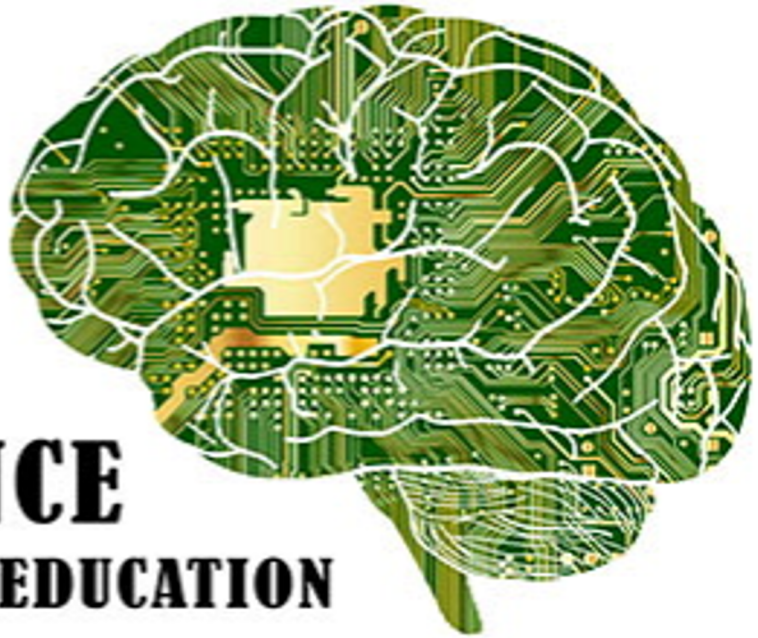
The National Academies of
SCIENCES • ENGINEERING • MEDICINE

JSM 2020 Late Breaking Session: Highlights from the National Academies of Sciences, Engineering, and Medicine's Roundtable on Data Science Postsecondary Education

August 4, 2020

https://nas.edu/dsert



ROUNDTABLE ON DATA SCIENCE POST-SECONDARY EDUCATION

Facilitated by:

Board on Mathematical Sciences and Analytics (BMSA)
And
Committee on Applied and Theoretical Statistics

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

Roundtable on Data Science Post-secondary Education

Sponsors

GORDON AND BETTY MOORE FOUNDATION

ASA
AMERICAN STATISTICAL ASSOCIATION

NIH
National Institutes of Health

MAA
MATHEMATICAL ASSOCIATION OF AMERICA

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

Today's Panelists

- Kathy McKeown - Columbia University (DSERT co-chair)
- Nicholas Horton - Mathematics & Statistics, Amherst College
- Mark Krzysko - Acquisition Data and Analytics, OUSD(A&S) Acquisition Enablers
- Rachel Levy - Mathematical Association of America
- Nina Mishra – Amazon

- Moderator: Eric Kolaczyk - Boston University (DSERT co-chair)

Eric Kolaczyk – Boston University (DSERT co-chair)

- Professor, Department of Mathematics & Statistics
- Founding Member, Faculty for Computing & Data Science
- Affiliated Faculty, Bioinformatics, Computational Neuroscience, and Systems Engineering

- Director, Hariri Institute for Computing
- Founding Director, MS in Statistical Practice program
- Former Director, Program in Statistics

DSERT Goals

- Bring together representatives from CS, Mathematics, Statistics, and Domain Fields;
- In coordinated discussion among academia, government, and industry;
- To share and disseminate information on new courses, programs, and related initiatives;
- As a means towards identifying both common challenges and emerging best practices in post-secondary data science education.

DSERT Members

Co-Chairs: Eric Kolaczyk (BU) and Kathy McKeown (Columbia)

- John Abowd, Cornell (Census)
- Deb Agarwal, LBNL
- Ron Brachman, Cornell Tech
- Jeff Brock, Yale
- Alok Choudhary, Northwestern
- Tom Ewing, Virginia Tech
- Emily Fox, U Washington
- James Frew, UCSB
- Constantine Gatsonis, Brown U
- Lise Getoor, UCSC
- Mark Green, UCLA
- Alfred Hero, U Mich
- Nick Horton, Amherst Col.
- Eric Horvitz, Microsoft
- Bill Howe, U Washington
- Charles Isbell, Ga Tech
- Mark Krzysko, DOD
- Rachel Levy, MAA
- Brandeis Marshall, Spelman
- Chris Mentzel, (former) Moore Found
- Nina Mishra, Amazon
- Deborah Nolan, UC Berkeley
- Peter Norvig, Google
- Antonio Ortega, USC
- Claudia Perlich, Dstillery
- Patrick Perry, NYU
- Mehran Sahami, Stanford
- Victoria Stodden, UIUC
- Duncan Temple-Lang, UCD
- Uri Treisman, UT Austin
- Mark Tygert, Facebook
- Jeff Ullman, Stanford
- Jessica Utts, UC Irvine
- Jane Ye, NIH

Modus Operandi

- Twelve quarterly meetings, w/ pre-defined themes
- Combination of speakers, panels, and break-out sessions
- All proceedings web-cast live
- Slides and video archived on NASEM website
- NASEM-prepared written summaries available by the following quarter.

