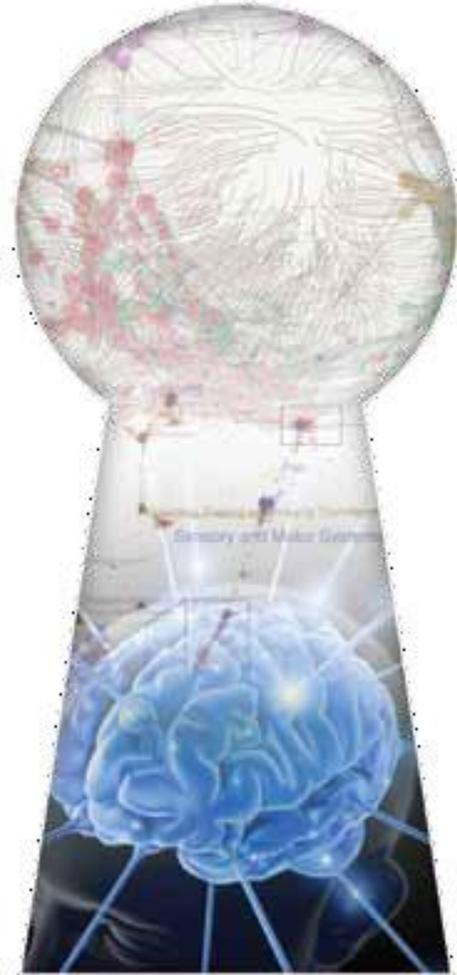


From the Keeping Data Science Broad Report: Data Literacy Goals



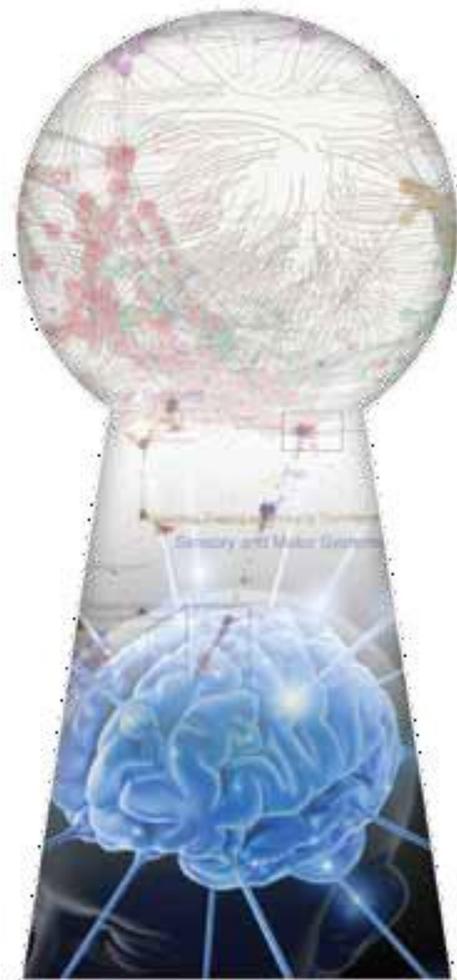
- A **commonly understood definition of DL** and how it differs from data science
- A **published handbook** including examples from journalism and academic papers
- **Inclusion** of a DL component in **every introductory course**
- Ensuring that DL is **pervasive throughout curriculum** (similar to writing across the curriculum)
- **Teacher training** to use DL approaches
- **All students, starting in pre-K**, will receive DL through curriculum. This approach will continue through middle and high school so that all high school graduates are data literate

From the Report: Data Literacy Goals



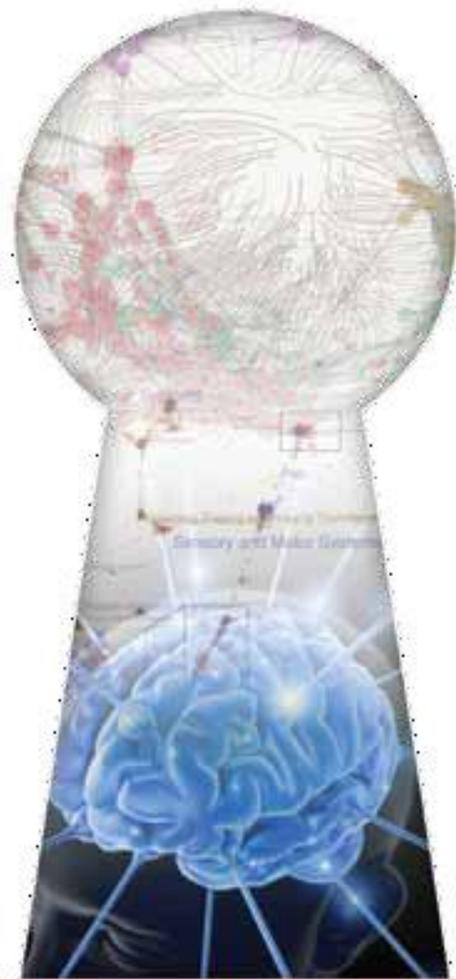
- Using backward design from **career pathways**, e.g. data businessperson, data engineer, data creative and data researcher, to arrive at **skills needed**
- Define the **relationship DL has to domain science**
- Define other terms used to identify **related sets of skills and educational experiences**, e.g. Data Analytics, Business Intelligence, Big Data, Machine Learning, Deep Learning
- Data science will be seen as an **engaging, creative and imperative** body of knowledge
- DL will be the **foundation of fact-based critical thinking**.

