



---

## Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements

Gail Steinhart, Eric Chen, Florio Arguillas, Dianne Dietrich, Stefan Kramer\*

Cornell University, Ithaca, NY, USA

### Abstract

**Objective:** Cornell University's Research Data Management Service Group (RDMSG) surveyed NSF principal investigators (PIs) at Cornell in order to understand how well-prepared researchers are to meet the new NSF data management planning requirement, to build our own understanding of the potential impact on campus services, and to identify service gaps.

**Methods:** We administered a 43-question online survey, which included questions about the respondents' research and research data, their interest in assistance with the creation of data management plans, and questions for each of the five general areas cited in the NSF's Grant Proposal Guide (2011) section on data management plans.

**Results and Discussion:** Respondents produce a wide variety of types and formats of data, al-

though most expect to share relatively small amounts of data. Respondents are generally uncertain as to whether the data they produce conforms to disciplinary standards. The majority create no metadata; of those that do, most do not create metadata according to a particular standard. Most researchers do not express a need for advice regarding intellectual property issues. Researchers report using a variety of strategies (on-campus and commercial) for backing up and for providing access to their data sets.

**Conclusions:** The overarching finding from our survey is that there is much uncertainty about what the new requirement means and how to meet it, and researchers welcome offers of assistance. To the extent that Cornell researchers are representative of NSF PIs, our findings reveal something about researchers' readiness to meet the new requirement, and their attitudes towards it.

---

### Introduction

In 2009, the report of the Interagency Working Group on Digital Data, a group including representatives of more than two dozen federal agencies, made clear the desire of U.S. government research funders to maximize the return on the research they fund by developing a strategic framework to promote preservation of and access to digital data. The report included the recommendation

that agencies promote data management planning, and offered specific suggestions for the kinds of information data management plans might include. This general sentiment has since been echoed in The Open Government Partnership's National Action Plan for the United States of America (2011). In May 2010, the U.S. National Science Foundation (NSF) issued a press release announcing its intention to require data management plans with all grant proposals (NSF

---

**Correspondence to** Gail Steinhart: gss1@cornell.edu

**Keywords:** Data management, data policies, data sharing, data management plans

2010a); and made the specifics of the requirement public in October, 2010 (NSF 2010b).

Cornell University has been building capacity and exploring opportunities in the area of data management and curation for several years. With the NSF's 2010 announcement that all grant proposals would require a data management plan (DMP), Cornell realized it was time to formalize collaborations among units across campus to more effectively support the data management needs of researchers. Under sponsorship of the Office of the Vice Provost for Research and the University Librarian, the Research Data Management Service Group (RDMSG; <http://data.research.cornell.edu/>) was formed. Consisting of multiple units at Cornell, including the Center for Advanced Computing (CAC; <http://www.cac.cornell.edu/>), the Cornell Institute for Social and Economic Research (CISER; <http://ciser.cornell.edu/>), Cornell Information Technology (CIT; <http://www.it.cornell.edu/>) and the Cornell University Library (CUL; <http://www.library.cornell.edu/>), the RDMSG is a virtual organization aimed at making Cornell's diverse and distributed data management resources more easily and seamlessly accessible to researchers (Block et al. 2010).

One of the first activities of the RDMSG was a survey of current NSF principal investigators (PIs) in order to understand how well-prepared researchers should meet the new NSF requirement, to build an understanding of the potential impact on campus services, and to identify service gaps. To the extent that Cornell researchers can be considered representative of NSF PIs, our findings reveal something about researchers' general readiness to meet the new requirement, and their attitudes towards it.

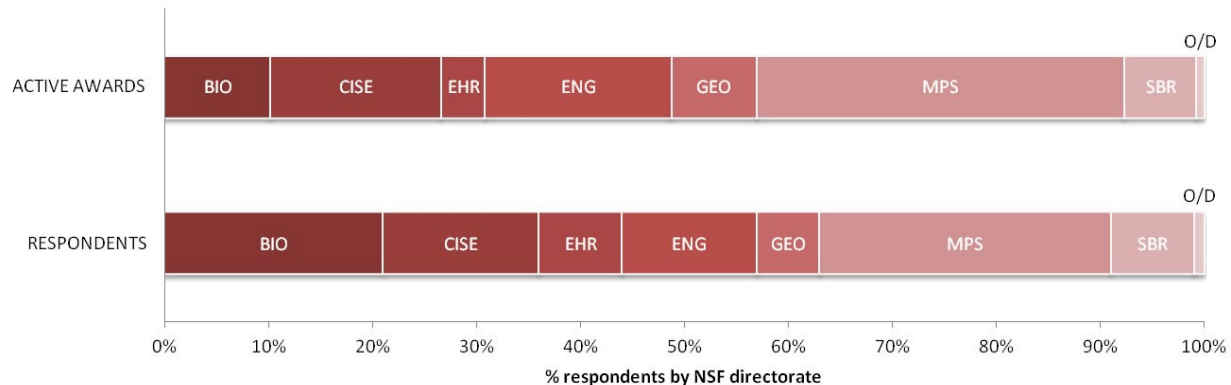
## Methods

We were interested in Cornell PIs' understanding of the concepts introduced in the