Accessing Content

Every Meeting was Live Streamed, Archived & Summarized

See: https://www.nationalacademies.org/our-work/roundtable-on-data-science-postsecondary-education

DSERT Meetings

- Dec 14, 2016: **Foundations** of Data Science from Statistics, Computer Science, Mathematics, and Engineering
- Mar 20, 2017: Examining the Intersection of **Domain Expertise** and Data Science
- May 1, 2017: Data Science Education in the **Workplace**
- Oct 20, 2017: **Alternative Mechanisms** for Data Science Education
- Dec 8, 2017: Integrating **Ethics and Privacy** Concerns into Data Science Education
- Mar 5, 2018: Increasing **Reproducibility** by Teaching Data Science as a Scientific Process

DSERT Meetings

- June 13, 2018: Programs and Approaches for Data Science Education at the **PhD Level**
- Sept 17, 2018: Challenges and Opportunities to Better **Engage Women and Minorities** in Data Science Education
- Dec 10, 2018: Motivating Data Science Education through **Social Good**
- Mar 29, 2019: Improving **Coordination between Academia and Industry**
- June 12, 2019: Data Science Education at **Two-Year Colleges**
- Sept 27, 2019: **Steps Forward**

Themes that Emerged from our Three Years of Discussions

- Foundations of Data Science
- Data Science Across the Curriculum
- Data Science Education and Non-Academics
- Social and Ethical Dimensions of Data Science

Foundations of Data Science

- Arguably most mature area of DS education.
- Currently lots of variability in the learning experience of foundations, both within ‘core’ areas and across domains.
- Despite connotations of permanence in ‘foundations’, organizational agility/flexibility is key.
- Long term goal: Intellectual common core

E.g., see meetings on

- **Foundations (#1)**
- **Domain areas (#2)**
- **Reproducibility (#6)**
- **PhD Programs (#7)**

Data Science Across the Curriculum

- What do we want our students to learn?
- What is pedagogical 'best practice' for DS?
- Implications for institutional capacity.
- Different external/internal rates of evolution.
- Data science for all – what role the university?

E.g., see meetings on

- **Foundations (#1)**
- **Domain areas (#2)**
- **Alternative Mechanisms (#4)**
- **Engaging Women & Minorities (#8)**
- **Two-year Colleges (#11)**

Data Science Education and Non-Academics

- Industry/government can augment relevance of DS education through partnership with academia.
- Productivity/value in industry/government can help inform academia of impact of education.
- In situ immersion of student learning in complex, fluid environments is critical for heightened employment preparedness.

E.g., see meetings on

- **Workplace (#3)**
- **Alternative Mechanisms (#4)**
- **Social Good (#9)**
- **Academic/Industry Collaboration (#10)**

Social and Ethical Dimensions of Data Science

- Ethics should be integrated throughout any DS curriculum.
- How best to teach DS ethics needs research.
- Because DS impacts all, all have to be involved in DS education.
- Use the experience of using data science to draw as large a body of students as possible.

E.g., see meetings on

- **Foundations (#1)**
- **Domain areas (#2)**
- **Ethics & Privacy (#5)**
- **Engaging Women & Minorities (#8)**
- **Social Good (#9)**

Our Goal for Today

- To not only summarize and help orient you to our 3 years of discussions, but also give you a feel for them in action;
- And, in doing so, tie the various themes that emerged to current events.

Kathy McKeown – Columbia University (DSERT co-chair)

- Henry and Gertrude Rothschild Professor of Computer Science at Columbia University
- Founding Director of Columbia's Data Science Institute, serving as Director from 2012 to 2017
- Research area: natural language processing
  - Text summarization, social media analysis, multilingual analysis, NLP for social good

Why were the DSERT conversations relevant to you and your work?

- For universities building educational programs, revealed the different ways that data science programs can be developed at the undergraduate and graduate levels
- Highlighted the need for interaction between all disciplines, particularly between foundational and domain areas (e.g., political science, history, science)
- Brought out interests in data science for social good and ethical issues of learning from data

Kathy McKeown – Columbia University (DSERT co-chair)

Key takeaways

- To get a program in place, compromise is important – especially at the PhD level
- Important to build interdisciplinary bridges in a way that works at your university
- Ethics is an important component of every data science program

What's next?