From the Report: Next Steps



- Develop a **globally accepted definition of Data Literacy**. This can be achieved through a distributed and collective effort, engaging a wide range of data scientists and data science educators, policymakers, community activists, and learning researchers. This process will require several iterations. The resulting commonly held definition will not belong to any one institution.
- Data literacy may be combined with a course in propositional logic to **create a critical thinking course**.
- **Couple data literacy with a discussion of ethics** using case studies.

From the Report: Next Steps

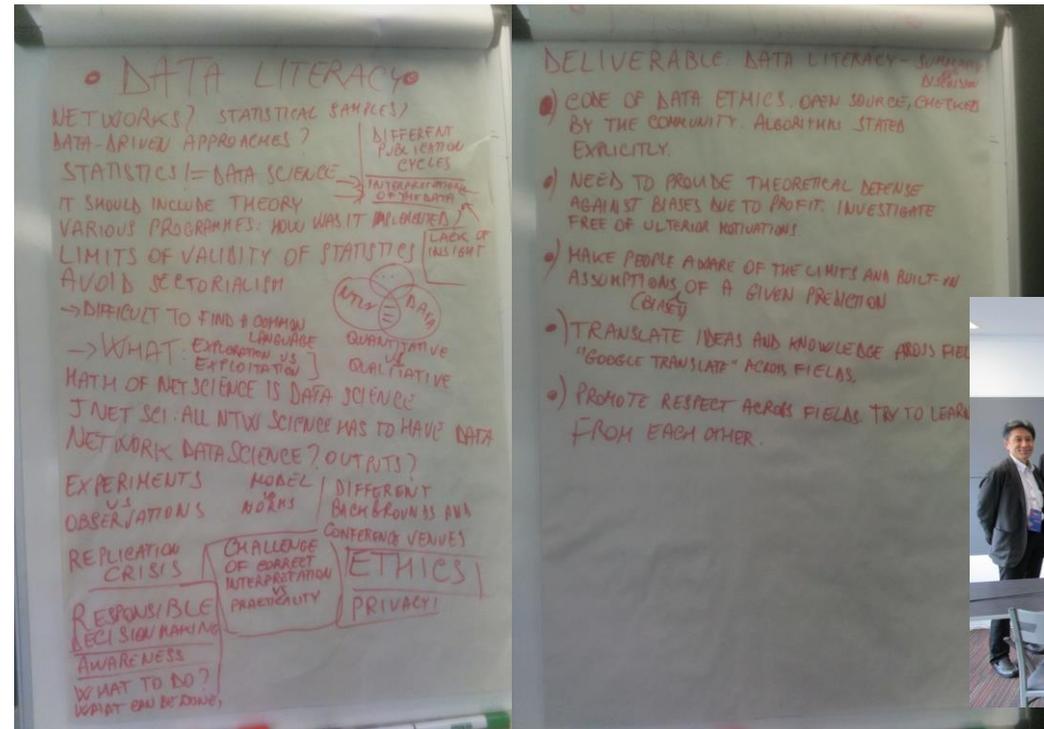


- Demonstrate a **portfolio of work** analyzing and synthesizing articles, blog posts, videos, etc. in a particular field of interest. Student lightning talks may be used to facilitate this.
- Offer middle and high school **teacher workshops** building on elementary school work
- Develop **critical thinking and computational thinking curricula**
- **Strengthen foundational math skills**
- **Strengthen application-based problem-solving skills**

Recent Developments



- Northeast Big Data Innovation Hub Summit
March 2018 - Data Literacy Breakout
- International Network Science Conference Satellite
June 12, Paris, France – Data Literacy Brainstorm



Recent Developments



Draft Concepts or Buckets:

- What are data?
- Data interpretation requires critical thinking
- Data can be useful
- Interpreting visualizations
- Data are neutral - but algorithms are not
- To be data literate means having specific skills
- Data can be preserved, stored and retrieved
- Data quality
- Data privacy and protection
- Data ethics

Next Steps



- Support for looking at learning that is currently happening
- Engaging libraries
- Growing capacity and community
- Cross-Hub collaborations
- EU and UK collaborations

Join us!

Thank you!

Catherine Cramer (cc4287@columbia.edu)



COLUMBIA UNIVERSITY
Data Science Institute