NSF data management plan requirement, how they anticipate approaching the requirement, what concerns they have with the new policy, the extent to which they are likely to utilize existing services, and what gaps in services might exist. We developed a 43-question online survey, which began with general questions about the respondents' research and research data (NSF directorate of their most recent award, general types of data produced), and interest in assistance with creating data management plans. We then developed a set of questions for each of the five general areas cited in the NSF's Grant Proposal Guide (2011) section on data management plans: types of data and other materials to be produced, standards used for data and metadata, policies and provisions for access, sharing, confidentiality, security, and intellectual property, policies and provisions for re-use and re-distribution, and plans for archiving and preserving access to data and other materials. Each of the sets of questions addressing NSF's five general DMP areas was introduced with a direct quote from the Grant Proposal Guide for that section. Sets of questions allowed for multiple-choice responses and an open-ended response, and to the greatest extent possible, used the terminology of the NSF Grant Proposal Guide. Where appropriate, we also included questions about specific campus services related to each of the NSF policy areas. Finally, we included questions designed to assess researchers' need for additional assistance or services, including data management planning. The authors tested the survey and requested feedback on its design from the RDMSG management group; the survey was not developed with input from researchers nor was it tested by any researchers.

We invited current and prospective NSF PIs from Cornell (researchers with funded proposals or proposals in review, approximately 1,650 individuals total, identified by Cornell's Office of Sponsored Programs) to participate in the survey. It was administered online using Cornell's installation of the Qualtrics

**Figure 1:** Distribution of current awards by directorate (“ACTIVE AWARDS”), and distribution of respondents by directorate to which they submitted their most recent proposal (“RESPONDENTS”). BIO=Directorate for Biological Sciences, CISE=Directorate for Computer & Information Science & Engineering, EHR=Directorate for Education & Human Resources, ENG=Directorate for Engineering, GEO=Directorate for Geosciences, MPS=Directorate for Mathematical & Physical Sciences, SBE=Directorate for Social, Behavioral & Economic Sciences, and O/D=Office of the Director (includes Office of Cyberinfrastructure, Polar Programs, and others).



(<http://www.qualtrics.com/>) web survey tool. Responses were accepted from December 20, 2010 through January 31, 2011. The complete survey instrument is available at <http://hdl.handle.net/1813/25624> (Steinhart et al. 2011).

## Results and Discussion

### Overview

We received 86 responses to the survey (excluding respondents with zero responses to questions), for a response rate of 5.2%. Responses were not required for every question. The average time spent taking the survey was 17.4 minutes, and only two respondents answered fewer than 20 of the 43 questions. Survey data, with identifying information and free-text responses removed, are available online at the following location: <http://hdl.handle.net/1813/25624> (Steinhart et al. 2011). We removed all free-text responses from the results because selective removal was potentially too subjective and carried the risk of identifying respondents to

readers knowledgeable of the respondent's discipline, or of the research environment at Cornell. We did, however, include in this paper a small number of quotes that, in our judgment, are not easily attributed to any particular individual.

Respondents were asked to reply with their most recent NSF proposal and the directorate to which they submitted it in mind. Overall, the distribution of respondents across NSF directorates was fairly representative of the distribution of current Cornell NSF awards (Figure 1); directorates with a higher survey response rate relative to active awards included Biology (BIO) and Education & Human Resources (EHR). Directorates with a somewhat lower response rate relative to active awards included Engineering (ENG) and Math & Physical Sciences (MPS).

The majority (62%, Table 1) of respondents indicated they were interested in help with writing a DMP; only 13% of respondents said they were not interested in guidance on writ-

**Table 1:** Responses to selected survey questions. Some questions have been edited for brevity. The survey instrument and results are available (Steinhart et al. 2011).

