

The evolving U.S. policy environment for open research data

Heather Joseph

Executive Director, SPARC, Dupont Circle NW, Suite 800, Washington, DC 20036, USA

Tel.: +1 202 296 2296; E-mail: heather@sparcopen.org

Abstract. This paper presents a brief overview of emerging policies to open up access to research data in the United States. It provides a summary of the drivers behind policy development, outlines some key precedents that serve as a foundation for an emerging research data policy framework, examines the latest policies and suggests a few ways that the academic and research communities can work to ensure that these policies continue to evolve in a direction that actively supports the best interests of the research enterprise.

Keywords: Open Data, Data Sharing Policies, Open Access

1. Drivers towards open research data sharing policies

Around the world, research funders invest tens of billions of dollars each year on basic and applied scientific research. In the United States, the federal government invests over \$60 billion on such research. This investment is made with the expectation that it will stimulate new ideas, accelerate scientific discovery, improve educational outcomes, fuel innovation, grow the economy and create jobs, and in general, improve the welfare and well-being of the public.

Increasingly, research funders have realized that these outcomes can only be fully achieved if the outputs of this research – including data – can be freely accessed and used by the widest possible audience. Over the past decade, there has been a trend toward developing policies that utilize the power of networked digital technology, specifically the Internet and World Wide Web, to expand the dissemination and utility of research outputs. The working theory is that our expected outcomes will be accelerated and significantly improve if there are policies in place that encourage open access to the results of publicly funded research.

Funders – and the academic and research community at large – also see additional benefits to opening up access research data. These benefits include the opportunity to engage new communities in interacting with data, encouraging collaboration, avoiding duplication, improving reproducibility, preventing/improving response to crises (the Zika virus, earthquakes, climate change, etc.), and speeding the translation of research results into products and services, and getting those to market more quickly as crucial motivations for increased research data sharing.

In an era where improving the transparency and accountability of government has taken center stage, U.S. policymakers have recognized the need to create a policy framework that supports all stakeholders in a transition to a more open system of sharing research results.

2. Policy precedent/development

Fortunately, in the United States, we have a long history of information policy precedents that have laid a strong foundation for the creation of an effective research data-sharing framework. Dating back to the mid-1960s, expectations for sharing government-produced/funded information have been articulated in a number of well-known, key policy documents, including:

- Freedom of Information Act, 1966;
- Copyright Act, 1976;
- Paperwork Reduction Act, 1980;
- Office of Management and Budget (OMB) Circular No. A-130, 1985;
- Electronic FOIA Amendments, 1996;
- Paperwork Elimination Act, 2003.

In particular, the language contained in OMB Circular A-130 plays a key role in helping to define expectations for sharing digital data of all kinds. While the regulation was crafted prior to the Internet era, and was not designed to specifically address data sharing, it clearly outlines key principles for sharing government information that speak to the heart of the objectives of U.S. federal research funders. The circular underscores the benefits of broad access to information, directly noting:

“...Government information is a valuable national resource, and the economic benefits to society are maximized when government information is available in a timely and equitable manner to all.”

Additionally, Circular A-130 addresses the need for access to this critical information to be provided with cost barriers that are as low as possible, specifically calling for:

“Open and unrestricted access to public information at no more than the cost of dissemination”...

This has proven to be an extremely useful foundation to build upon, and one that the Obama Administration has taken full advantage of. On his first day in office in January 2009, President Obama issued a sweeping Open Government Directive, outlining guidelines for all federal agencies to adhere to in an effort to promote a more transparent and participatory government. The first concrete step that agencies were required to take was to publish government information online – and in open formats – to increase its accessibility and utility to the public.

From that moment forward, the Administration moved rapidly (in policy-making terms) toward issuing ever-more granular policies that tightened the focus to all digital government data. By 2013, the Administration had successfully issued a new Executive Order making “Open and Machine Readable” the default for all government data. Additionally, he honed directly in on research data, with an additional Directive from the White House Office of Science and Technology Policy (OSTP) requiring Public Access to Federally Funded Research Outputs in 2013.

3. Current policy environment

The OSTP Directive was a landmark statement. For the first time, there were specific requirements for Agencies to create policies to ensure that all digitally formatted scientific data resulting from unclassified

research supported by U.S. federal funding be stored and made accessible for the public to search, retrieve, and analyze.

