

RESEARCH PAPER

An Analysis of Federal Policy on Public Access to Scientific Research Data

Adam Kriesberg¹, Kerry Huller¹, Ricardo Punzalan¹ and Cynthia Parr²¹ College of Information Studies, University of Maryland, College Park, MD, US² National Agricultural Library, US Department of Agriculture, Beltsville, MD, USCorresponding author: Adam Kriesberg (akriesberg@gmail.com)

The 2013 Office of Science and Technology Policy (OSTP) Memo on federally-funded research directed agencies with research and development budgets above \$100 million to develop and release plans to increase and broaden access to research results, both published literature and data. The agency responses have generated discussion and interest but are yet to be analyzed and compared. In this paper, we examine how 19 federal agencies responded to the memo, written by John Holdren, on issues of scientific data and the extent of their compliance to the directives outlined in the memo. We present a varied picture of the readiness of federal science agencies to comply with the memo through a comparative analysis and close reading of the contents of these responses. While some agencies, particularly those with a long history of supporting and conducting science, scored well, other responses indicate that some agencies have only taken a few steps towards implementing policies that comply with the memo. These results are of interest to the data curation community as they reveal how different agencies across the federal government approach their responsibilities for research data management, and how new policies and requirements might continue to affect scientists and research communities.

Keywords: open data; office of science and technology policy; data access; metadata

Introduction

On February 22, 2013, John Holdren, Director of the Obama Administration's Office of Science and Technology Policy, issued a memo to the heads of other government agencies entitled "Increasing Access to the Results of Federally Funded Scientific Research" (Holdren, 2013). This document outlines a vision for both academic papers and scientific data requiring federal agencies with annual research and development budgets above \$100 million to draft a plan explaining how they will increase access to the research products produced with federal support. The memo identified eight elements to be included in agency responses. These elements reflect the values and perspective of the Obama administration on federal science, and suggest how the affected agencies, ranging from the Department of Defense to the US Geological Survey, might respond to the Holdren memo. The effects of new agency policies will be felt by intramural (government employed) and extramural (grant-funded) researchers into the future, therefore making these plans and responses important harbingers of what is to come when plans are fully implemented.

The memo has generated much discussion in academia and the popular press. Many of these are speculative or editorial in nature (e.g. Berman & Cerf, 2013; Corneliussen, 2016; Franceschi-Bicchierai, 2013; Murphy, 2016; Van Noorden, 2013), but some cite the Holdren memo as inspiration or evidence of the importance of research data. In particular, the memo stimulated the examination of data sharing practices and open access policies (Bishoff and Johnston, 2015; Van Tuyl and Whitmire, 2016). Current published literature, however, has not presented an analysis and comparison of what the plans indicate the agencies will do to increase access to scientific data and literature.

In the memo, Holdren outlines a vision for dissemination of the knowledge generated through the federal government's research funding infrastructure. He charges agencies funding research above the \$ 100 million threshold to develop eight specific elements of a plan around federally-funded research. These are:

1. A strategy for leveraging existing archives, where appropriate, and fostering public-private partnerships with scientific journals relevant to the agency's research;
2. A strategy for improving the public's ability to locate and access digital data resulting from federally funded scientific research;
3. An approach for optimizing search, archival, and dissemination features that encourages innovation in accessibility and interoperability, while ensuring long-term stewardship of the results of federally funded research;
4. A plan for notifying awardees and other federally funded scientific researchers of their obligations (e.g., through guidance, conditions of awards, and/or regulatory changes);
5. An agency strategy for measuring and, as necessary, enforcing compliance with its plan;
6. Identification of resources within the existing agency budget to implement the plan;
7. A timeline for implementation; and
8. Identification of any special circumstances that prevent the agency from meeting any of the objectives set out in this memorandum, in whole or in part.

In this paper, we focus on the scientific data produced through federal research funding. Data-specific issues are part (either implicitly or explicitly) of each point above, but items #2 and #3 are most directly relevant for the data curation community. Here, Holdren asks federal agencies to identify how they will ensure preservation and access to federally-sponsored research data. While much of the research in data curation focuses on individual researchers, institutions, and disciplinary approaches to data curation and management, this memo and its responses provide an opportunity to examine how the federal government of the United States understands data curation and what it plans to do to address some of the major obstacles to making research data accessible, discoverable, and interoperable.

The Holdren memo, and its subsequent responses from 19 agencies for which we were able to locate publicly available plans, reveals the current state of data curation thinking and activity in the federal research and development community. Actions taken to date and plans for the future reveal that, while the word "curation" was not frequently used, the issues that are important to the digital curation community are very much at the forefront of the conversation about scientific research data in the federal government. Our results establish a framework for future analyses of the current and future work of federal agencies to satisfy the memo, and will be useful to policy-makers as well as the data curation research and practice communities, as members of these groups will increasingly interact with researchers who will be affected by new protocols around data management and sharing. However, the current administration's approach regarding public access to research data may change; thus the results presented in this paper cannot yet be compared to agency activities beyond 2016 until the OSTP issues additional policies.

Research Questions

The Holdren memo has generated significant amounts of discussion but relatively little comprehensive research has been undertaken to analyze its effects. Our study is motivated by the following research questions:

1. What are the ways federal science agencies responded to the Holdren memo?
2. How do these responses satisfy the requirements of the Memo?
3. What do the Holdren memo responses say about the state of access to federal scientific research data?

To answer these questions, we analyzed and compared agency responses to the 2013 memo, the details of which are discussed in the Methods section later in the paper.

Background

The United States federal government has had a long history of supporting scientific research, dating back to the establishment of the Smithsonian Institution in 1846 and its early support for research in such fields as natural science and meteorology (Fleming, 1989). During the 20th century, federal funding for

scientific research vastly increased with the establishment of the National Science Foundation in 1950 (National Science Foundation, 2016) and the expansion of other federal funding agencies such as the National Institutes of Health (Harden, 2009). Since 1957, the President has included a Science Advisor in the administration (Killian, 1977); in 1976 the Office of Science and Technology Policy was founded to broaden the scope of advice available to the executive branch on a range of science and technology related topics. Part of OSTP's mission includes the development and enforcement of science policy (Office of Science and Technology Policy, 2016).

