IJDC | *Research Paper*

Analysis of U.S. Federal Funding Agency Data Sharing Policies: 2020 Highlights and Key Observations

Reid I. Boehm
Purdue University

Patricia B. Condon
University of New Hampshire

Hannah Calkins
Sage Bionetworks

Jonathan Petters
Virginia Tech

Rachel Woodbrook
University of Michigan

Abstract

Federal funding agencies in the United States (U.S.) continue to work towards implementing their plans to increase public access to funded research and comply with the 2013 Office of Science and Technology memo *Increasing Access to the Results of Federally Funded Scientific Research*. In this article we report on an analysis of research data sharing policy documents from 17 U.S. federal funding agencies as of February 2021. Our analysis is guided by two questions: 1.) What do the findings suggest about the current state of and trends in U.S. federal funding agency data sharing requirements? 2.) In what ways are universities, institutions, associations, and researchers affected by and responding to these policies? Over the past five years, policy updates were common among these agencies and several themes have been thoroughly developed in that time; however, uncertainty remains around how funded researchers are expected to satisfy these policy requirements.

Submitted 9 December 2021 ~ *Accepted* 9 August 2022

Correspondence should be addressed to Reid I. Boehm, riboehm@purdue.edu

Boehm and Condon provided equal authorship contributions.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

In the United States, the National Institutes of Health (NIH) and the National Science Foundation (NSF) implemented data sharing policies in 2003 and 2011 respectively, requiring grant applicants to submit details of how they plan to manage and share data originating from federally funded research (National Institutes of Health, 2003; National Science Foundation, 2010). Motivations for the development of such policies generally acknowledge "maximising the impact and reach of research" and "concerns around citizen access and engagement with research" (Neylon, 2017a). In 2013, the U.S. Office of Science and Technology (OSTP) published a memo, *Increasing Access to the Results of Federally Funded Scientific Research*, requiring "federal agencies with more than $100M in R&D expenditures to develop plans to make the results of federally funded research freely available to the public" (OSTP, 2013). This call to action resulted in many U.S. federal funding agencies publishing plans and policies outlining expectations for data management planning, data sharing, and timely public access to research outputs from funded projects including data and associated metadata.

Navigating and understanding the various data sharing plans and policies from funding agencies, however, proved to be challenging for researchers and data management professionals. In a 2016 joint project, the Scholarly Publishing and Academic Resources Coalition (SPARC) and Johns Hopkins University Libraries developed the Data Sharing Policy resource (SPARC, 2021), a tool for understanding and comparing elements of the United States federal funding agencies' research data sharing policies for extramural research. Since the tool's release, funding agencies have continued to publish new or updated policies, plans, and guidance in response to the 2013 OSTP memo. The tool needed to be updated to keep this community resource current and valuable. In 2020, a group of research data management professionals working under the auspices of the Research Data Access and Preservation Association (RDAP) updated the SPARC Data Sharing Policy resource to reflect the then-current policies of these funding agencies.

In this paper we report on key observations from the project's updated dataset (Calkins, et al., 2021) and discuss what these observations suggest about the current state of U.S. federal data sharing requirements. Our analysis is guided by two questions:

1. What does the data sharing policy resource update suggest about the current state of and trends in U.S. federal data sharing requirements?

2. In what ways are universities, institutions, associations, and researchers affected by and responding to these policies?

The authors of this paper are the contributors to the 2020 update to the SPARC Data Sharing Policy resource and include the lead from the original 2016 project to create the resource. We use our experience working with researchers in their data management; literature on data management, sharing, and curation; our deep reading of the policy and guidance documents during the resource update; and the tabulation of key observations from that dataset to discuss gaps and trends in U.S. funding agency data sharing policy development.

Literature Review

There are few studies that provide content analysis of U.S. federal funding data sharing policies. One of these, Kriesberg et al. (2017), conducted a comprehensive analysis of the public access plans of 19 U.S. federal funding agencies. Using 15 themes pulled from the 2013 OSTP memo, the researchers scored the plans as to whether the theme was not discussed, briefly discussed, or

discussed in detail. Some themes such as attribution, data management planning, and preservation overlapped with the analysis presented here, while others such as use of digital object identifiers and training did not. The researchers found that most of the themes were discussed at least superficially in the agencies' responses to the memo. However, the plans "read more like 'plans to plan' than an actual roadmap to meeting the memo's directives." The researchers further note, "compliance with the memo and meaningful action towards increasing access to the results of federally-funded research are not the same," and that progress will be slow without infrastructure and resources to implement the plans and support researchers.

Two other less comprehensive studies were also conducted. Williams, Bagwell, and Zozus (2017) conducted an analysis of data management and data sharing plan requirements by U.S. funders. The researchers found that more funders were requiring data sharing plans, emphasizing post-publication sharing of data rather than active management of data throughout the research project. Additionally, the researchers noted a lack of consistency in required elements of a data management plan. Finally, Gaba et al. (2020) conducted a survey of data sharing policies from non-commercial and commercial funders of randomized controlled trials. From their sample, less than half of the funders had data sharing policies in place, and for those that did, the policies lacked consistency and details on key components.

