

Enriching Data Science R&D Projects through Strategic Engagements

Chaitan Baru

Senior Advisor for Data Science

Computer and Information Science and Engineering Directorate

National Science Foundation



Opportunities for International Engagements

- Translational Data Science
 - Ethics and Policy
 - “Blue Collar” data science
- The Open Knowledge Network
 - Data and knowledge as infrastructure
- Data Science Education and Training
- Data Science Corps



Putting it all Together: Translational Data Science

Application of data science techniques, tools, and technologies in science and other applications domains

Foundations

Cyber

- 1st Workshop on Translational Data Science, June 26-27, 2017, U.Chicago (Robert Grossman, Chicago; Raghu Machiraju, OSU)
- 2nd Workshop on TDS, November 13-14, UC Berkeley (David Culler, Berkeley)
- 3rd Workshop on TDS, planned for ~March 2018 (Juliana Freire, NYU)

algorithms

Workforce



Translational Data Science challenges: From Workshop 1

- Responsible data science
- Data quality
- Best practices around data triage and cost planning with respect to scale, quality, freshness, and heterogeneity
- Data and model commons



TDS Challenges: From Workshop 2

- Recognizing the data science researcher and professional
- Telling the data science story
 - Explaining what analysis/analyses were performed, and why
- Telling the “application story” with the data
 - What the data could possibly do for you
- “Blue collar” data science
 - Not Google, Amazon, Facebook, Twitter problems
 - E.g Honda in the US Midwest



Big Data and Data Policies

- Ethics and policies reflect local norms and regulations
 - Divergence of issues
- Examples
 - EU has passed the GDPR (General Data Protection Regulation)
 - US has Federal Information Security Modernization Act (FISMA) and Health Insurance Portability and Accountability Act (HIPAA)
 - India: Supreme Court passed the citizens “Right to Privacy”
 - Implications for the country's biometric identification program (Aadhaar)
 - Japan, China, etc ...
- Data Policy: A ripe area for international research



Data and Knowledge as Infrastructure: The Open Knowledge network

- An open, community-driven web-scale knowledge network
 - Initiated by the community: Andrew Moore, CMU; Ramanathan Guha, Google, et al
- Semantically-linked concepts, data
 - Foster research on a new class of applications leveraging data, context, and inferences from data
- Rich interfaces to data/knowledge
 - Question/answer; Dialog-based; Explanatory and Story-telling interfaces



The Open Knowledge Network

- Joint academia, industry, government workshops
 - July 2016, Washington, DC
 - Feb 2017, Sunnyvale, CA
 - Oct 4,5, 2017, National Library of Medicine, Bethesda, Maryland
 - System architecture/software; data representation/curation; research related to representation and use of massive knowledge graphs
 - Domains discussed:
 - Biomedical, Finance, Geoscience, Manufacturing



Data Science Education and Training

- Envisioning the Data Science Discipline: The Undergraduate Perspective, National Academy of Sciences, study/workshops
 - <https://www.nap.edu/catalog/24886/> (Interim Report)
- Keeping Data Science Broad: Negotiating the Digital and Data Divide, Oct.31-Nov.1, 2017, Atlanta, GA (Renata Rawlings-Goss, GaTech)
- Fostering community groups for:
 - Defining Data Science curriculum (undergraduate and graduate levels)
 - Developing program evaluation and assessment methods



Data Science Corps

Getting your hands dirty with data!

- **VISION**

- Provide practical experiences, teach new skills, and offer teaching opportunities in data science to U.S. data scientists and data science students, in the service of science and society

- **MISSION:**

- Enable U.S. data scientists and data science students to obtain practical experience with data-intensive applications;
- Promote a better understanding of the power of data, and the role that data can play in addressing issues at the local, regional, national, and international levels;
- Teach data literacy and provide basic training in data science to the existing workforce in communities, organizations, and institutions at the local, state, national, and international levels



Volunteer Organizations

Project Organizations

Graduate programs

Undergraduate programs

Universities, other

Skills and expertise
Varying levels of skills, expertise, and experiences

- First Data Science Corps workshop, Dec 7-8, 2017, at Georgetown University
- Attendees included: US academic institutions; IBM, Intel Foundation, SAS Foundation, Bloomberg, ESRI, Kaplan, DataKind, as well as Asian Development Bank, UNICEF, World Bank, ITU, Data for Sustainable Development Goals,

in:
research
&
ected
unities
h
Criminal Justice

Industry

NGOs

Professionals from Industry, NGOs

Federal Governments

International Organizations, e.g., WorldBank, UNICEF, ITU,

- Transportation,
- Energy,
- ..



Thank You!

- cbaru@nsf.gov