Since OSTP issued its memorandum in February 2013, affected agencies have been developing and releasing final access plans to the public. Despite language in the memo prompting affected agencies to submit draft plans within six months of the memo being published, to date not all of the final plans have been made public. Meanwhile some agencies not required to comply with the memo have voluntarily completed access plans to address both literature and data (CENDI, 2016).

Historically, scientists have sought access to the journal articles that result from research projects (English and Raphael, 2006; Kaiser, 2006; Nickum, 2006; Peek, 2009). Nickum describes the history of access to federal scientific and technical literature and evaluates the access that the Department of Energy and the Environmental Protection Agency have provided over time. Just as the Holdren memo is considered to be another unfunded mandate, she explains that Congress has fallen short of providing sufficient support for scientific and technical publishing since the 1920s. Given this lack of support for disseminating the results of research, researchers began turning to commercial publishers and professional societies for opportunities to publish scientific papers and reports (Nickum, 2006).

In recent years, discussion has increasingly focused on public access to the articles resulting from federal research as well as ongoing legislative efforts regarding public access to research products (Berman and Cerf, 2013; English and Raphael, 2006; Frankel, 1999; Kaiser, 2006; Peek, 2009; Van Noorden, 2013). The Wellcome Trust in the UK was the first funding agency to require open access to research it supported (Wellcome Trust, 2016); the National Institutes of Health soon followed suit as the first U.S. federal agency to adopt a voluntary public access policy for peer-reviewed literature in 2005 (English and Raphael, 2006). A report by NIH in January 2006, however, found that less than 4% of grantees were complying with the voluntary policy, so an NIH advisory committee recommended that it be made mandatory (Kaiser, 2006). Congress subsequently required NIH to carry out the mandatory deposit of published articles in the Consolidated Appropriations Act of 2008 (National Institutes of Health, 2008).

Before NIH's mandatory policy was enacted, Senators John Cornyn and Joe Lieberman introduced the Federal Research Public Access Act of 2006 (FRPAA), which would have codified into law the NIH policy, but also would have applied to all agencies spending over \$100 million in research (English and Raphael, 2006; Harvard Open Access Project (HOAP), 2015; Kaiser, 2006). FRPAA would require researchers to submit final peer-reviewed articles into an agency or agency-approved repository within 6 months after publication (English and Raphael, 2006). FRPAA was introduced again in 2009 and 2012, but was never brought up for a vote (Harvard Open Access Project (HOAP), 2016; Peek, 2009). In 2015, a modified version of FRPAA, the Fair Access to Science and Technology Research (FASTR) Act was introduced, and is working its way through Congress (Harvard Open Access Project (HOAP), 2016).

Coinciding with NIH's policy and the repeated introduction of legislation, patient groups and other stakeholders in the medical research community increasingly began to demand free public access to journal articles that their taxpayer dollars had funded (Van Noorden, 2013; English and Raphael, 2006; Kaiser, 2006). Members of the public not affiliated with an academic library or research institution would have to spend upwards of \$25 per article or find someone with university credentials to log in to a journal or database offering the article (English and Raphael, 2006). Additionally, English and Raphael (2006) point out that researchers do not have access to every existing journal and sometimes must use interlibrary loan, which becomes a costly and time-consuming enterprise for articles that may end up being irrelevant to their research.

Despite the overall focus on public access to literature, some, (e.g. Barata, 1996) have contended that federal research, including records of laboratory proceedings, should be available not only for scholars, but "to provide a means of ensuring project accountability both fiscally and ethically" (p. 129). Senator Richard Shelby, while arguing for increased requirements making public the data which inform changes to federal policy, highlighted the role government could play in changing the culture of scientific data management (2000), and Heafey (2011) mentioned the importance of data in the overall ecosystem of the research lifecycle and public access to peer-reviewed papers. These points reflect similar arguments regarding the management of federal research made in the Holdren memo (Holdren, 2013), such as reducing duplication, guaranteeing accountability and compliance.

Following the release of the Holdren memo, there have been numerous guidance materials published to help grantees navigate the new requirements at federal agencies. Many universities are providing as much instruction as possible from the agency plans regarding papers and data, but this is limited to how much the plans explain. A crowd-sourced Google spreadsheet that summarizes all of the federal agency responses is continuously updated, primarily by data specialists at academic libraries, but anyone may add to it (Whitmire et al., 2015). CENDI (The Commerce, Energy, NASA, Defense Information Managers Group), a federal group consisting of scientific and technical information managers from many federal agencies beyond those mentioned in the organization's title, provides access to the available public access plans on its website, as well as the effective dates that researchers must abide by for each agency (CENDI, 2016). A joint project between the Scholarly Publishing and Academic Resources Coalition (SPARC) and Johns Hopkins University provides an online tool for comparing agency policies, which pulls out specific references to things like data management planning, data preservation, and metadata from a variety of agency documents (Scholarly Publishing and Academic Resources Coalition, 2016). All of these resources provide useful direction to grantees during this period of transition, but do not actually evaluate the agency plans themselves or provide analysis. Furthermore, they largely focus on peer-reviewed literature at the expense of discussing research data management, curation, preservation, and access. In this paper, we present a more comprehensive analysis of data-related elements of the agency plans that have been released, with the goal of providing insight into each agency's approach to the Holdren memo as well as a broader understanding of data management and curation policy across the federal government.

Methods

This project analyzes the 19 publicly available responses to the 2013 Holdren memo. We collected these documents from agency websites, and used two sources to verify that we had discovered the correct versions for each agency: the Commerce, Energy, NASA, Defense Information Managers Group (CENDI) list "Public Access (PA) Plans of U.S. Federal Agencies" (CENDI, 2016) and the Scholarly Publishing and Academic Resources Coalition (SPARC) browser and comparison tool for data sharing requirements by federal agencies (Scholarly Publishing and Academic Resources Coalition, 2016). The documents, issued by each federal agency with over \$100 million annual budget for research and development as well as additional agencies with a science mission but budget below this threshold, outline how the agency will make the published results of federally-funded research available within a year of publication and how they will make data associated with those publications publicly accessible (Howard, 2013). The agencies can be found in **Table 1** below. Links to the 19 agencies studied as well as versions of the documents used in our analysis can be found in Appendix 1. Our analysis focused on the data portion of each plan.

