



Building New Partnerships

Breakout Reports

November 12, 2013

Education and Workforce Development



Moderator:

Michael Rappa

**Director, Institute for Advanced Analytics
North Carolina State University**

Session Presenters:

Deborah Nolan

**Professor, Department of Statistics,
University of California, Berkeley**

R.H. Campbell

**Director, Institute for Advanced Analytics,
North Carolina State University**



Education and Workforce Development

Objective:

To discuss collaborative projects that focus on either improving the Big Data workforce or educating the general public about Big Data. These could include: undergraduate and graduate curriculum and practicum; training and certification programs; K-12 education; informal Big Data and data analytics education; and improving access to the Big Data talent pool for organizations needing data scientists.

Education and Workforce Development



Guiding Questions:

- 1) What models for expanding the Big Data workforce should be scaled or replicated? Why?
- 2) Are there challenges associated with the Big Data workforce where no "promising practices" exist?

Education and Workforce Development



- Prepare students at all levels (K-12, undergrad, grad, professionals)
- Need include many fields (social sciences, computer science, statistics, etc.)
- UCSD - Boot camps for professionals
- MOOCs
- Berkeley revamping – prepare MA statistics students for the workforce. Less focus on producing PhD's
- North Carolina State University – multidisciplinary data science education
- University of Illinois – Experimenting with Big Data requirement for engineers

Research and Development



**Moderator:
Stanley Ahalt**

**Session Presenters:
Michael Franklin and Robert Ross**

Research and Development



Objective:

To discuss new research partnerships across federal, academic, private, and non-profit sectors. Topics could include: core techniques and technologies; scientific and technical computing; cross-sector research centers; and cyberinfrastructure to support Big Data.

Research and Development

Guiding Questions:

1) What are the technical challenges associated with Big Data that should be high priorities for federal science agencies? For a given challenge, does the research community have any ideas for making progress?

2) What are new modes for various organization conducting R&D to collaborate or coordinate? (i.e. academic, federal, and industry researchers) How can we foster those collaborations to better bring R&D findings and discoveries to use every day?

Research and Development



What are the technical challenges associated with Big Data that should be high priorities for federal science agencies? For a given challenge, does the research community have any ideas for making progress?

Research and Development

Opportunities, Challenges, or Barriers with Big Data

- The key is asking what technical challenges , those that bridge data science and domain science
 - Rob → 3 categories: analysis/algorithms, user environment, and hardware/software
 - Michael: many of the data science problems are problems where you need to optimize across time , money and quality. And need to that automatically
 - We aren't creating tools the fundamental challenge of collecting and cleaning data
- Discussion of what is data science, what is the new science in data science
 - Inherent scalable
 - Data precedes model
 - Valid extractions

Research and Development



What are new modes for various organization conducting R&D to collaborate or coordinate? (i.e. academic, federal, and industry researchers) How can we foster those collaborations to better bring R&D findings and discoveries to use every day?

Research and Development



Actions or Ideas for collaborations

- Michael : bring industry together and then bring federal sponsors. Because this established the IP framework. Allowed NSF to see value in project, virtuous cycle
- Discussion around utility of open source
- Discussion on the diminished level of industrial research, which has given rise to collaborations between industry and academia
- Federal agencies could provide seed funding, but must insist on value creation

Innovation



**Moderator:
Christy Wilson**

**Session Presenters:
Eng Lim Goh and Joris VanDam**

Innovation



Objective:

To discuss how the data community can come together to foster the innovation ecosystem and promote the use of data science and engineering to address the challenges our nation faces. These challenges could include: national and regional economic development, education, healthcare, energy, climate change, transportation, and others.

Innovation



Guiding Thought:

What is Innovation?

Where should innovation go?

Innovation



Definition of Innovation

- A challenge that cannot be solved by one party on their own
 - Requires partnerships to solve
 - New models/tools/thoughts to better production
- A circular, wandering thought process rather than linear path to a solution
- Creativity x Adoption

Innovation



Where should innovation go?

- Sharing of work flows, dynamic software, models, not just information and expertise
- Remove barriers to combining practice and research
- Real time access to data and innovative approaches to solve problems
- Provide data and questions that are consumable to others
- Identify the important data
- Thinking from serendipitous level
 - Blue skies research
 - Focus on unknown unknowns