Big Data Funding Opportunities

Jill Morris
Institute for Population Research
Slides updated Aug 2015
614-688-5423/ morris.856@osu.edu

Digging into Data Challenge

http://www.diggingintodata.org

- 10 research funders representing Canada, the Netherlands, the United Kingdom, and the United States **(NSF)**
- projects that explore how computationally intensive research methods can be used to ask new questions about and gain new insights into our world, funding organizations that are working together to focus the attention of the social sciences, humanities, library, archival, information, computer, mathematical, and statistical science communities on large-scale data analysis and its potential applications.
- we have massive databases of materials available for research in the humanities and the social sciences--ranging from digitized books, newspapers, and music to information generated by Internet-based activities and mobile communications, administrative data from public agencies, and customer databases from private sector organizations-—what new, computationally-based research methods might we apply? As the world becomes increasingly digital, new techniques will be needed to search, analyze, and understand these materials.
- Digging into Data challenges the research community to help create the new research infrastructure for 21st-century scholarship.
- Applicants will form international teams from at least two of the participating countries. Winning teams will receive grants from two or more of the funding agencies and, two years later, will be invited to show off their work at a special conference sponsored by the ten funders.

NSF Big Data Initiatives

NSF bigdata@nsf.gov

Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA) (NSF-15-544)

- The solicitation includes participation by the Mathematical and Physical Sciences (MPS) and Geosciences (GEO) Directorates at NSF
- The solicitation invites two types of proposals: *"Foundations" (F):* those developing or studying fundamental theories, techniques, methodologies, technologies of broad applicability to Big Data problems; and *"Innovative Applications" (IA):* those developing techniques, methodologies and technologies of key importance to a Big Data problem directly impacting at least one specific application. http://www.nsf.gov/cise/news/bigdata.jsp

What has been funded:
http://www.nsf.gov/awardsearch/advancedSearchResult?ProgEleCode=8083&BooleanElement=ANY&BooleanRef=ANY&ActiveAwards=true&

Big Data webcast
http://www.nsf.gov/news/news_videos.jsp?cntn_id=123607&media_id=72174&org=NSF

NSF Ideas Lab-Data Intensive Education-Related Research

Building Community and Capacity in Data Intensive Research in Education (BCC-EHR) (NSF 15-563)

- enable research communities to develop visions, for data-intensive EHR areas of research. In some cases large scale data repositories may already exist, but the infrastructure such as tools and communities to utilize the data may be in need of development. In other cases appropriate activities may include the design of large scale data repositories and/or associated analytic tools.

- Submitted proposals for FY 2015 should focus on the development of communities, or the utilization of existing communities, to develop plans for data repository design or utilization, and to develop infrastructure (including analytic tools) within which identified research may effectively proceed. http://www.nsf.gov/pubs/2015/nsf15563/nsf15563.pdf

Cyberlearning and Future Learning Technologies (NSF 14-526)

- NSF seeks to integrate opportunities offered by emerging technologies with advances in what is known about how people learn to advance three interconnected thrusts:

Innovation: inventing and improving next-generation genres (types) of learning technologies, identifying new means of using technology for fostering and assessing learning, and proposing new ways of integrating learning technologies with each other and into learning environments to foster and assess learning;

Advancing understanding of how people learn in technology-rich learning environments: enhancing understanding of how people learn and how to better foster and assess learning, especially in technology-rich learning environments that offer new opportunities for learning and through data collection and computational modeling of learners and groups of learners that can be done only in such environments; and

Promoting broad use and transferability of new genres: extracting lessons from experiences with these technologies that can inform design and use of new genres across disciplines, populations, and learning environments; advancing understanding of how to foster learning through effective use these new technologies and the environments they are integrated into.