We examined the public access plans using 15 thematic codes (Miles and Huberman, 1994, p. 58) developed by the research team (Appendix 2). The initial codes were based on the structure and format of the Holdren memo itself. A preliminary reading of sample plans suggested the value in adding additional codes for issues relevant to digital curation and preservation research, such as "Digitization/Legacy Data" and "Intramural/Extramural Policy". For each theme, the agency plan was assigned a score on a three-point scale, indicating that a theme was not mentioned (1), a theme was briefly discussed (2), or a theme was thoroughly discussed (3). To ensure consistent analysis, the two primary coders (KH and AK) each began by analyzing the same plan and calculating Cohen's Kappa to measure inter-coder reliability. The coders achieved a score of 0.92, indicating strong agreement on coding decisions. Following this test, each coder worked independently on half of the remaining plans. In addition to thematic coding, we identified passages from agency plans that illustrated different approaches taken to address one of the identified themes. These quotes are included in the following sections of the paper to show additional evidence from the plans beyond the results of the coding.

Results

Overall, federal science agency public access plans that have been developed in response to the Holdren memo demonstrate a range of approaches. While some agency plans indicate thoughtful policy roadmaps and procedures already in place for managing scientific data, our analysis suggests that a number of agencies are not currently able to provide access to federally-funded research data in the manner suggested by the Holdren memo.

An overview of the qualitative analysis conducted by the research team is displayed in **Figure 1** below. For each of the 15 themes, we determined whether agency plans do not cover the given concept at all, partially

Table 1: Agencies whose Public Holdren memo Responses were analyzed in this study.

Department of Agriculture (USDA)

National Institute of Standards and Technology (NIST)

Department of Energy (DoE)

National Oceanic and Atmospheric Administration (NOAA)

Agency for Healthcare Research and Quality (AHRQ)

Smithsonian Institution

Department of Transportation (DOT)

Centers for Disease Control (CDC)

National Science Foundation (NSF)

Department of Health and Human Services (HHS): Overview plan

Department of Veterans Affairs

HHS – Assistant Secretary for Preparedness and Response

National Aeronautics and Space Administration (NASA)

Food and Drug Administration (FDA)

National Institutes of Health (NIH)

Department of Defense (DOD)

US Agency for International Development (USAID)

US Geological Survey (USGS)

Department of Education Institute of Education Sciences (IES)

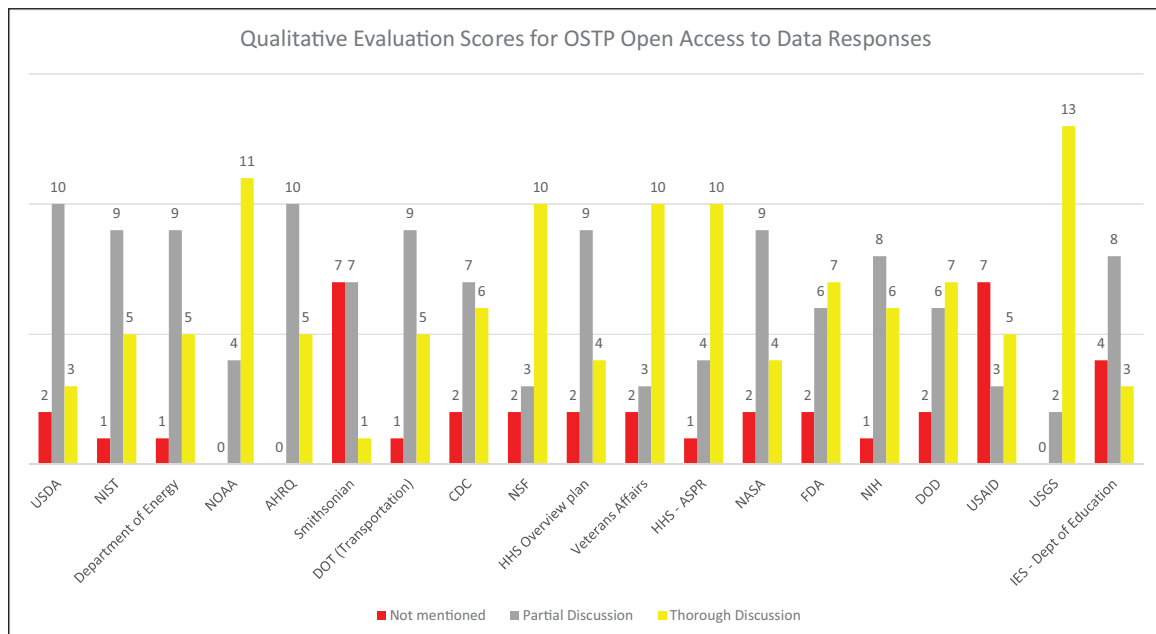


Figure 1: Qualitative Evaluation Scores for by Agency.

discuss it, or fully discuss the given concept. These results are shown below and allow for comparison of the coverage given each issue by responding agencies.

The five highest-scoring agencies (26% of those studied) received 10 or more scores of “Thorough Discussion” are: the National Oceanic and Atmospheric Administration (NOAA); the Department of Veterans Affairs (VA); the National Science Foundation (NSF); the Department of Health and Human Services Office

of the Assistant Secretary for Preparedness and Response (HHS-ASPR); and the United States Geological Survey (USGS). The two agencies with the most scores indicating “No mention” of a given theme are the United States Agency for International Development (USAID) and the Smithsonian. From our analysis, we determined that each of these agencies has omitted a discussion of seven themes of the Holdren memo. The Smithsonian’s plan was brief, particularly in its coverage of requirements for research data. It directs researchers to deposit data in the Smithsonian’s repository, Smithsonian Research Online (<http://research.si.edu>), or in a discipline-specific repository. While it lacks specific details about digital preservation standards or the ability for the public to provide feedback on the plan, it does highlight existing infrastructure that the institution plans to leverage in support of the Holdren memo’s mandates.