While these federal, private, and international data sharing policies are works in progress, their implementation still impacts the behavior of researchers. Several studies suggest that funding agency mandates for data sharing and data management planning have improved research data sharing practices and increased the likelihood of researchers sharing data (Piwowar & Chapman, 2010; Kim, Schuler, & Pechenina, 2018; Wallis, Rolando, & Borgman, 2013). However, developing the infrastructure to support data sharing and the cultural change required to widely adopt responsible data sharing practices is not without challenges. The ability of funding agencies to manage these challenges as well as track research outputs has been underwhelming (Couture, Blake, MacDoland, & Ward, 2018; Neylon, 2017a; Neylon, 2017b). This disconnect between funder mandates, infrastructure and resources, and researcher follow-through is described as the "data stewardship gap" (Gutmann, York, & Berman, 2019; York, Gutmann, & Berman, 2018).

Understanding the data stewardship gap is relevant for researchers, information and data management professionals, funding agencies, and other stakeholders. Data stewardship is defined as the "activities to maintain the integrity of and preserve access to data" (York, Gutmann, & Berman, 2018) and encompasses a range of activities taken by the researchers and by information and IT professionals. The data stewardship gap is exposed in several studies that focus on researcher behaviors and data availability (e.g., Tenopir, et al., 2011; Tenopir, et al., 2020; Zuiderwijk, Shinde, & Jeng, 2020). For example, one recent study of 315 publications from studies funded by the Exxon Valdez Oil Spill Trustee Council (EVOSTC), who required data be made publicly accessible, found that only 26% of the data underlying those studies was recoverable. The researchers suggest that "To mitigate data losses, funders and publishers should provide more support and more stringent requirements for data sharing" (Couture, Blake, McDonald & Ward, 2018).

As part of the Alfred P. Sloan-funded Stewardship Gap Project,[1] researchers conducted a comprehensive literature review to investigate the gap in data stewardship policies and practices that drives the incongruence between the amount of data produced by grant-funded research and the amount that is made publicly accessible, and how we measure that disparity. The researchers found six gap areas: culture, knowledge, responsibility, commitment, resources, and stewardship actions. They concluded that "there is not a single stewardship gap, but rather numerous and diverse components that contribute to and influence whether research data are responsibly stewarded" (York, Gutmann, & Berman, 2018). These gap areas are explored in studies about researcher data sharing behaviors and attitudes that highlight researcher motivations, responsible data sharing practices, and barriers to sharing such as lack of incentives, insufficient infrastructure and resources, and anxiety about ownership and reuse of

[1] The Stewardship Gap Project: https://cupc.colorado.edu/stewardship_gap/

data (Chawinga & Zinn, 2019; Couture, Blake, McDonald & Ward, 2018; Tenopir, et al., 2020; Zuiderwijk, Shinde, & Jeng, 2020).

This paper contributes to the research that explores the content of data sharing policies and the research that probes the data stewardship gap. Through an analysis of the updated dataset from the SPARC Data Sharing Policy resource project, we identify key observations about U.S. federal grant-making agency data sharing policies, and highlight challenges that influence the stewardship gap and opportunities for addressing it, particularly as they pertain to researchers complying with federal mandates. We present an overview of agencies' progress as reflected in their policies, with the aim of providing insight into how these policies are understood by, and impact, the broader research and data professional communities.

Approach

For the analysis of data sharing policies presented here, we use data that was collected for the primary purpose of updating the SPARC Data Sharing Policy resource. For the update, we gathered data about 17 U.S. federal funding agencies' data sharing policies for extramural research related to 13 thematic areas of policy focus, or facets. To maintain consistency with the original data sharing resource developed in 2016, we used the same 13 facets; we did not add, remove, or modify these during the update. Table 1 provides a description of the 13 facets.

Table 1. Descriptions of facets used to code funder data sharing policies.

Facet	Instructions for Data Entry - Focusing on policies affecting extramural researchers, not those only affecting intramural researchers
Principles and Implementation Approach	Explicitly stated purpose and/or foundational principles that guided the policy's development.
Exclusions and Limitations in Policy Scope	Any specific kinds of data to be included and excluded in data sharing or management policy coverage. While the federal government has a general definition for research data (U.S. Office of Management and Budget, 2013), a specific agency's policy may differ in what research data are explicitly covered.
When Data are to be Publicly Shared	Any requirements for when research data are to be made more widely available, and any circumstances under which a researcher might delay making their data more widely available.
How Data are to be Publicly Shared	Any mechanisms for how data can be shared, including recommended data repositories or required data registries.

Facet	Instructions for Data Entry - Focusing on policies affecting extramural researchers, not those only affecting intramural researchers
Metadata and/or Documentation to be Included with Data	Recommendations or requirements for metadata and/or documentation to be included with shared research data. These recommendations and requirements could include specifics on content and descriptive metadata, specified formats or schema for this metadata, or they could be broad or general.
Recommendations or Requirements for Terms of Data 'Licensing'	Requirements or recommendations the agency has for terms of use for research data generated through its funding (e.g., their re-use, re-distribution, and subsequent production of derivatives; commercial use of the data; or researcher credit and attribution using data citations).
Data Management Planning	If the agency requires a data management plan, provide important properties of the DMP requirement.
Data Preservation	Description of long-term preservation plan/approach, if any.
Current or Planned Partnerships in Policy Enactment	Identification of partnerships with external entities (e.g., other federal agencies, non-profits, publishers, academic institutions) in enactment of this policy.
Compliance Measures for Policy Enforcement	Compliance measures for enforcement of this digital data policy.
Guidance and Support to Researchers in Following Policy	Guidance provided by the agency to assist researchers in following the policy (e.g., FAQs, DMP templates, other guides for reviewers, proposal writers).
Guidance on Data Sharing/Management for Researchers using Multiple Funding Sources	Any comments on how researchers should meet agency requirements when one or more agencies funded the research.