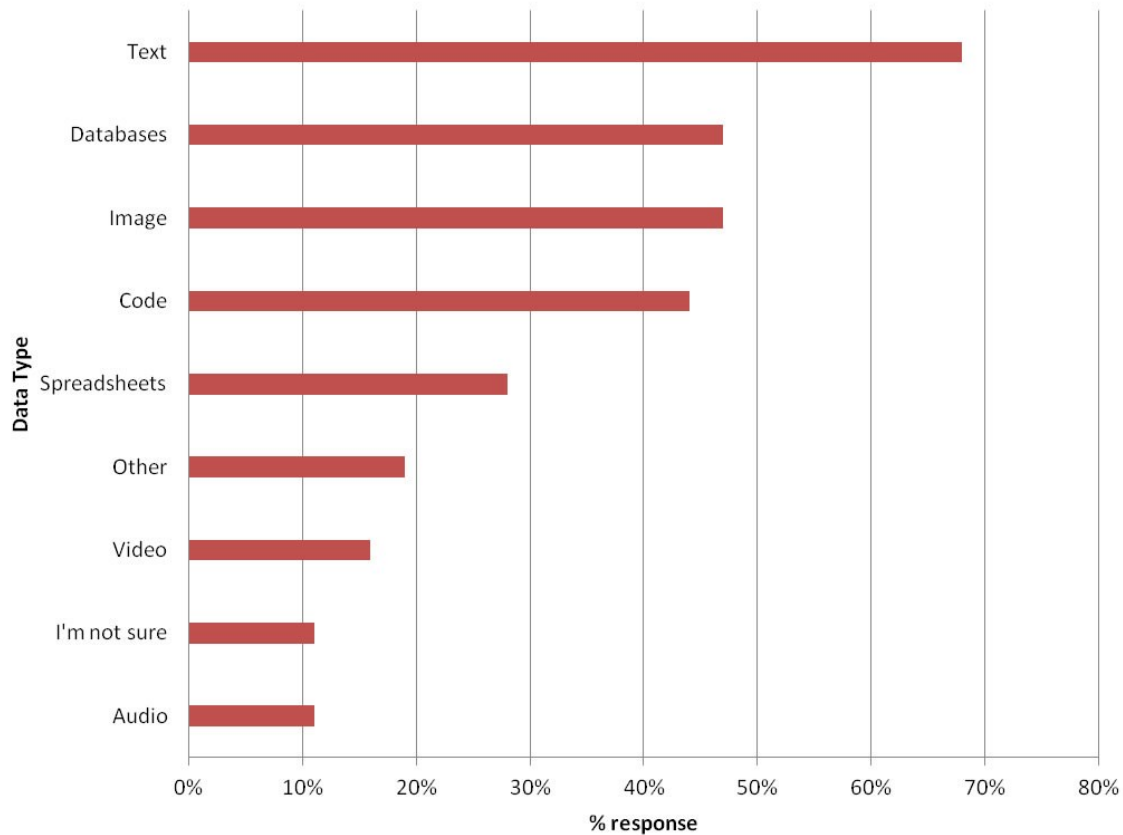
Question	Response		
	Yes	No	I'm not sure
Would you be interested in any sort of guidance, including consultation, for writing a data management plan in support of an NSF grant application?	53 (62%)	11 (13%)	22 (26%)
Does the data you have produced or intend to produce conform to known standards in your discipline?	37 (45%)	10 (12%)	35 (43%)
Have you produced or do you anticipate producing metadata for this project?	33 (42%)	20 (26%)	25 (32%)
Does the metadata you have produced or intend to produce conform to known standards in your discipline?	10 (30%)	3 (9%)	20 (61%)
Do you anticipate need to consult with an intellectual property specialist to create a license agreement or usage statement for the data you have produced or intend to produce?	7 (9%)	49 (65%)	19 (25%)
When you publish your findings from this research project, do you plan on submitting your supporting data to a journal publisher?	23 (31%)	28 (38%)	23 (31%)
Do you plan to deposit the data you have produced or intend to produce in Cornell's institutional repository, eCommons, or would you be interested in doing so to satisfy the NSF requirement?	18 (24%)	10 (14%)	46 (62%)
Do you plan to deposit the data you have produced or intend to produce in CISER's Data Archive, or would you be interested in doing so to satisfy the NSF requirement?	7 (9%)	30 (41%)	37 (50%)
Do you plan to utilize the Cornell Restricted Access Data Center (CRADC) to work with restricted access data or limited use licensed data, or would you be interested in doing so to satisfy the NSF requirement?	4 (5%)	35 (47%)	35 (47%)
Do you plan to store the data you have produced in the Center for Advanced Computing disk farm, or would you be interested in doing so to satisfy the NSF requirement?	9 (12%)	24 (32%)	41 (55%)
Do you plan to deposit the data you have produced or intend to produce in a data center or other non-Cornell repository, or would you be interested in doing so to satisfy the NSF requirement?	19 (26%)	20 (27%)	35 (47%)
Do you anticipate or would you be interested in any sort of guidance, including consultation or instruction, for any of the data management plan components mentioned above?	50 (68%)	9 (12%)	15 (20%)
Do you currently keep track of research outputs and their availability?	51 (69%)	23 (31%)	N/A
If there were a service where you could enter basic information about your data (description, availability on the web) to demonstrate compliance with NSF's policy, would you use it?	36 (49%)	8 (11%)	30 (41%)

ing one; the remainder responded "I'm not sure." Three respondents commented on the newness of the requirement and their uncertainty in how to meet it, while three others noted that they already share data or were aware of infrastructure for sharing in their discipline.