Of the five highest-scoring agencies, three (those from NOAA, the NSF, and USGS) are all commonly associated with government-sponsored scientific research. These agencies either had well-developed policies in place governing some of the concepts addressed in the Holdren memo, or they quickly adapted their existing plans to comply with the memo’s requirements. The other plans that fared best in our analysis are from the VA, which while not primarily associated with research, does manage significant medical studies in addition to being the largest healthcare network in the country. The complex regulatory framework in which the VA operates may have positioned it well to respond to executive memos. Finally, the HHS-ASPR is a relatively new office within HHS, having been established in 2006 with the passage of the Pandemic and All Hazards Preparedness Act (Office of the Assistant Secretary for Preparedness and Response, 2014). As such, it also may be more ready to respond to an administration document such as the Holdren memo. Furthermore, the office was established in the wake of Hurricane Katrina and is focused on increasing national readiness for public health emergencies, a goal to which data sharing contributes (Lurie et al., 2013).

A complimentary summary of scores by theme appears in **Figure 2** below. This figure shows whether agency plans addressed a concept at all, partially discussed it, or fully discussed each theme. For every theme with the exceptions of “DOIs/Unique Identifiers” and “Digitization/Legacy Data” very few agencies received scores indicating the absence of any discussion on a given theme. This may be because these two themes did not explicitly appear in the memo itself, but emerged during our initial analysis of the responses. For other themes related directly to the text of the memo such as “Access” and “Funding/Costs”, the scores indicate that most responding agencies sought compliance with the memo by mentioning the theme in some way.

Figure 3 provides an alternate visualization of our data. This figure displays a ranked list of the average scores given to every response for each of our 15 themes. In this format we highlight the areas in which the 19 agencies and offices analyzed have the most well-developed responses to the Holdren memo. The highest average score was 2.86, for Data Management Plans. For this code, agency responses received a “Thorough” score for clearly explaining the requirements for researchers around data management plans, the timeline for adding this to existing review and application processes, and for articulating how DMPs contribute to

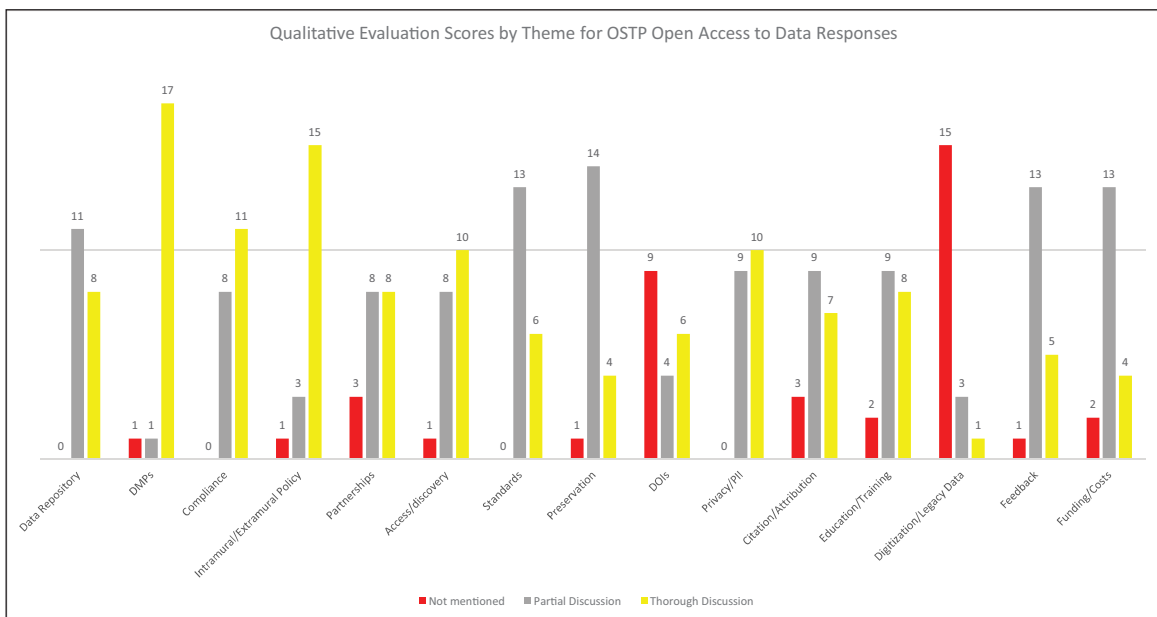


Figure 2: Qualitative Evaluation Scores by Theme.

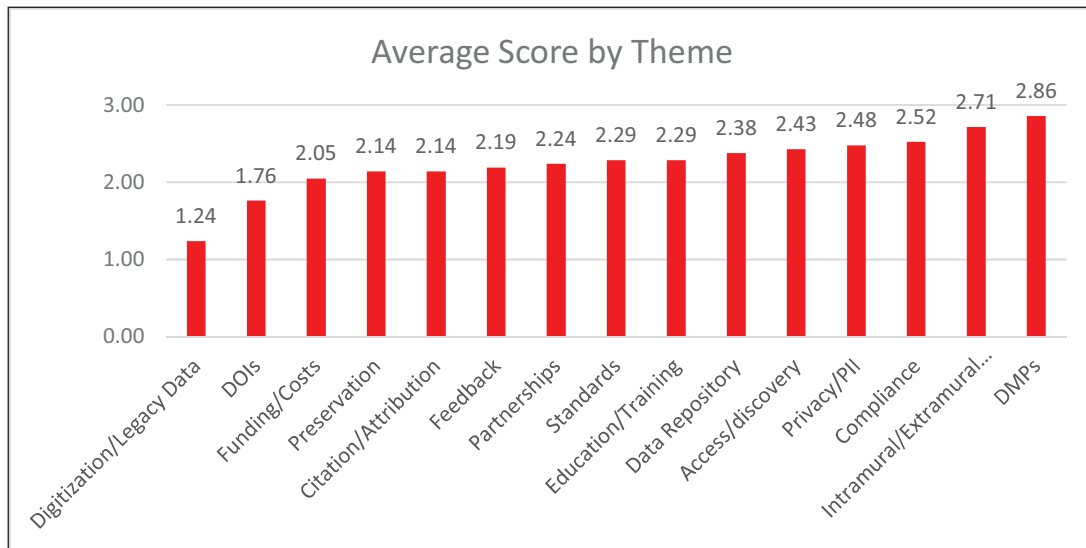


Figure 3: Average Score by Theme.