To update the resource, we identified 17 U.S. federal grant-making agencies that were required to respond to the 2013 OSTP memo with public access plans and had published policies addressing public access to funded data (see Table 2). The Department of Homeland Security released a public access plan but has yet to release a policy, so we did not include the agency in the final update and analysis. The new NIH data management and sharing policy was released during the course of the project. Since the SPARC resource is aimed at helping researchers understand currently applicable policy, and the new NIH policy is not effective until January 2023, the updated content from that policy is not represented in the dataset.

Agencies with new or updated policies were moved on to next steps of data collection, while agencies without new policy (NIH, FDA, ASPR) were marked as “no update” and existing facet data was left as-is. For agencies with evidence of new policy, we used a two-person approach for

analysis and review. The primary reviewer identified all relevant new or updated policies and guidance from their assigned agencies, coded those documents using the 13 facets, and transferred relevant coded passages from the documents to a spreadsheet for further analysis. Passages from new policy and guidance documents were compared to originally cited documents, and new language was crafted for the SPARC resource that included the most recent and complete guidance available. The secondary reviewer then verified each of these steps; if the secondary reviewer recommended a modification, the two reviewers collaborated to reconcile changes. A final draft of the updated resource was provided to SPARC for publication on their website. The dataset and process documentation are published on the Open Science Framework (Calkins, et al., 2021).

Table 2. U.S. federal funding agencies included in the SPARC Data Sharing Policy Resource.

U.S. Federal Grant-making Agencies included in SPARC Data Sharing Policy Resource
Agency for Healthcare Research and Quality (AHRQ)
Assistant Secretary for Preparedness and Response (ASPR)
Centers for Disease Control (CDC)
Department of Agriculture (USDA)
Department of Defense (DOD)
Department of Education - Institute of Education Sciences (IES)
Department of Energy (DoE)
Department of Transportation (DOT)
Environmental Protection Agency (EPA)
Food and Drug Administration (FDA)
National Aeronautics and Space Administration (NASA)
National Institute of Standards and Technology (NIST)
National Institutes of Health (NIH)
National Oceanic and Atmospheric Administration (NOAA)
National Science Foundation (NSF)
US Agency for International Development (USAID)
US Geological Survey (USGS)

Findings

For each of the 17 agencies included in the dataset for the updated SPARC resource, information was collected about 13 thematic areas or facets – totalling 221 facets across all agencies. Figure 1 provides a breakdown of how many facets had no updates, partial updates (some language from the 2016 resource remains), full updates (no language from the 2016

resource remains), or no documents found. 64% of the facets either had no updates since 2016 or no guidance/policy was located by the team.

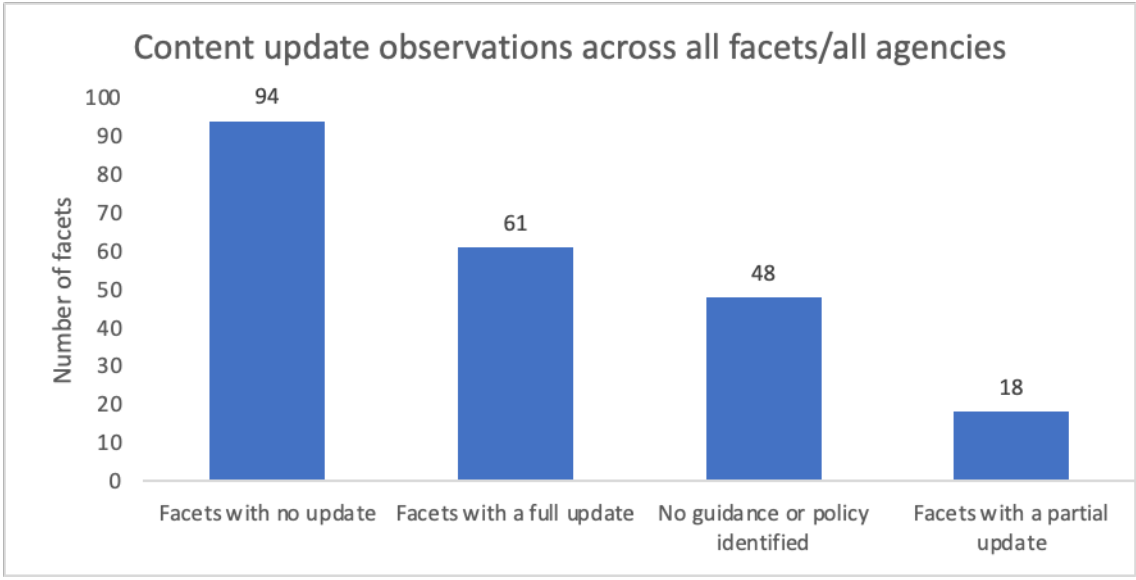

Figure 1: Total number of facets with content updates across all agencies

Table 3 shows the number of agencies that had new or updated policy or guidance documentation for each of the 13 facets since 2016.