*Types of data and other materials to be produced*

We asked researchers about general types of data (file types) as well as a list of specific file extensions to get a sense of the diversity of data researchers produce. In response to the more general question, text, image, data-

**Figure 2:** Responses to the question “Please specify the types of data you have produced or anticipate producing for this project that you intend to share with others.” Respondents were asked to select all that apply.



bases, and code were the most common answers (see Figure 2); 41 researchers (48%) reported generating three or more of the data types listed in Figure 2.

When asked for specific file extensions, researchers reported 77 unique file extensions, 39 of which were listed by only a single researcher (Table 2). Code or script files were the most common class of extensions, followed by numeric and image/graphic file extensions. Several researchers included Microsoft Word and Adobe PDF in their lists of file extensions for data. Three of the 64 respondents (4.7%) to this question named more than nine specific file extensions; 52 (81%) named five or fewer. Three researchers responded “??” or “What is data?” to this

question, indicating there is some confusion as to what research outputs are considered “data.”

The results suggest two challenges for providers of data management services. First, considerable confusion exists as to what “counts” as data, even among researchers who are likely among their discipline’s experts. Second, providers of data services will encounter a very broad array of digital content in the course of planning and delivering data management services. This will be particularly challenging for those working to preserve digital research data for the long term.

**Table 2:** Data file extensions reported by survey respondents. No attempt was made to verify reported extensions.

Reported extension	Count
pdf	26
txt	17
doc	15
xls	15
jpeg	14
tiff	9
docx	8
m	7
gif	6
csv, cvs, ppt, tex	5
c, mov, wav	4
cpp, html, mat, png, ps, py, wmv, xslx	3
custom, dat, dta, eps, f, fits, h, hdf, java, mpeg, pptx, psrfits, Rdata, sav, sql	2
1fl, 3d, ai, avi, bit, cc, do, dvi, dxf, emf, f90, fas, fna, gbk, gff, htm, IDL, img, jdx, kmz, m2, mp3, mtl, o, obj, pds, profile, Rd, rtf, sd2, sh, spit, spss, stata, svg, tar.gz, tsv, v, xmed	1