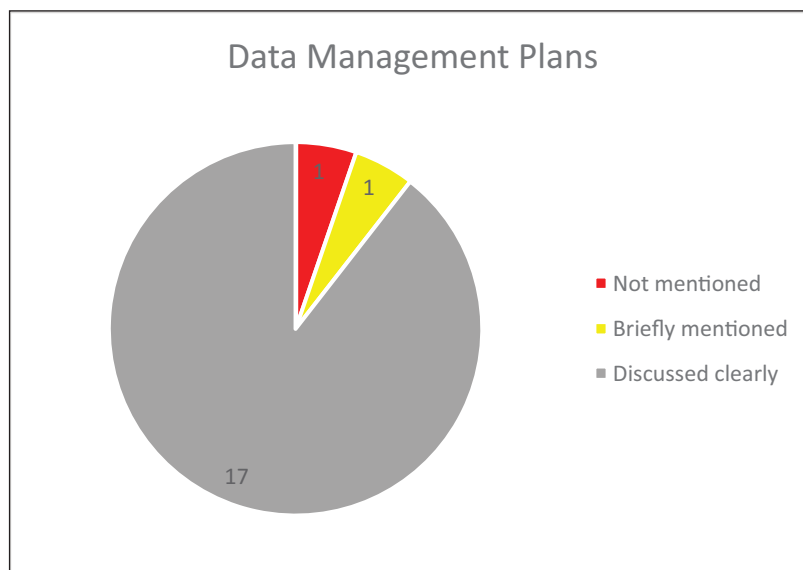


Figure 4: Agency Response Scores for “Data Management Plans”.

broader goals of information access. Seventeen plans were scored “Thorough” for this theme, one “Partial” and one was “Not mentioned”.

The language describing DMP requirements in NASA’s plan is representative of the language used in multiple agency responses. The plan specifies that the new policies will “include a requirement for all research proposers, intramural and extramural, to submit Data Management Plans (DMPs) with their proposals or project plans. DMPs will describe how the proposed research plan conforms to NASA policy on the dissemination and sharing of research results...” (NASA, 2014, p. 11). Following this short requirement, the plan goes on to specify what elements the agency expects in a DMP, such as an enumeration of file formats, metadata standards, and plans for depositing data in a repository. A breakdown of the scores for all agencies on this theme is presented in **Figure 4**.

The next three highest-scoring themes are “Privacy/PII”, “Compliance”, and “Intramural/Extramural Policy”. These themes all speak to some of the more generalizable activities of the agencies examined for this study. Many agencies, particularly those located in HHS or dealing with health and personnel information, were likely to already have policies governing the management of Personally Identifiable Information in place at the time the Holdren memo was issued. For example, the response for the Department of Health and Human Services Agency for Healthcare Research and Quality (ARHQ) outlines

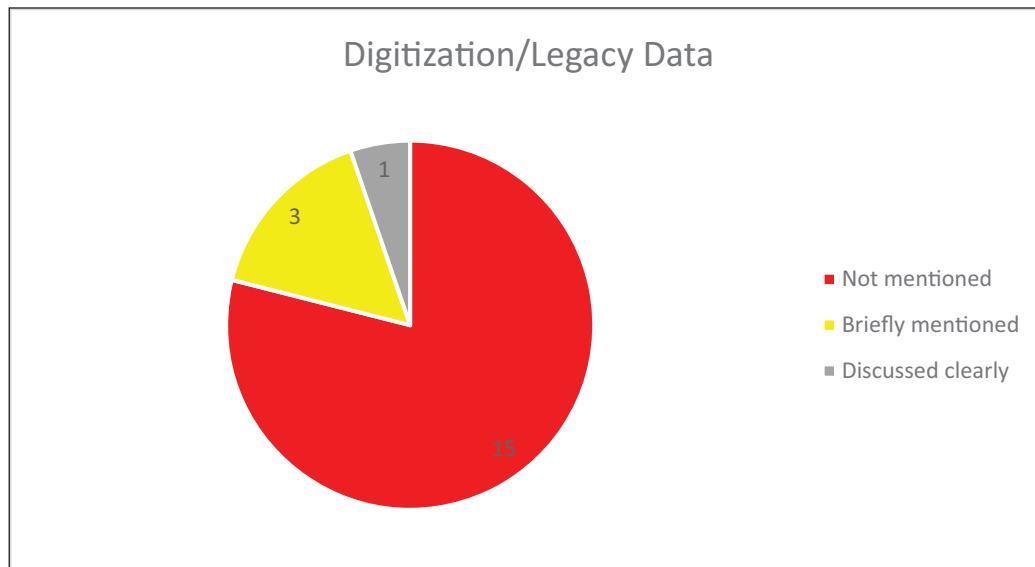


Figure 5: Qualitative Analysis Scores for "Digitization/Legacy Data".

the expectations and scope of data access for PII, stating: "AHRQ will make an attempt to make all research data available to the public by creating de-identified public use data files". The response goes on to provide more general guidance to grant applicants in their data management plan submissions: "AHRQ will require all data management plans to be consistent with applicable laws and regulations governing the privacy and confidentiality of individual human data". This language is typical of that used by other agencies for describing how to handle PII in federally-funded research. Other laws and regulations address these issues more directly and this is reflected in OSTP responses. This example illustrates the intersection between ongoing work in these agencies and their new responsibilities related to the Holdren memo.