Table 3. Number of agencies out of 17 with updates for each of the 13 facets

Facet	Agencies with Full Update	Agencies with Partial Update	Agencies with No Update	Agencies with no guidance available
Exclusions and Limitations in Policy Scope	4	4	9	0
Principles and Implementation Approach	4	2	11	0
Data Management Planning	7	2	7	1
Guidance and Support to Researchers in Following Policy	5	1	9	2
How Data are to be Publicly Shared	6	1	7	3

Facet	Agencies with Full Update	Agencies with Partial Update	Agencies with No Update	Agencies with no guidance available
When Data are to be Publicly Shared	4	2	8	3
Data Preservation	7	1	6	3
Metadata and/or Documentation to be Included with Data	4	3	7	3
Compliance Measures for Policy Enforcement	5	1	7	4
Data Citation and Attribution	4	0	7	6
Recommendations or Requirements for Terms of Data 'Licensing'	6	0	4	7
Current or Planned Partnerships in Policy Enactment	1	1	7	8
Guidance on Data Sharing/Management for Researchers using Multiple Funding Sources	4	0	5	8

We identified key observations about the facets, including the facets most often addressed across agencies, the facets in which most new content was added, and the facets in which little to no new content was added between 2016 and 2020. Table 4 summarizes these findings. The content shown in Tables 3 and 4 are referenced throughout the Discussion section.

Table 4. Key observations across 13 facets.

Category	Facets
Facets with the highest policy coverage	Principles and implementation approach Exclusions and limitations in policy scope Data management planning
Facets with the lowest policy	Current and planned partnerships in policy enactment

Category	Facets
coverage	Guidance on data management for researchers using multiple funding sources Recommendations / requirements for terms of data licensing Data citation
Facets with the most new or updated content	Data preservation Data management planning Recommendations or requirements for terms of licensing How data are to be publicly shared
Facets with little to no new or updated content	Principles and implementation approach Guidance and support to following policy Exclusions and limitations in policy scope
Facets with the most updates/changes	Exclusions and limitations in policy and scope Metadata and or/documentation to be included with the data

Discussion

Observations from the dataset

As we look at the facets with greatest coverage, it is clear that the agencies understand and communicate the importance of the policy, with attention to what is and is not covered in the policy scope (facets: "principles and implementation approach" and "exclusions and limitations in policy scope"). These facets also have minimal updates from the 2016 content, suggesting that these elements are quite standard, static across time, and remain the theoretical underpinning of policy development linking them to the 2013 OSTP memo. Similarly, the facet of "data management planning" is addressed by most agencies and typically includes guidance documentation in addition to what is contained in the official policy document. Presumably, this level of coverage across agencies exists because the plan itself is the main vehicle for evaluating how projects and research practices will adhere to the data sharing policies. While we analyzed data management planning as its own facet, we observed that in some cases, the data management plan (DMP) acted as a catchall for facets that were not thoroughly addressed elsewhere in the policy (e.g., the DMP should define when the researcher expects to share data).

The facets with lowest coverage are those concerned with issues of partnership and collaboration across agencies (facets: "current or planned partnerships in policy enactment" and "guidance on data sharing/management for researchers using multiple funding sources"). Little attention to guidance on current and planned partnerships may suggest that agencies see their operations as stand-alone entities or that investment into building shared infrastructure has not yet been prioritized, and as such they are not taking full advantage of existing or potential partnerships that would allow for connection and alignment. Similarly, there is little guidance on data management for those using multiple funding sources. This brings up many questions and implications for researchers faced with this situation. For example, a researcher funded by the Department of Education (IES) for behavioral health research who is required to register their study with clinicaltrials.gov will be subject to two data sharing policies that have different, and at times contradictory, expectations and requirements (e.g. differences in acceptable choices for

data repositories); it is unclear to what extent the researcher should comply with each policy and how they should resolve any conflicts between the two. While agencies do work with one another (Heckman, 2021; NSF, 2020), and projects do have multiple federal funders in certain situations (e.g., BIRDS Lab, 2021), this information on partnerships is simply not reflected in the policy resources. The consistent absence of this information suggests it is yet to be recognized as important to communicate to researchers engaged in such projects.

Additional facets with low coverage are "recommendations or requirements for terms of data licensing" and "data citation and attribution." In the absence of this content, we see that researchers are left to navigate the question of open licensing, intellectual property rights, and credit for their work either individually or with support from their organizations, which leads to fragmented and inconsistent approaches to these topics throughout the community and across disciplines (Boté, & Térmens, 2019; Dorta-Gonzalez, Gonzalez-Betancor, Dorta-Gonzalez, 2021; Labastida & Margoni, 2020). For those agencies where there is coverage, a closer look at the content reveals a primary focus on the implications of licensing and liability for the agency and less focus on researchers' rights, concerns, or actions. This is unfortunate considering that a widespread fear and barrier to data sharing among researchers is getting "scooped", losing out on publishing important findings, and failing to receive proper attribution for their contributions (Stieglitz et al., 2020; Zuiderwijk, Shinde, & Jeng, 2020). Funding agencies should provide clearer guidance regarding data licensing, attribution, and credit to the researchers they fund.