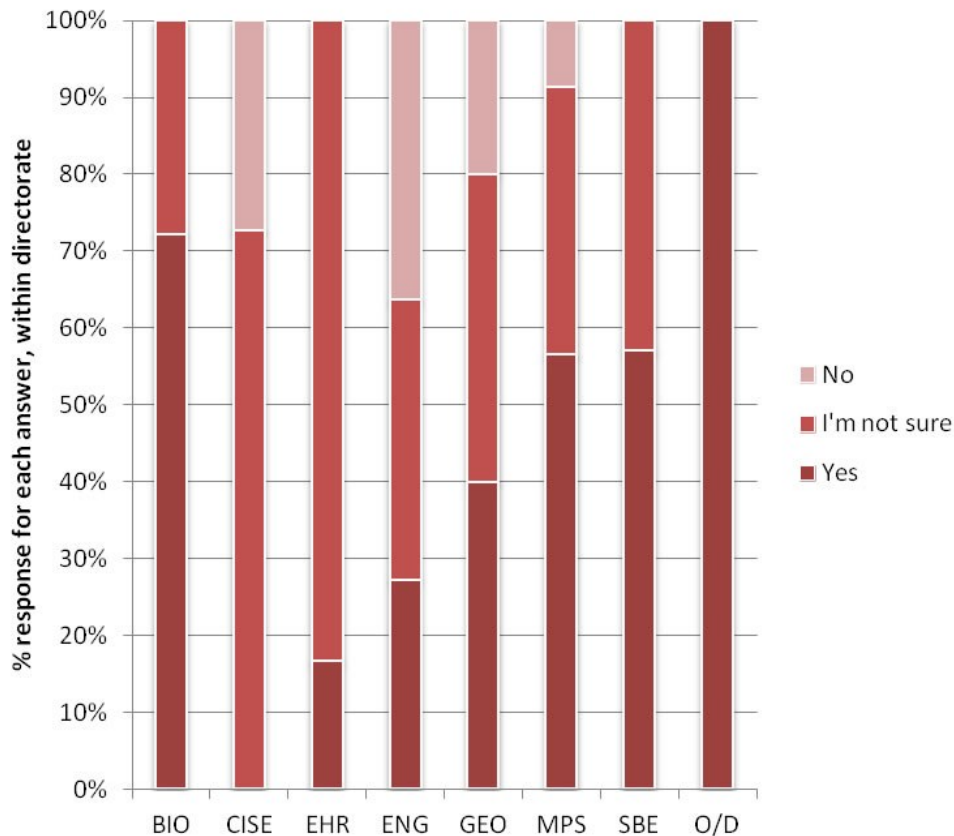
*Standards used for data and metadata*

Thirty-five respondents (43%) said they don't know if their data conform to disciplinary standards, and 10 (12%) indicated that their data do not conform to disciplinary standards, demonstrating a general lack of application of standards to data management (Table 1). This varies across directorates; however, with more than 50% responding that their data do conform to standards among submitters to BIO, MPS, and Social, Behavioral & Economic Sciences (SBE; Figure 3). Note that only one respondent submitting to the Office of the Director (O/D; includes Office of Cyberinfrastructure, Polar Programs, and others) answered this question; there were at least five (and as many as 23) respondents submitting to each of the other directorates. When asked to specify the standards they use (a free-text response), responses varied from specific known standards to generalized descriptions of standards ("library standards" and

"generic publishing expectations") to statements indicating that respondents found the question or the topic to be confusing ("don't know" and "no idea what you're asking").

Less than half the respondents (33, or 42%) reported that they have created or plan to create metadata for their data sets (Table 1). Of those that do create metadata (or plan to), slightly less than one-third indicated that the metadata they create conforms to disciplinary standards (Table 1). Adoption of standards appears to be highest (at least 33%) among responders submitting to the BIO, Computer & Information Science & Engineering (CISE), and SBE directorates (the sole respondent submitting to O/D also answered "Yes" to this question). As with data, some respondents' comments, such as "I don't know what metadata is," indicate confusion or lack of familiarity with these concepts. In spite of this uncertainty about what metadata are and best practices for producing it, nearly two-thirds of respondents re-

**Figure 3:** Responses by discipline to the question: “Does the data you have produced or intend to produce conform to known standards in your discipline?” BIO=Directorate for Biological Sciences, CISE=Directorate for Computer & Information Science & Engineering, EHR=Directorate for Education & Human Resources, ENG=Directorate for Engineering, GEO=Directorate for Geosciences, MPS=Directorate for Mathematical & Physical Sciences, SBE=Directorate for Social, Behavioral & Economic Sciences, and O/D=Office of the Director (includes Office of Cyberinfrastructure, Polar Programs, and others).



ported they would not use a metadata service, whether fee-based or free of charge (Figure 4).

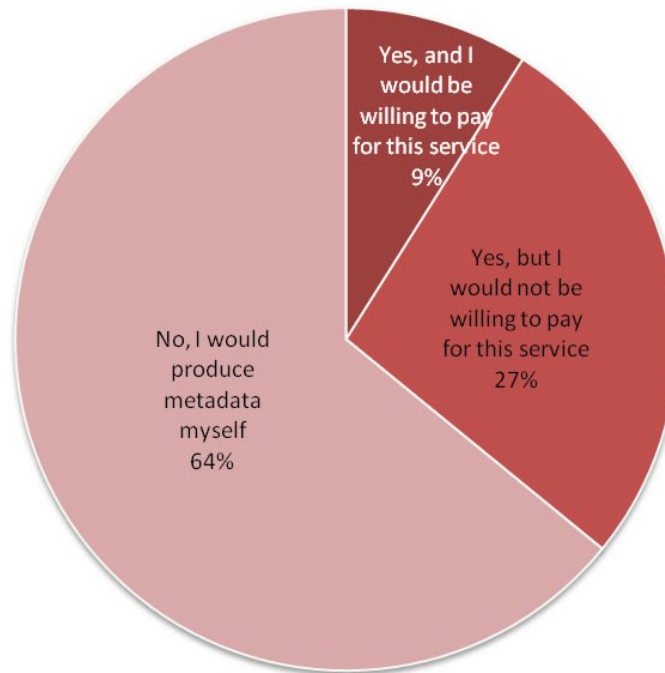
Responses to questions about standards for data and metadata showed some confusion among researchers as to the meaning and application of these terms. When asked, for example, to list the data and metadata standards used, several researchers described the use of methods or protocols that standardize data collection and management within a research group. We assume that the NSF’s recommendation that researchers specify “standards to be used for data and

metadata format and content” (National Science Foundation 2011) has more to do with formally recognized (or at least de facto) standards within a scientific discipline, not just within a research group or laboratory.

For service providers, this same general confusion suggests an opportunity for outreach and education to help researchers understand the nature of data standards as well as the value and utility of metadata for research data. The apparent reluctance to make use of a metadata service, whether free or fee-based, may be a product of this lack of understanding of the real value of



**Figure 4:** Responses to the question “Would you make use of a service to produce metadata for this project?”



metadata. Service providers may also wish to consider other service delivery strategies, such as face-to-face or virtual training on data and metadata standards and metadata creation.