The Directive provides an interesting and accurate window into the current state of research of data sharing policies. It reflects the deep diversity of data generated in various research disciplines, but (understandably) also contains a high level of ambiguity and often seems to be contradictory. For example, the Directive tasks agencies with the goal of “maximizing access to data,” while still fully protecting confidentiality and personal privacy, and recognizing proprietary interests, business confidential information and IP rights. Another example is the need to balance the value of long-term preservation and access with costs and administrative burdens. It is important to note that these apparent contradictions are not intended to confuse producers of research data but actually to serve as important indicators of areas where tension between the potential benefits of full open sharing of data runs directly into the potentially negative consequences of such sharing.

As of June 2016, draft or final policy plans for access to research have been released by fifteen of the nineteen U.S. science agencies covered by the OSTP Directive. It’s clear from their content that creating final policies will be an evolutionary process, requiring significant community involvement and input to work towards acceptable policies. Unlike in many other cases, creating research data sharing policies will not be a “one-and-done” drafting process.

To date, all of the plans released differ somewhat in interpretation of the OSTP guidelines, as well as in the implementation processes that they propose. On the positive side, there are lots of significant commonalities. Most U.S. agency policy plans:

- Require the submission of Data Management Plans at the proposal stage;
- Provide direction for approved locations for data deposit/storage;
- Acknowledge the need for routine attribution for data;
- Require data the creation of agency inventories & indices of data to aid discovery;
- Support public/private collaboration to achieve aims of data sharing policies;
- Recognize the importance of/need for robust long-term preservation strategies.

However, there is not yet a common set of standards for any of these policy components. For example, while all agencies require Data Management Plans to be submitted at the funding proposal stage, a common set of attributes/expectations for these plans do not exist. This is probably to be expected and will be addressed as agencies work through additional community consultations and rounds of input, but in the short term, this will make policy development more labor intensive for funders, and more labor intensive for funding recipients to comply with.

With a high level of ambiguity still a hallmark of these emerging policies there is an undercurrent of confusion over compliance from the institutions that are the primary recipients of federal research funding. However, there is also a general willingness on their part to work together with funders and the larger community on solutions to decrease compliance friction.

This confusion and concern might be interpreted as a signal that more granular policy direction is needed (i.e., via more prescriptive legislative language.) No bills directly related to this issue have been introduced in Congress, but it is certainly an area to watch. Additionally, the fluid political climate in this presidential election year also poses challenges. For example, we have a potential presidential candidate who has indicated that all of the current Administration’s Executive Actions would be overturned – and that would include the key Open Government and Data Directives described earlier.

4. What can we do to keep things moving in the right direction?

This complex and somewhat volatile environment highlights the need for regular and close collaboration not only among funders, but also with the academic and research community at large, in order for research data sharing policies to evolve at paces and in directions that are acceptable to and sustainable by the research enterprise.

It also underscores the necessity of an evolutionary policy development approach to produce effective research data policies in different disciplines. Taking an approach that emphasizes regular pilots to test assumptions, and also provide mechanisms to gather and incorporate community feedback on those pilots could be effective in this environment.

“Harmonizing” policy components across U.S. Federal agencies wherever possible to reduce operating friction should be a key shared goal. Creating consistent policy components, and ongoing interagency collaboration on implementation requirements is also vital for the ultimate success of these policies.

Collaboration with the community will be helpful in promoting a reasonable level of standardization to facilitate smoother compliance process. As an example, SPARC and the Johns Hopkins University Libraries recently produced a free, online resource that tracks research data sharing policy components across all of the U.S. Federal Agencies, and allows users to compare requirements, building an understanding of these policies and helping to reduce confusion.

Finally, all of this will be moot if we are unable to effectively support and sustain the infrastructure necessary to make data sharing a reality. Community collaboration in developing advocacy initiatives to call for additional investments in the infrastructure needed to support access to and use of research data for the long term will be essential. We cannot build an effective sustainable infrastructure to support vital national interests without additional investments – and without this infrastructure, our collective ability to achieve the laudable objectives of open research data sharing will be negated.