The four facets that emerge with the most new content across the agencies reflect areas of increasing need for attention by both data stewards and researchers. The first, "data preservation," is clearly coming to the fore in policies as a matter of great importance for sustaining the record of federally funded research endeavors. While the digital/data preservation field is well-established, in practice preservation is a complex multi-leveled set of actions and is not always easy for researchers and institutions to define, facilitate, or fund (Barrueco & Termens, 2021; Johnston, 2020; Lindlar, Rudnik, Jones, & Horton, 2020). The addition of content related to the other three facets ("how data are to be publicly shared," "recommendations or requirements for terms of data licensing," and "data management planning") clearly demonstrates agency progress towards implementing policy from plans set in action by the 2013 OSTP memo. Each of these four facets also represent new hurdles for researchers unaccustomed to data sharing. Such challenges are necessarily accompanied by the need for more information and guidance, such as how to select an appropriate repository, what it means to archive data, and how to understand and enact levels of preservation. Unfortunately, as we see from our observations, while agencies are beginning to add content in these areas, the change across agency data sharing policies from the 2016 resource is not yet sufficient to effect large-scale change around these requirements.

Two facets arise as areas with the most full or partial changes from the 2016 content. These updates provide insight into shifts in the federal research community around public access to data and reproducibility. The observed changes in "exclusions and limitations in policy scope," the facet addressing ways that research data are defined and content excluded from that definition, suggests that agencies have a better understanding of how the plans they created to address the 2013 OSTP memo can be realistically implemented. This may reflect a greater attention to what is required to fully realize the reproducibility of research (Merz, et al., 2020). The updates/partial changes to "metadata and/or documentation to be included with data" follow this same improved understanding. With a wider definition of shared research materials, (e.g., code, software, and workflows) metadata elements that describe these materials will also need to be included to facilitate research transparency and reproducibility. Deeper exploration of the content in these facets may surface the nature and alignment of perspectives across agencies.

Addressing the guiding questions

What does the data sharing policy resource update say about the current state of and trends in federal data sharing requirements across agencies?

Fourteen of the U.S. federal grant-making agencies included in the SPARC resource update have published a newer version of their policy or guidance documents since 2016, when the original SPARC resource was created. This suggests that agencies continue to work towards elements addressed in their plans to comply with the 2013 OSTP memo and continue to hone guidance for researchers. While content updates were common and several facets explored were thoroughly developed in the policies, questions remain around how researchers are to comply with these policies. Additionally, the lack of harmonization of data management and sharing requirements across U.S. funding agencies makes for a more challenging environment for researchers to work within. This lack of consistency across agency policies has been an enduring challenge, with Joseph noting in a 2016 article about the U.S. data sharing policy landscape that "there is not yet a common set of standards for any of these policy components" (Joseph, H., 2016).

The agencies are clearly able to talk about the sentiment behind the scope and purpose of their policies, but when it comes to specific details about how researchers are to effectively enact the requirements defined for management and sharing, there are still a lot of gaps and a lack of clarity. In dissecting these gaps, the agencies' responses to the OSTP memo remain focused on inward-looking benefits to the agency. Currently, we can see that most of the agencies are still in the initial response phases and are not yet externally focused on what is feasible and sustainable across the research data landscape.

As we know from many aspects of federal government and other institutional oversight and policy establishment practices, change is slow, incremental, and not always congruent across the elements of the policy in a way that facilitates full understanding or benefits compliance (Neylon, 2017a). This is exemplified in these findings through the influx of new content in the data sharing policy facets of "data preservation," "recommendations or requirements for terms of data licensing," and "how data are to be publicly shared," in contrast with there being little to no change in the guidance for researchers. While agencies are responding to theoretical principles of transparency, integrity, and reproducibility -- perhaps even reflecting the prominent findable, accessible, interoperable, and reproducible or FAIR principles (Wilkinson, et al., 2016) -- implementation of theory into practice is yet to be fully addressed. This observation is reflected in the lack of practical direction within policies and evolving state of accompanying guidance documents from funders. Overall, the federal agencies appear to be actively engaged in refining their understanding and policies, but much is still in flux and researchers cannot be faulted for finding the landscape a confusing one to navigate globally.

In what ways are universities, institutions, associations, and researchers affected by and responding to these policies?

This disconnect between requirements and guidance can leave the researcher unclear about how to comply with mandates and what they can and cannot propose to do. For example, a federal agency may require the researcher to submit the data to a repository for public access and to preserve the data for an extended period. Researchers may not understand that many available data repositories are designed for open access but not preservation. Some researchers have multiple terabytes of data generated over the course of their research, and this scale of archiving and preservation requires a level of funding and skills they may not be able to easily access. Further, without awareness of the need to allocate funds up front, they may face challenges and expenses once they reach the preservation and sharing stages. Additionally, funding generally must be used within the grant period, but requirements for data archiving and long-term preservation extend beyond the funded period and beyond the life of the project.

Another example concerning the "recommendations or requirements for terms of data licensing" and "data citation and attribution" facets demonstrates the gap between researcher needs and agency priority. In the policy and guidance documents in which these topics are addressed, information is geared toward how the use of licenses and citations affects the agencies. For the most part, it does not take into account the very real effect this has on the researchers who are generating the content. A funding agency requiring a researcher to share data publicly, and requiring the researcher choose a license without communicating the implications or providing guidance on licensing and ownership, may put the researcher's livelihood or integrity in a precarious position if they are unable to find their own assistance and/or guidance on this front.

Universities and research institutions are responding in several ways to the changes and challenges that these data sharing policies bring to the preparation and execution of federally funded grants. One clear shift is in the academic library with the addition of positions or changing roles to accommodate research data management, data archiving, and data curation (Gowen & Meier, 2020). Further, many research centers and hospitals have professionals embedded in their institutions who work directly with researchers on aspects related to metadata, curation, archiving, sharing, and compliance. Libraries within academic institutions are increasingly building research data services, such as reviewing data management plans, providing data curation support, and implementing open access data repositories that comply with federal funding requirements for archiving and preservation of data. However, libraries face challenges growing data services programs because of lack of resources, staffing, and organizational support (Cox, Kennan, Lyon, & Pinfield, 2017).