By contrast, the lowest scoring theme was "Digitization/Legacy Data". This was not explicitly mentioned in the memo, which focused on current proposals and future research, but is a relevant topic for data access. Preliminary readings of some plans suggested that multiple agencies were concerned with the issue of what to do with historical data in analog or obsolete formats. Digitization of analog datasets can enable new research and allow scientists to understand past phenomena (Michellini et al., 2005). Ultimately we found that of the 19 agency plans analyzed for this study, four mention legacy data (**Figure 5**). Of those that did address this theme, the USGS covered it most completely. Its response reads:

Given the long history of the USGS, there is a wide array of legacy data assets... which could be made more accessible in digital forms. Programs that conduct data rescue and preservation activities constantly weigh the relative costs and benefits of... preserving these resources on the basis of current research needs and public demand for them. (U.S. Geological Survey, 2016, pp. 15–16)

Here, the USGS gives individual programs autonomy to decide if digitization efforts for legacy data would be worthwhile while acknowledging the broad utility of such work. Including this language in their public access plan allows the agency to return to digitization when desired and tie the work back to the Holdren memo. NOAA's approach to digitization presents a different perspective on efforts to convert analog data into machine-readable formats:

The Plan does not apply to non-archived legacy results that were produced by NOAA Programs which no longer exist, unless those results are deemed worthy of preservation by an existing Program which identifies the necessary resources for archiving. NOAA recognizes that this means some historical results may be lost, but NOAA does not have the resources necessary for comprehensive data rescue. (NOAA Research Council, 2015, p. 3)

While digitization and the provision of access to historical scientific data were not thoroughly addressed in agency responses, the sentiment expressed by NOAA on legacy data reflects an underlying theme running

through all 19 responses. While agencies are amenable to the idea of increasing access to publications and data resulting from federally-funded research, they continue to operate in resource-limited environments and recognize that they may not have the financial means to accomplish some of the goals outlined in the Holdren memo. This sentiment only appears explicitly in a few places across the responses but is nevertheless an emergent theme from our results. The Department of Transportation recognized their lack of resources and does not appear to be planning a repository of its own for data. The agency response reads “To minimize costs, DOT will encourage researchers to use publicly accessible databases for the deposit of their data” (U.S. Department of Transportation, 2015, p. 11). This reflects a recognition that the agency will need to rely on other infrastructure to fulfill the OSTP mandate and will need to craft additional policy instruments to incentivize researcher deposit into external repositories.

Discussion

The results of our analysis revealed that many federal agencies that support scientific research have taken a number of steps, in response to the Holdren memo, to broaden access to research data. We found that for 13 of the 15 themes coded, agency responses gave at least “partial discussion” to the topic. These results suggest an increased focus on research data management, preservation, curation, and access in the federal government. Ultimately, while the plans point to some work already underway at the time of the memo’s release to achieve its goals, these are fundamentally reactionary documents issued in response to an executive mandate. The 19 responses read more like “plans to plan” than an actual roadmap to meeting the memo’s directives because without clear funding mechanisms, additional infrastructure development will be incremental at best. Some report on activities already underway but others read more as plans to develop further plans that will support OSTP’s goals as articulated in the memo. While some agencies, particularly those with a longer history of conducting or supporting science, directly addressed many topics relevant to the goals of the Holdren memo, other responses suggest that these agencies are still in the beginning stages of planning when it comes to increasing access to research data.

One concrete policy change that many agencies indicated plans to enact is an expansion of requirements around Data Management Plans. Agencies where DMPs were not formerly required will now expect them with new funding applications, and existing requirements will be revised. For example, the FDA’s discussion of DMPs explains, “[the] FDA will create or modify agency policies to require that data management plans be developed and ... followed by both intramural and extramural researchers”. (Food and Drug Administration, 2015, p. 9). Discussions such as these occur throughout the responses. While these are widely considered to be a positive step forward and have opened up the possibility of gaining deeper understanding of researcher behavior (Rolando et al., 2015), they have also been identified as inconsistent and not necessarily representative of any widespread change in researcher behavior around data (Bishoff and Johnston, 2015; Parham et al., 2016; Parham and Doty, 2012). While the impact of DMPs remains uncertain, they still function as a valuable tool to raise awareness about data management issues and will continue to evolve, reflecting changing understandings across the research community.

Close readings of the agency responses to the Holdren memo indicate that many of the details describing how agencies plan to provide access to research data require further development. Additionally, the question of funding the activities outlined in the plans is one of the lowest scoring themes from our analysis. Some agencies did not mention funding the work to make data publicly accessible at all, while others such as NIST indicated that they would use existing appropriations to implement their plans. NIST’s response claims that the institute “has identified base funding resources within its Scientific and Technical Research and Services Appropriations in order to ensure initial implementation, scale-up, and continued operation of the NIST system to make publications and data publicly available” (National Institute of Standards and Technology p.16). While it underscores the implicit idea that the vision introduced in the Holdren memo is an unfunded mandate for federal funding agencies, the response from NIST reflects the original memo, which specified that activities to increase access to federally-funded scientific research should utilize “resources within the existing agency budget” (Holdren memo p. 2).

What do these agency responses tell us about the state of data curation and access in the federal government? While the increased focus of these agencies on data curation issues is encouraging, the lack of specifics around data management and preservation practice suggests that much more work is necessary to fulfill the mandate of the 2013 Holdren memo. At this stage, compliance with the memo and meaningful action towards increasing access to the results of federally-funded research are not the same. These agencies complied with the memo by responding to its eight prompts, and now have the opportunity to flesh out the details. However, the lack of detail in the compliant plans raises the question of what compliance

means for the Holdren memo. The July 2016 update provided to Congress by John Holdren states that “departments and agencies with approved plans are making steady progress in policy implementation” (Holdren, 2016a, p. 1). This update goes on to report developments in access to research results across the federal government, stating that “Several departments and agencies are exploring the development of new data repositories and data catalogs to improve access to data associated with research publications and make such data easier to find” (Holdren, 2016a, p. 2). The most recent update, from October 2016, includes additional anecdotes from the agencies on progress around data management, preservation, and access, such as the number of Data Management Plans received from Veterans Affairs grant applications and the release of data management software by USGS (Holdren, 2016b, p. 2). Despite a lack of systemic metrics for evaluating success in efforts to increase access to federally-funded scientific research, the OSTP notes this progress as encouraging and concludes the update by reiterating its commitment to helping agencies work towards the goals of the Holdren memo.

