MEASURING THE IMPACT OF DIGITAL REPOSITORIES: RECOMMENDATIONS

Prepared by the

BIG DATA INTERAGENCY WORKING GROUP

NETWORKING & INFORMATION TECHNOLOGY
RESEARCH & DEVELOPMENT SUBCOMMITTEE

COMMITTEE ON S&T ENTERPRISE

of the
NATIONAL SCIENCE & TECHNOLOGY COUNCIL

JULY 2018

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the Federal research and development enterprise. A primary objective of the NSTC is to ensure science and technology policy decisions and programs are consistent with the President's stated goals. The NSTC prepares research and development strategies that are coordinated across Federal agencies aimed at accomplishing multiple national goals. The work of the NSTC is organized under committees that oversee subcommittees and working groups focused on different aspects of science and technology. More information is available at http://www.whitehouse.gov/ostp/nstc.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976 to provide the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, homeland security, health, foreign relations, the environment, and the technological recovery and use of resources, among other topics. OSTP leads interagency science and technology policy coordination efforts, assists the Office of Management and Budget with an annual review and analysis of Federal research and development in budgets, and serves as a source of scientific and technological analysis and judgment for the President with respect to major policies, plans, and programs of the Federal Government. More information is available at https://www.whitehouse.gov/ostp.

About the Big Data Interagency Working Group

Federal agency members of the Big Data Interagency Working Group (BD IWG) coordinate research and development (R&D) focused on improving the management and analysis of large-scale data to extract knowledge and insight from large, diverse, and disparate data sources, including mechanisms for data capture, curation, management, visualization, and access. The BD IWG works under the auspices of the Networking and Information Technology Research and Development (NITRD) Subcommittee of the NSTC Committee on Science and Technology Enterprise to identify current big data R&D activities across the Federal Government and to offer opportunities for coordination of R&D activities among agencies, academia, and the private sector. More information about the BD IWG is available at https://www.nitrd.gov/groups/bd.

Copyright Information

Measuring the Impact of Digital Repositories

Background

There is a strong consensus in the scientific community that digital repositories are a critical component of the Nation's research infrastructure; at the same time, there is very little consensus on how to assess their impact.[1] Traditionally, digital data repositories have been relatively basic online storage spaces for raw data, but with changing data demands, they are expanding to include complex functionality and specific analytic tools. New incentives for data sharing are driving data repositories to become more interconnected and interdependent just as more researchers seek trusted places to deposit their research data. Best practices and methods to assess data repositories' impacts and effectiveness, as well as to consider innovative ways to ensure their long-term financial viability, are critical for advancing scientific research and for supporting the integrated Data Strategy under the President's Management Agenda.[2]

The Big Data Interagency Working Group (BD IWG) of the Networking and Information Technology Research and Development Subcommittee of the National Science and Technology Council's Committee on Science and Technology Enterprise identified the literature on these topics; held a workshop with representatives from academic institutions, journal editors, nonprofits, and government entities to share their experiences and perspectives; and evaluated and integrated that information into recommendations for research and development (R&D) to address the impact assessment and sustainability needs for data repositories.[3]

Recommendations of the Big Data Interagency Working Group

The BD IWG's principal recommendation for R&D related to data repositories is to develop new procedures and technologies to generate open, transparent metrics that can measure both the usage of data in digital repositories and the resulting impact they have on society. To accomplish this goal, the BD IWG suggests the following steps:

- Establish a representative body of leaders (managers, maintainers, users, and funders of major data repositories), possibly within a community-led organization, to recommend best practices, metrics, stewardship, preservation, and investment policies. Members should reflect both breadth and depth of expertise.
- Develop and implement methods for attribution of datasets to track how data is used and reused, and to establish these datasets as first-class research objects worthy of academic citation.
- Expand the use of trustworthy digital repository certification initiatives to ensure that a repository's digital material is authentic, reliable, accessible, and usable on a continuing basis.

[1] "Impact" is defined here to mean both a measure of a data repository's success (or value) to the various user communities and as a driver of the innovation that broadly benefits the U.S. economy and its citizens.

[2] https://www.performance.gov/PMA/PMA.html.

[3] The workshop was held on February 28 and March 1, 2017, in Arlington, VA. More information about the workshop is available at https://www.nitrd.gov/nitrdgroups/index.php?title=DigitalRepositories, including a list of the referenced resources and the participants' names and affiliations.

- Commission an organization such as the National Academy of Sciences to collect and organize information on the scientific, social, and economic impacts of data repositories. This organization would then develop a community consensus that considers the appropriate context for a given research community's repositories.
- Develop innovative strategies that ensure the financial sustainability of data repositories, such as the formation of an international repository consortium or the use of public-private partnerships between funding agencies and private organizations to build and maintain the data management infrastructure.