Cross-institution initiatives and dialogues are growing that bring together expertise from libraries, IT, research compliance, and other research data stakeholders (for example, AAU/APLU, 2021). In many cases, training and workshops are offered to researchers at both a general level and with more granularity, such as when new policies are introduced. Institutions are increasingly establishing their own data management and sharing policies that reflect large scale compliance considerations and respond in kind to those at the federal level. As we have witnessed at our own institutions, money, time, and resources are increasingly devoted to expanding infrastructure and staffing in response to the needs of the researchers engaged in federally funded endeavors. However, meeting the needs of the researchers and the complex requirements of their data management work remains elusive and difficult to align with current support service capacity. Often an institution may be unaware or unable to track these gaps because they are focusing on other pressing situations such as securing equipment for labs and research misconduct cases. Over the last several years we have each experienced new dialogues with stakeholders and service providers at our institutions regarding how we can best clarify and respond to the expanding needs of research data generated on our campuses.

These examples are only a few of many that arise when aspirational ideas taken from the broader realm of data governance and stewardship are implemented without managing the real-world needs of the researcher and the practical capacity to execute such requirements. This, in turn, contributes to the data stewardship gap and highlights challenges faced by researchers across all six gap areas: culture, knowledge, responsibility, commitment, resources, and stewardship actions (York, Gutmann, & Berman, 2018).

Addressing the stewardship gap

One might ask, is it the role of the funding agency to provide this support and be the change-leaders? Herein lies a fundamental challenge of increasing public access to research data by relying largely on policy compliance. Funders, institutions, and researchers all play a critical role in fostering change and building the infrastructure needed to advance public access to funded research and implement more open science practices. Funders, however, have the reach and power to set norms across research agendas to broadly effect change.

Funders can make significant improvements in both the accessibility and clarity of policy they create; for example, researchers can benefit greatly from additional guidance such as FAQ

pages, example data management plans, lists of approved data repositories, and detailed information about how funds might be requested and utilized in support of data sharing and preservation. Further, in developing guidance and navigating the line between being too prescriptive and too vague, providing researchers and data management professionals increased opportunities to engage in dialogue with the agencies can improve community influence. While this currently happens in some cases, such as the NIH requests for feedback during their nascent policy release NIH, n.d.), this process could be extended, continued, and made more systematically accessible after the release. In many instances, issues may begin to arise only after a policy is active.

Another important element is the delivery of information about the policy and guidance. As we searched for policy and guidance for the SPARC Data Sharing Policy resource update, we ran into broken links, ambiguous context, lack of versioning, and overall difficulty clarifying the applicability of information. This made it difficult for us to determine which policies superseded each other and precisely which funding opportunities were affected by particular policies. Funders can and should provide clear, explicit guidance about the applicability of their policies and make sure such guidance and the most recent versions of all relevant policies are readily findable and accessible online.

At the institutional level, libraries and other specialized units (e.g., within research compliance and information security offices) often can and do play a major role in facilitating awareness and interpretation of policy, as well as providing resources for planning and adhering to policies. One source of potential support is comprehensive instruction and consultation services for data management planning, sharing, and reuse. Research data management and data curation specialists working from within a centralized unit, with strategic action, outreach, cross-campus connections, and support from administration are well positioned to raise awareness, advocate for researchers, and create pathways to existing services.

Associations such as SPARC, RDAP, the Research Data Alliance (RDA), and many others within the research data and data stewardship communities, play a major role as a space for connection, collaboration, and shared inquiry. There are initiatives in these associations where diverse working groups are advocating for specific needs through the development of guidance, documentation, and best practices. For example, the Association of Public and Land-grant Universities (APLU) and the Association of American Universities (AAU) conducted a series of summits on advancing public access to research data within academic research institutions. They convened stakeholders from university libraries, information technology, research offices, and funding agencies towards a guide to help institutions establish and grow structures for sharing research data (AAU/APLU, 2021). Another example, a Research Data Alliance (RDA) working group is developing a FAIR data maturity model, so that researchers and institutions can better measure whether their data meets an acceptable standard to be labeled as "FAIR" (RDA FAIR Data Maturity Model Working Group, 2020). Ongoing, dynamic, and long-term efforts in these realms will serve to build stronger connections between the researcher, the institution, and the federal funding agencies.

Limitations

There are limitations in our analysis. The main constraint is that the dataset we used was originally collected not to support a rigorous content analysis of policy documents but to update the SPARC Data Sharing Policy resource. If we had designed a study that looked at the U.S. federal funding agencies' policies, our methods and codes may have been different. While the group conducted and documented our update in a way that it could be repeated for future updates, the original purpose of the dataset was not for research.

Another limitation is a lack of the funding agency perspective in both the original resource update and the analysis. While we analyzed both official policy documents and associated guidance and communications, we did not have funding agency representation as part of the

analysis and so we are necessarily missing that perspective. However, this limitation is in some ways useful as it honors the real-world dynamic by which we are only privy to the concerns of the funders via the communications they release and do not always have avenues to understand or challenge the motivations and decisions behind their policies and guidance.