The Obama administration’s emphasis on increasing access to federally-funded scientific research results is an opportunity for the data curation community to contribute its expertise to an emerging policy priority for the federal government. Much of the research on data curation has focused on individual scientists, research communities, repositories, and institutional (university) efforts. The activity across the federal government following the 2013 Holdren memo necessitates an increased focus on federal scientific data curation in the research community. As a key funder of scientific research, the actions of government agencies around management and curation of scientific data require further study.

The focus of this study was on the United States, but issues around public policy and government-supported scientific research are increasingly relevant internationally. Public sector data curation initiatives such as the European Commission’s Horizon 2020 Framework (“Horizon 2020 – European Commission”, 2016), Canada’s Action Plan on Open Government 2014–2016 (Canada and Treasury Board, 2014), and the new data policies of UK government research funding councils (“Funders’ data policies|Digital Curation Centre”, 2016) all contain their own directives to researchers in their constituent countries with respect to legal mandates and best practices for management and curation of research data. We hope that this project leads to more work examining how agencies and researchers are shaping the future of research data management through policy interventions.

Conclusion

In this study, we compared the open access plans of federal scientific agencies and presented analysis about the different steps proposed by these agencies to provide greater access to the results of federally-funded scientific research, specifically the data generated during the course of this research. The issues raised in the Holdren memo, and outlined in the responses, remain relevant for the data curation community as it continued to seek new ways to provide and broaden access to research data.

Our results demonstrated that the 19 agencies that responded to the memo broadly complied with its directives. Overall, we found that most of the themes coded in our analysis were addressed in some way by responding agencies. Those agencies with a history of and reputation for supporting scientific research scored higher than peer agencies with less experience managing research and development activities. We found that most responding agencies agreed that requiring research proposals from intramural and extramural scientists to include data management plans is an important component of their plans to comply with the memo. We reported on emergent themes outside of the scope of the Holdren memo, such as efforts to digitize historical data, because their inclusion in a few responses suggests that agencies understand the complexities involved in fulfilling the spirit of the memo. Ultimately, these 19 responses leave questions unanswered about agency plans, as implementation details were not included in every response.

While this study provided a comprehensive and broad analysis of agency responses to the 2013 Holdren memo, this issue continues to evolve as agencies refine and implement elements of their initial responses. Some will build their own data management infrastructures while others are likely to partner with each other or with third party organizations in their efforts to provide greater access to research data. Future research should continue to examine how the Holdren memo shapes federal agency policies around data management on federally-funded research projects. Will the commitment to open data and broadening access to publicly-funded research continue beyond the Obama administration and the tenure of John Holdren as Director of the Office of Science and Technology Policy? As of this writing, his successor has yet to be nominated and the future direction and priorities of the OSTP are undefined. Nevertheless, the Holdren memo prompted agencies to consider how they might broaden access to research results, efforts which stand to benefit scholarly communities, the government, and society as a whole.

Additional Files

The additional files for this article can be found as follows:

- **Appendix 1.** List of URLs for OSTP Plan Responses. DOI: <https://doi.org/10.5334/dsj-2017-027.s1>
- **Appendix 2.** List of Codes and Definitions Used in Analysis. DOI: <https://doi.org/10.5334/dsj-2017-027.s2>

Acknowledgements

The authors wish to acknowledge the USDA National Agricultural Library for supporting this work through a Cooperative Agreement.

Competing Interests

AK, KH, and RP have received funding from the USDA National Agricultural Library through a cooperative agreement. CP is an employee at the USDA National Agricultural Library.

References

- Barata, K J** 1996 Managing Intellectual Assets: The Identification, Capture, Maintenance, and Use of the Records of Federally Sponsored Scientific Research. *Archival Issues*, 21: 129–143.
- Berman, F** and **Cerf, V** 2013 Who Will Pay for Public Access to Research Data? *Science*, 341: 616–617. DOI: <https://doi.org/10.1126/science.1241625>
- Bishoff, C** and **Johnston, L** 2015 Approaches to Data Sharing: An Analysis of NSF Data Management Plans from a Large Research University. *Journal of Librarianship and Scholarly Communication*, 3. DOI: <https://doi.org/10.7710/2162-3309.1231>
- Canada and Treasury Board** 2014 Canada's Action Plan on Open Government 2014–16.
- CENDI** 2016 Implementation of Public Access Programs in Federal Agencies [WWW Document]. URL: http://cendi.gov/projects/Public_Access_Plans_US_Fed_Agencies.html (accessed 6.14.16).
- Corneliusson, S T** 2016 New York Times commentary argues that all research papers should be free. *Physics Today*. DOI: <https://doi.org/10.1063/PT.5.8169>
- English, R** and **Raphael, M** 2006 The next big library legislative issue. *American Libraries*.
- Fleming, J R** 1989 Meteorology at the Smithsonian Institution, 1847–1874: the natural history connection*. *Archives of Natural History*, 16: 275–284. DOI: <https://doi.org/10.3366/anh.1989.16.3.275>
- Food and Drug Administration** 2015 Plan to Increase Access to Results of FDA-Funded Scientific Research.
- Franceschi-Bicchierai, L** 2013 White House Directive Expands Access to Scientific Research. www.Mashable.com.
- Frankel, M S** 1999 Public access to data. *Science*, 283: 1114–1114. DOI: <https://doi.org/10.1126/science.283.5405.1114>
- Funders' data policies Digital Curation Centre [WWW Document]** 2016 URL: <http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies> (accessed 11.10.16).
- Harden, V A** 2009 A Short History of the NIH [WWW Document]. URL: <https://history.nih.gov/exhibits/history/index.html>.
- Harvard Open Access Project HOAP** 2016 Notes on the Fair Access to Science and Technology Research Act [WWW Document]. <http://cyber.law.harvard.edu>. URL: http://cyber.law.harvard.edu/hoap/Notes_on_the_Fair_Access_to_Science_and_Technology_Research_Act.
- Heafey, E** 2011 Public Access to Science: The New Policy of the National Institutes of Health in Light of Copyright Protections in National and International Law. *UCLA Journal of Law and Technology*, 14. Retrieved from http://www.lawtechjournal.com/articles/2010/02_100216_heafey.pdf.
- Holdren, J P** 2013 Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research.
- Holdren, J P** 2016a Public Access- Report to Congress- July 2016. Office of Science and Technology Policy.
- Holdren, J P** 2016b Public Access- Report to Congress- October 2016. Office of Science and Technology Policy.
- Horizon** 2020 – European Commission [WWW Document], 2016. Horizon 2020. URL: <https://ec.europa.eu/programmes/horizon2020/> (accessed 11.10.16).
- Howard, J** 2013 White House Delivers New Open-Access Policy That Has Activists Cheering. *The Chronicle of Higher Education*.
- Kaiser, J** 2006 Bill Would Require Free Public Access to Research Papers. *Science*, 312: 828a–828a. DOI: <https://doi.org/10.1126/science.312.5775.828a>