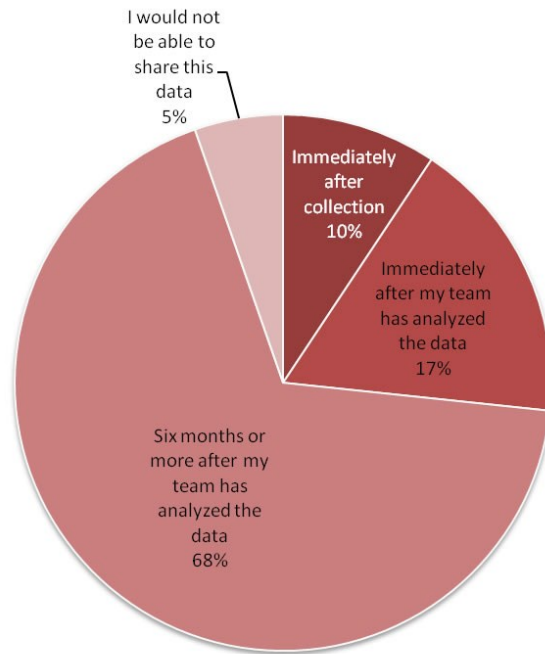
*Policies and provisions for access, sharing, confidentiality, security, and intellectual property*

A majority of respondents (65%) reported no need for assistance from an intellectual property specialist to develop usage statements for or apply licenses to their data sets (Table 1). This is perhaps consistent with a general willingness to share data; 95% reported they would be able to share their data at some stage in their research. The majority (68%) said they would prefer to wait until at least six months after analyzing their data within their research group to share their da-

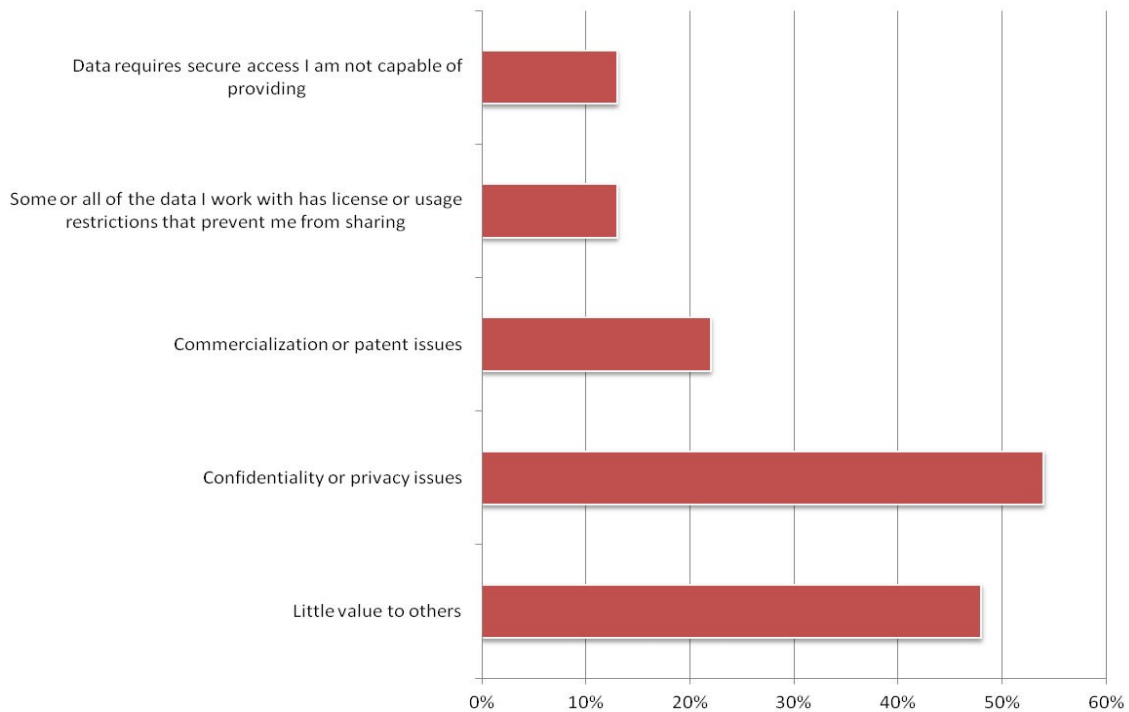
ta (Figure 5). While most respondents expressed a willingness to share, 46 respondents indicated there are circumstances that would prevent them from sharing at least some of their data (Figure 6). The most often cited reasons for not sharing are confidentiality and privacy issues (54%) and a sense that the data hold little value to others (48%). Free text comments following these questions suggested some additional concerns. Some researchers reported that they themselves use data that may be subject to restrictions that would preclude sharing, or they work with collaborators that might not permit data sharing. Concerns over being “scooped” as well as outright fraud were also expressed. Practical issues – lack of resources and the overall volume of data among them – were also raised. Considered together, it appears that even though 95% of respondents indicate a willingness to share