Conclusion

In this article we outlined key observations about U.S. federal funding agencies' data sharing policies as recorded in the dataset used to update the SPARC Data Sharing Policy resource. We found that 14 of the U.S. federal grant-making agencies included in the SPARC resource update had published a newer version of their policy or guidance documents since 2016. While several facets ("principles and implementation approach," "exclusions and limitations in policy scope," and "data management planning") have been thoroughly developed, some that address researcher recognition remain underdeveloped ("recommendations or requirements for terms of data licensing" and "data citation and attribution"). Progress continues on facets that address how researchers are expected to satisfy policy requirements (including "data preservation" and "how data are to be publicly shared"). We discussed what these observations suggest about the current state of federal data sharing requirements and ways in which universities, institutions, organizations, and researchers are responding to these mandates. Finally, from our experience as data support professionals we considered how various stakeholders within the research sphere act (or do not act) to support researchers in policy compliance and to fill in the gaps where policy may be lacking.

This analysis contributes to the conversation in the literature by providing a look at data sharing policies and how they have evolved with time, and by discussing the ways in which policy and guidance documentation affect researchers and institutions. It also considers deficits in policy and where and how those shortfalls fail to support the culture of data sharing that these policies are aimed at promoting. By adding our perspectives to the conversation around data sharing in this manner, we hope to prompt further discussion with funders and industry partners and ultimately encourage broad involvement by stakeholders in the development of data sharing policy and guidance. In this circumstance, the necessary infrastructure and cultural change to embed data sharing into research practice must follow policies that mandate it. Researchers, institutions, and funders all play a role in the successful implementation of responsible data sharing practices and helping to narrow the stewardship gap.

There are a variety of next steps which we feel would be helpful in continuing this conversation. As mentioned in the limitations section, we are writing from very specific perspectives as research support professionals. Perspectives shared by researchers, funders themselves, or publishers and other industry members would provide useful and constructive context. Additionally, we encourage funding agencies to establish forums or other mechanisms of engaging with researchers and industry professionals or expand the use of those already established. We recommend funders continue to refine policies in response to lessons learned via those avenues, continue to create and refine guidance to assist researchers in complying with policy requirements, and fund infrastructure to help narrow the data stewardship gaps.

Acknowledgements

The authors would like to thank SPARC and RDAP for sponsoring the update to the Data Sharing Policy resource which produced the dataset used in this analysis. In particular, the authors would like to acknowledge the hard work and support from Nick Shockey and Lese Fandel at SPARC who made the update possible.

References

Association of American Universities and Association of Public and Land-grant Universities (2021). Guide to Accelerate Public Access to Research Data. Washington, DC. https://www.aplu.org/library/guide-to-accelerate-access-to-public-data/file

Barrueco, J. M., & Termens, M. (2021). Digital preservation in institutional repositories: a systematic literature review. *Digital Library Perspectives*. doi.org/10.1108/DLP-02-2021-0011

BIRDS Lab, U. (2021). BigAnt v6 robot motion tracking data - RAW dataset [Data set], University of Michigan - Deep Blue Data. doi.org/10.7302/024q-kk06

Boté, J. J., & Térmens, M. (2019). Reusing Data: Technical and Ethical Challenges. *Journal of Library & Information Technology*, *39*(6). doi.org/10.14429/djlit.39.06.14807

[dataset] Calkins, H., Condon, P., Petters, J., Woodbrook, R., & Boehm, R. I. (2021, February 12). SPARC Data Sharing Resource Update 2020 [Data set]. Open Science Framework https://doi.org/10.17605/OSF.IO/C4PD8

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research*, *41*(2), 109-122. oi.org/10.1016/j.lisr.2019.04.004

Couture, J. L., Blake, R. E., McDonald, G., & Ward, C. L. (2018). A funder-imposed data publication requirement seldom inspired data sharing. *PLoS One*, 13(7), e0199789. doi.org/10.1371/journal.pone.0199789

Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, *68*(9), 2182-2200. doi.org/10.1002/asi.23781

Dorta-González, P., González-Betancor, S. M., & Dorta-González, M. I. (2021). To what extent is researchers' data-sharing motivated by formal mechanisms of recognition and credit? *Scientometrics*, *126*(3), 2209-2225. doi.org/10.1007/s11192-021-03869-3

Gaba, J. F., Siebert, M., Dupuy, A., Moher, D., & Naudet, F. (2020). Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PloS One*, *15*(8), e0237464. doi.org/10.1371/journal.pone.0237464

[journal article] Gowen, E. & Meier, J. J., (2020) "Research Data Management Services and Strategic Planning in Libraries Today: A Longitudinal Study", *Journal of Librarianship and Scholarly Communication* 8(1), eP2336. doi: https://doi.org/10.7710/2162-3309.2336

Gutmann, M.P., York, J., & Berman, F. (2019) *The Stewardship Gap: A Challenge in Long-Term Access to Data*. https://hdl.handle.net/2027.42/146759

Heckman, J. (April 7, 2021). HHS QSMO sees $6B more in grants handled through shared services this year. *Federal News Network.* Retrieved from https://federalnewsnetwork.com/agency-oversight/2021/04/hhs-qsmo-sees-6b-more-in-grants-handled-through-shared-solutions-this-year/

[journal article] Joseph, H. (2016). The evolving US policy environment for open research data. Information Services & Use, 36(1-2), 45-48. doi.org/10.3233/ISU-160806