- http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504984

NSF Computational and Data-Enabled Science and Engineering in Mathematical and Statistical Sciences (CDS&E-MSS) (PD 11-8069)

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504687&org=NSF

- supported by the CDS&E-MSS program accepts proposals that confront and embrace the host of mathematical and statistical challenges presented to the scientific and engineering communities by the ever-expanding role of computational modeling and simulation on the one hand, and the explosion in production of digital and observational data on the other.

- support fundamental research in mathematics and statistics whose primary emphasis will be on meeting the aforementioned computational and data-related challenges.

NSF Metadata for Long-standing Large-Scale Social Science Surveys (META-SSS) (NSF 11-583)

http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504705&org=NSF

- proposals that will develop tools to bridge data collection and dissemination by first, collecting and coding metadata associated with future waves of the ANES, GSS, and PSID surveys as collection and processing techniques evolve; and second, migrating (or "retrofitting") metadata associated with earlier (i.e., legacy) waves of these surveys into formats and schema that are compatible with current and future collection efforts.

- goal is to fund projects that will help make the many years of legacy data available to researchers who seek to answer current scientific questions.

- What Has Been Funded (Recent Awards Made Through This Program, with Abstracts) **http://www.nsf.gov/awardsearch/advancedSearchResult?ProgEleCode=1320,1331,1371&BooleanElement=ANY&BooleanRef=ANY&ActiveAwards=true&#results**

NIH Big Data Initiatives

NIH Big Data to Knowledge (BD2K) Initiative Research Education: Open Educational Resources for Sharing, Annotating and Curating Biomedical Big Data (R25)

http://grants.nih.gov/grants/guide/rfa-files/RFA-LM-15-002.html

- support educational activities that complement and/or enhance the training of a workforce to meet the nation's biomedical, behavioral and clinical research needs, by creating educational resources to foster sharing, curating, and annotating of biomedical Big Data.

- accomplish the stated over-arching goal, this FOA will support creative educational activities with a primary focus on:

  Curriculum or Methods Development: Development of open educational resources that cover concepts, approaches, relevant use cases and requirements for sharing, annotating and curating biomedical Big Data research resources, for use by librarians and other instructors to train researchers and graduate students for active roles in the connected biomedical enterprise.

NIH Big Data to Knowledge (BD2K) Development of Software Tools and Methods for Biomedical Big Data in Targeted Areas of High Need (U01) (RFA-CA-15-017)

http://grants.nih.gov/grants/guide/rfa-files/RFA-CA-15-017.html

Research Objectives and Scope

- This FOA solicits development of innovative analytical methods and software tools with the objective of addressing critical current and emerging needs of the biomedical research community for using, managing, and analyzing the larger and more complex data sets inherent to biomedical big data, focusing on the three topic areas listed below.
  - ***1. Data Privacy***
  - ***2. Data Repurposing***
  - ***3. Applying Metadata***

Accelerating the Pace of Drug Abuse Research Using Existing Data (R01)

http://grants.nih.gov/grants/guide/pa-files/PAR-13-080.html

- analyze cross-sectional or longitudinal behavioral, neuroimaging, and social science data from NIDA, NIAAA, or NCI funded studies, or derived from other sources..
- collaborate with investigators holding private data sets, use innovative statistical strategies to link methodologically comparable datasets, or utilize public use and administrative data readily available.
- analyses of data that have been harmonized and merged across multiple different datasets, such as those with PhenX measures (https://www.phenxtoolkit.org/).
- strongly encourage utilization of the National Addiction & HIV Data Archive Program (NAHDAP) hosted by ICPSR (http://www.icpsr.umich.edu/icpsrweb/NAHDAP/), as well as data from NIDA's Clinical Trials Network http://www.ctndatashare.org/.
- secondary analysis may be appropriate to all types of data, including qualitative information, and also covers the integration of quantitative and qualitative data.

THE OHIO STATE UNIVERSITY

INSTITUTE FOR POPULATION RESEARCH

Acknowledgments

This research was supported by grant P2C-HD058484 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development awarded to the Ohio State University Institute for Population Research.