- Apply data science for social good in a timely fashion
  - How to identify misinformation about COVID-19?
  - How to identify bias in data when attempting to address racial equity?
  - How to avoid developing applications that encode gender and racial bias?

Nick Horton - Amherst College

- Beitzel Professor of Technology and Society (Statistics and Data Science)
- Co-chair, National Academies Committee on Applied and Theoretical Statistics
- Member, Consensus study committee "Data Science for Undergraduates: Opportunities and Options"
- Former chair, Committee of Presidents of Statistical Societies (COPSS)

- Why were the DSERT conversations relevant to you and your work?
  - Data acumen as a liberal art (making sense of the data in our world)
  - Importance of diversity and inclusion as we build new programs
  - How to think about data science for social good/data ethics

Nick Horton - Amherst College

- Key takeaways
  - Faculty development is a huge challenge
  - Cloud computing revolution impact only just starting
  - We risk repeating past mistakes if we don't act in a planful manner

- What's next?
  - Focus on data ethics: what does every citizen need to know?
  - Focus on diversity and inclusion: making data science education available to all
  - Focus on two year colleges: role of certificates, transfer programs, and terminal associates degree to address workforce challenges
  - Focus on assessment: how do we ensure that students are learning what we teach?

Mark Krzysko – Acquisition Data and Analytics, OUSD(A&S) Acquisition Enablers

- Krzysko serves as the Principal Deputy for Acquisition Data and Analytics. In this senior leadership role, Mark enables the Department to make sound business decisions with data. He is leading a philosophical and technical transformation within the Department to make timely, authoritative acquisition information available to support insight and decision-making on the Department of Defense's major programs—a portfolio totaling approximately $2 trillion of investment funds. He holds a Bachelor of Science Degree in Finance and a Master of General Administration, Financial Management, from the University of Maryland University College, and numerous certificates from Harvard University.

- Why were the DSERT conversations relevant to you and your work?
  - Mark's focus is on transforming the way the Department thinks about and uses data related to the business of defense acquisition. As the senior executive for all aspects of management of the Department's acquisition data, Mark leads data governance, provides access to data, and delivers and supports data science initiatives to enable acquisition analysis and decision making across the Department.

Mark Krzysko – Acquisition Data and Analytics, OUSD(A&S) Acquisition Enablers

- Key takeaways
  - Government and industry can help shape a curriculum for data science education beyond the “intellectual common core”
  - Data-driven enterprises have diverse challenges that could allow for practical application of data science education
  - There are ethical considerations around data collection and transparency that influence decision making

- What’s next?
  - Train the existing government workforce
  - Continue key government partnerships with academia
  - Cultivate a culture of analytics

Ray Levy – Mathematical Association of America

- Former Professor of Mathematics and Associate Dean at Harvey Mudd College.
- Recruited and advised industrial mathematics capstone projects.
- Co-founded the Business, Industry and Government (BIG) Mathematics Network (bigmathnetwork.org) and Mathematical Modeling Hub (qubes.mmhub.org).
- Author of BIG Jobs Guide and BIG Career Card Game.

Why were the DSERT conversations relevant to you and your work?

- Exchanged information between programs such as MAA Preparation for Industrial Careers in Mathematics (PIC Math) and programs at other institutions.
- Gained insight into interrelated workforce issues in Academia and Industry.
- Dove into the complexity of addressing ethical and social issues in the classroom.

Ray Levy – Mathematical Association of America

Key takeaways

- Data science education needs **policies that facilitate cooperation** between disciplinary silos.
- We need to teach mathematics as if computers exist, while providing mathematical modeling as a way of making predictions/recommendations and proof as a fundamental way of knowing.
- The future will likely hold more fluidity between academic and industry jobs.

What's next?

- Increase in internship opportunities and sabbatical options
- Earlier math modeling, statistics and data science learning opportunities
- Policies that foster communication and cross-disciplinary collaboration
- More attention on ethics, reproducibility, transparency

Nina Mishra – Amazon

- Industry: Scientist at Amazon, Microsoft Research, HP Labs
- Academia: University of Virginia CS Faculty; Acting Faculty, Stanford
- Interests: ML algorithms and data

- Why were the DSERT conversations relevant to you and your work?
  - Influence that industry could have on data science education

Nina Mishra – Amazon

- Key takeaways
  - To thrive, data science requires diversity. Lack of diversity can perpetuate into products.
  - Variety of Data Science + X majors. Importance of Π vs T-shaped programs
  - Sincere desire that students and professors spend more time in industry

- What's next?
  - Find new ways to attract diverse talent to data science
  - Encourage students to apply for internships and faculty to hold part-time, concurrent appointments in industry

Q&A

Please submit your questions below!