Johnston, L. (2020, April). The Ever-Changing Work That is Digital Preservation. In *Archiving 2020 Final Program and Proceedings*. doi.org/10.2352/issn.2168-3204.2020.1.0.1

Kim, J., Schuler, E. R., & Pechenina, A. (2018). Predictors of data sharing and reuse behavior in academic communities. In D.G. Alemneh, J. Allen, & S. Hawamdeh (Eds.), *Knowledge Discovery and Data Design Innovation: Proceedings of the International Conference on Knowledge Management (ICKM 2017)* (pp. 1-25). doi.org/10.1142/9789813234482_0001

Kriesberg, A., Huller, K., Punzalan, R., & Parr, C. (2017). An Analysis of Federal Policy on Public Access to Scientific Research Data. *Data Science Journal*, *16*(0), 27. https://doi.org/10.5334/dsj-2017-027

Labastida, I, & Margoni, T. (2020). Licensing FAIR Data for Reuse. *Data Intelligence*; 2 (1-2), 199–207. doi: https://doi.org/10.1162/dint_a_00042

Lindlar, M., Rudnik, P., Jones, S., & Horton, L. (2020). "You say potato, I say potato" Mapping Digital Preservation and Research Data Management Concepts towards Collective Curation and Preservation Strategies. *International Journal of Digital Curation 15(1).* doi.org/10.2218/ijdc.v15i1.728

Merz, K.M., Amaro, R., Cournia, Z., Rarey,M., Soares, T., Tropsha, A., Wahab, H.A., & Wang, R. (2020). Editorial: Method and Data Sharing and Reproducibility of Scientific Results. *Journal of Chemical Information and Modeling*, *60* (12), 5868-5869. doi.org/10.1021/acs.jcim.0c01389

National Institutes of Health (NIH). (n.d.). *NIH Data Management and Sharing Activities Related to Public Access and Open Science*. NIH Office of Science Policy. Retrieved from https://osp.od.nih.gov/scientific-sharing/nih-data-management-and-sharing-activities-related-to-public-access-and-open-science/

National Institutes of Health (NIH). (February 26, 2003). Final NIH Statement on Sharing Research, Notice: NOT-OD-03-032DATA. Retrieved from https://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html

National Science Foundation (NSF). (May 10, 2010). Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans. News Release 10-077. Retrieved from https://www.nsf.gov/news/news_summ.jsp?cntn_id=116928

National Science Foundation (NSF). (December 3, 2020). *NSF Partnerships: Landscape Study*. https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf21201

Neylon, C. (2017a). Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers. *Research Ideas and Outcomes*, *3*, e21705. doi.org/10.3897/rio.3.e14673

Neylon, C. (2017b). Building a culture of data sharing: policy design and implementation for research data management in development research. *Research Ideas and Outcomes*, 3, e21773. https://doi.org/10.3897/rio.3.e21773

Piwowar, H. A., & Chapman, W. W. (2010). Public sharing of research datasets: a pilot study of associations. *Journal of informetrics*, *4*(2), 148-156. doi.org/10.1016/j.joi.2009.11.010

Research Data Alliance FAIR Data Maturity Model Working Group. (2020). *FAIR Data Maturity Model: specification and guidelines*. DOI: 10.15497/rda00050

Ross, J. S., Ritchie, J. D., Finn, E., Desai, N. R., Lehman, R. L., Krumholz, H. M., & Gross, C. P. (2016). Data sharing through an NIH central database repository: a cross-sectional survey of BioLINCC users. *BMJ open*, 6(9), e012769. dx.doi.org/10.1136/bmjopen-2016-012769

SPARC. (2021). Browse Data Sharing Requirements by Federal Agency. Retrieved from http://datasharing.sparcopen.org/data

Stieglitz, S., Wilms, K., Mirbabaie, M., Hofeditz, L., Brenger, B., López, A., & Rehwald, S. (2020). When are researchers willing to share their data?–Impacts of values and uncertainty on open data in academia. *Plos One*, *15*(7), e0234172. doi.org/10.1371/journal.pone.0234172

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, et al. (2011) Data Sharing by Scientists: Practices and Perceptions. *PLoS One* 6(6): e21101. doi:10.1371/journal.pone.0021101

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorg, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. PloS One, *15*(3), e0229003. doi.org/10.1371/journal.pone.0229003

U.S. Office of Management and Budget. (2013). *Federal Register: Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards.* Retrieved from https://www.federalregister.gov/documents/2013/12/26/2013-30465/uniform-administrative-requirements-cost-principles-and-audit-requirements-for-federal-awards#sectno-citation-%E2%80%89200.315

U.S. Office of Science and Technology (OSTP). (2013). *Increasing Access to the Results of Federally Funded Scientific Research.* Retrieved from https://obamawhitehouse.archives.gov/blog/2016/02/22/increasing-access-results-federally-funded-science

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS One*, *8*(7), e67332. doi.org/10.1371/journal.pone.0067332

Wilkinson, M., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Santos, L., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Williams, M., Bagwell, J., & Zozus, M. N. (2017). Data management plans: the missing perspective. *Journal of Biomedical Informatics*, *71*, 130-142. DOI: 10.1016/j.jbi.2017.05.004

York, J., Gutmann, M., & Berman, F. (2018). What do we know about the stewardship gap. *Data Science Journal*, 17. doi.org/10.5334/dsj-2018-019

Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PloS One,* 15(9), e0239283. doi.org/10.1371/journal.pone.0239283