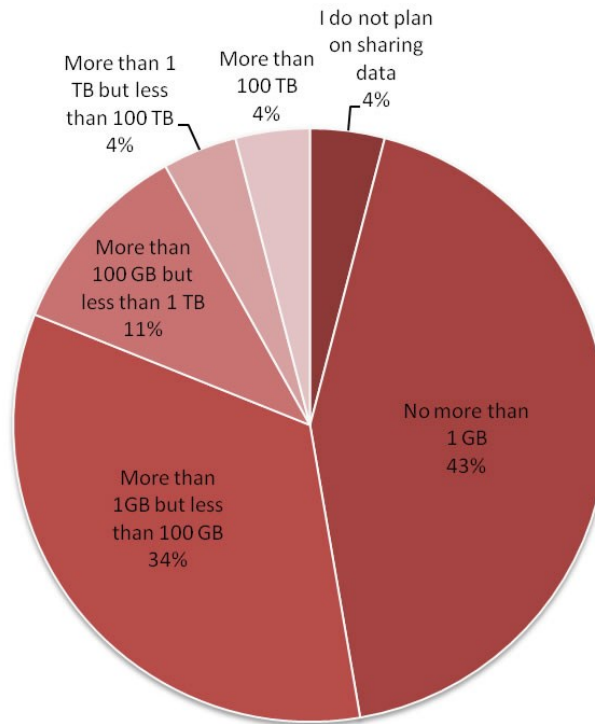
**Figure 5:** Responses to the question: “When would you be able to share the data you have produced or intend to produce for this project?”



**Figure 6:** Responses to the question: “What might prevent you from sharing the data you have produced or intend to produce for this project? Respondents were asked to select all that apply.”



**Figure 7:** Responses to the question: “Given the NSF expectation to share data with other researchers, how much data would you intend to share?”



their data at some point, at least some of those respondents would withhold some portion of their data.

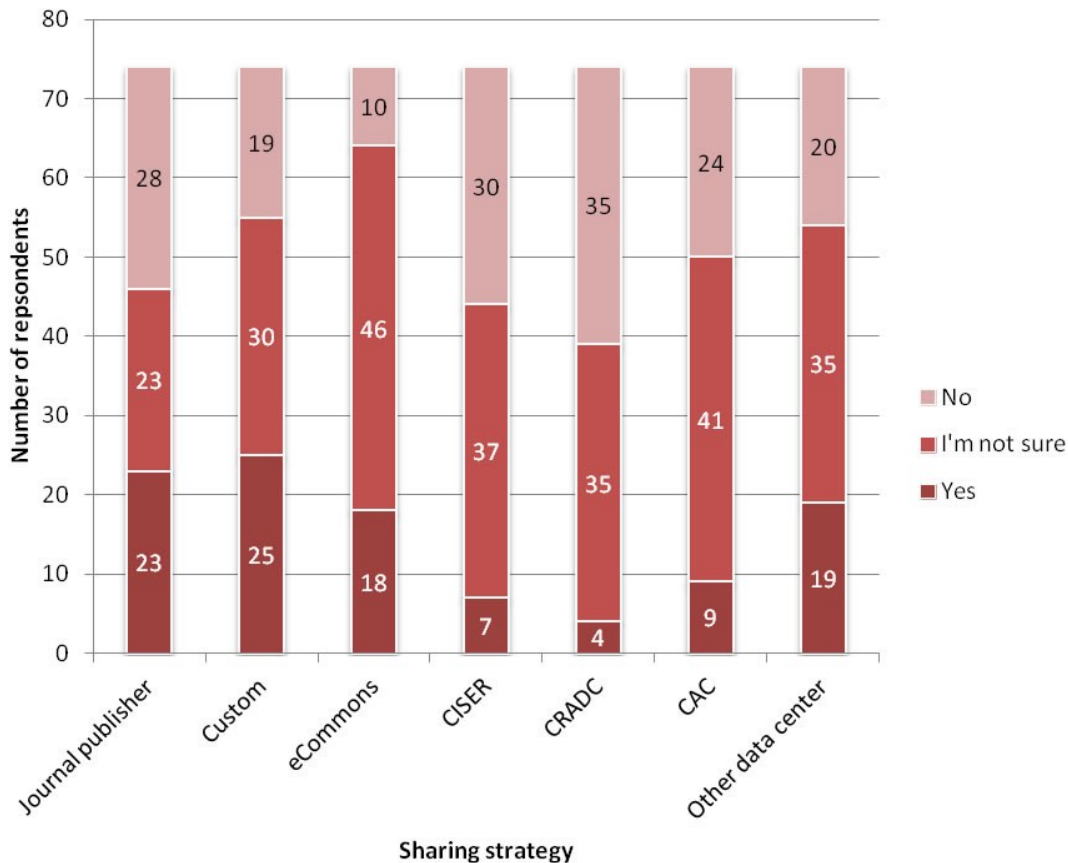
The primary challenges this set of responses suggest are a need to assist researchers in identifying the data that would provide the most benefit if shared, and to address privacy and confidentiality concerns that may stand in the way of sharing data. One possible strategy to address either concern would be to create a forum for researchers to share strategies and success stories. For privacy and confidentiality concerns, infrastructure that meets all compliance requirements would clearly be useful, but many institutions will likely find developing and sustaining such infrastructure to be a significant undertaking.