- Killian, J R** 1977 Sputnik, scientists, and Eisenhower: A memoir of the first special assistant to the president for science and technology. MIT Press (MA).
- Lurie, N, Manolio, T, Patterson, A P, Collins, F and Frieden, T** 2013 Research as a Part of Public Health Emergency Response. *New England Journal of Medicine*, 368: 1251–1255. DOI: <https://doi.org/10.1056/NEJMs1209510>
- Michelini, A, De Simoni, B, Amato, A and Boschi, E** 2005 Collecting, digitizing, and distributing historical seismological data. *Eos, Transactions American Geophysical Union*, 86: 261. DOI: <https://doi.org/10.1029/2005EO280002>
- Miles, M B and Huberman, A M** 1994 Qualitative data analysis: an expanded sourcebook. Sage Publications, Thousand Oaks.
- Murphy, K** 2016 Should All Research Papers Be Free? *The New York Times*.
- NASA** 2014 NASA Plan: Increasing Access to the Results of Scientific Research. National Aeronautics and Space Administration.
- National Institute of Standards and Technology** 2015 Plan for Providing Public Access to the Results of Federally Funded Research, NIST.
- National Institutes of Health** 2008 Analysis of Comments and Implementation of the NIH Public Access Policy. NIH.
- National Science Foundation** 2016 NSF At a Glance [WWW Document]. URL: <https://www.nsf.gov/about/glance.jsp>
- Nickum, L S** 2006 Elusive No Longer? Increasing Accessibility to the Federally Funded Technical Report Literature. *The Reference Librarian*, 45: 33–51. DOI: https://doi.org/10.1300/J120v45n94_04
- NOAA Research Council** 2015 NOAA Plan for Increasing Public Access to Research Results. National Oceanic and Atmospheric Administration, Silver Spring, MD.
- Office of Science and Technology Policy** 2016 About OSTP [WWW Document]. URL: <https://www.whitehouse.gov/administration/eop/ostp/about>.
- Office of the Assistant Secretary for Preparedness and Response** 2014 Pandemic and All Hazards Preparedness Act [WWW Document]. URL: <http://www.phe.gov/preparedness/legal/pahpa/pages/default.aspx>.
- Parham, S W, Carlson, J, Hswe, P, Westra, B and Whittier, A** 2016 Using data management plans to explore variability in research data management practices across domains. *International Journal of Digital Curation*, 11. DOI: <https://doi.org/10.2218/ijdc.v11i1.423>
- Parham, S W and Doty, C** 2012 NSF DMP Content Analysis: What Are Researchers Saying? [WWW Document]. URL: <https://smartech.gatech.edu/handle/1853/48707> (accessed 2.25.16).
- Peek, R** 2009 A Potential OA Policy for U.S. Agencies. *Information Today* 26.
- Rolando, L, Carlson, J, Hswe, P, Parham, S W, Westra, B and Whitmire, A L** 2015 Data Management Plans as a Research Tool. *Bul. Am. Soc. Info. Sci. Tech.*, 41: 43–45. DOI: <https://doi.org/10.1002/bult.2015.1720410510>
- Scholarly Publishing and Academic Resources Coalition** 2016 Browse Data Sharing Requirements by Federal Agency [WWW Document]. SPARC. URL: <http://datasharing.sparcopen.org/> (accessed 5.2.16).
- Shelby, R** 2000 Accountability and Transparency: Public Access to Federally Funded Research Data Policy Essay. *Harvard Journal on Legislation*, 37: 369–90.
- U.S. Department of Transportation** 2015 U.S. Department of Transportation Public Access Plan: Increasing Access to Federally Funded Research Results [WWW Document]. URL: <https://www.transportation.gov/sites/dot.gov/files/docs/Official%20DOT%20Public%20Access%20Plan%20over%201.1.pdf> (accessed 11.10.16).
- U.S. Geological Survey** 2016 Public Access to Results of Federally Funded Research at the U.S. Geological Survey: Scholarly Publications and Digital Data. U.S. Geological Survey.
- Van Noorden, R** 2013 US science to be open to all. *Nature*, 494: 414–415. DOI: <https://doi.org/10.1038/494414a>
- Van Tuyl, S and Whitmire, A L** 2016 Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia. *PLOS ONE*, 11: e0147942. DOI: <https://doi.org/10.1371/journal.pone.0147942>
- Wellcome Trust** 2016 Open Access Publishing [WWW Document]. URL: <https://wellcome.ac.uk/what-we-do/our-work/our-policy-work-open-access>.
- Whitmire, A, Briney, K, Nurnberger, A, Henderson, M, Atwood, T, Janz, M, Kozlowski, W, Lake, S, Vandegrift, M and Zilinski, L** 2015 A table summarizing the Federal public access policies resulting from the US Office of Science and Technology Policy memorandum of February 2013. DOI: <https://doi.org/10.6084/m9.figshare.1372041>

How to cite this article: Kriesberg, A, Huller, K, Punzalan, R and Parr, C 2017 An Analysis of Federal Policy on Public Access to Scientific Research Data. *Data Science Journal*, 16: 27, pp. 1–13, DOI: <https://doi.org/10.5334/dsj-2017-027>

Submitted: 09 January 2017 **Accepted:** 11 May 2017 **Published:** 14 June 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 