*Policies and provisions for re-use, re-distribution, archiving and preservation of data and other materials*

The amount of data to be shared can impact which strategies for providing access to data are available. Seventy-seven percent of respondents indicated they plan to share less than 100GB of data. Eleven percent reported plans to share more than 100GB and less than 1 TB, 4% reported plans to share between 1 and 100TB, and 4% reported plans to share more than 1TB (Figure 7).

We asked researchers what infrastructure they plan to use for sharing their data. Possible responses included several systems at Cornell as well as external systems. With the exception of custom solutions developed by researchers themselves, the possible responses are likely to satisfy both access (sharing) and preservation functions. We did

**Figure 8a:** Strategy for making data accessible. “Journal” indicates respondent will submit data with manuscript for publication. “Custom” indicates the respondent will develop their own solution for sharing data. “eCommons” is Cornell University Library’s institutional repository; “CISER” is the Cornell Institute for Social and Economic Research data archive; “CRADC” is the Cornell Restricted Access Data Center; and “CAC” is the Cornell Center for Advanced Computing.

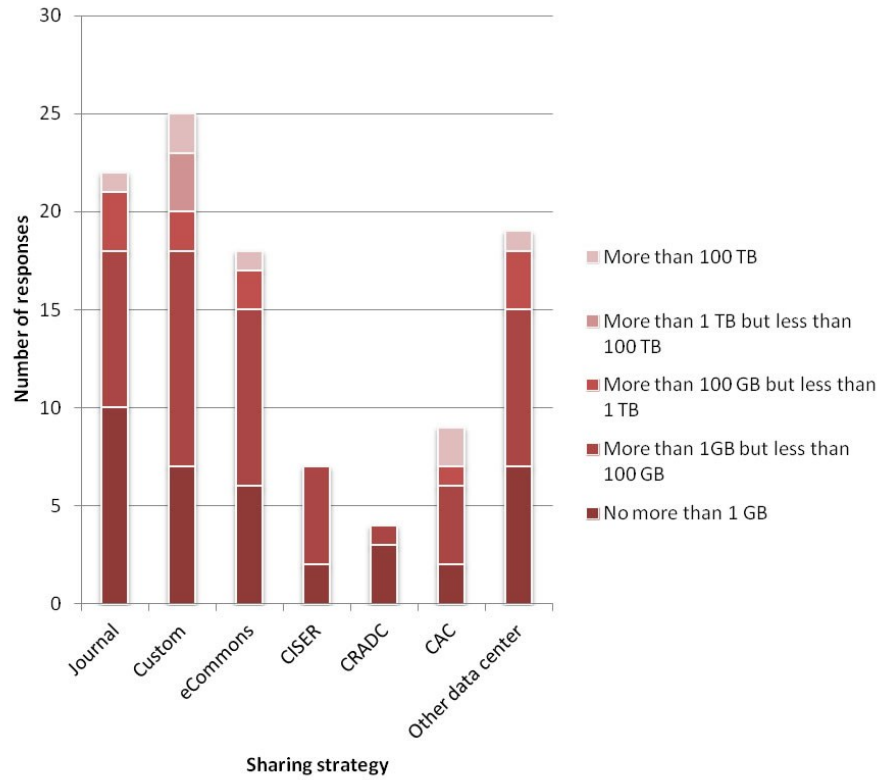


not ask explicitly about researchers’ plans for ensuring preservation of their data. Seventy-four researchers addressed this set of questions. “Custom” solutions developed by the researchers themselves was the most common answer, followed by submitting data to journal publishers as supplemental materials along with manuscripts, depositing data in disciplinary data centers, and using the Cornell University Library’s institutional repository, and other more specialized facilities at Cornell (Figure 8a). The “Custom” strategy actually had two options: one indicating researchers planned to handle the

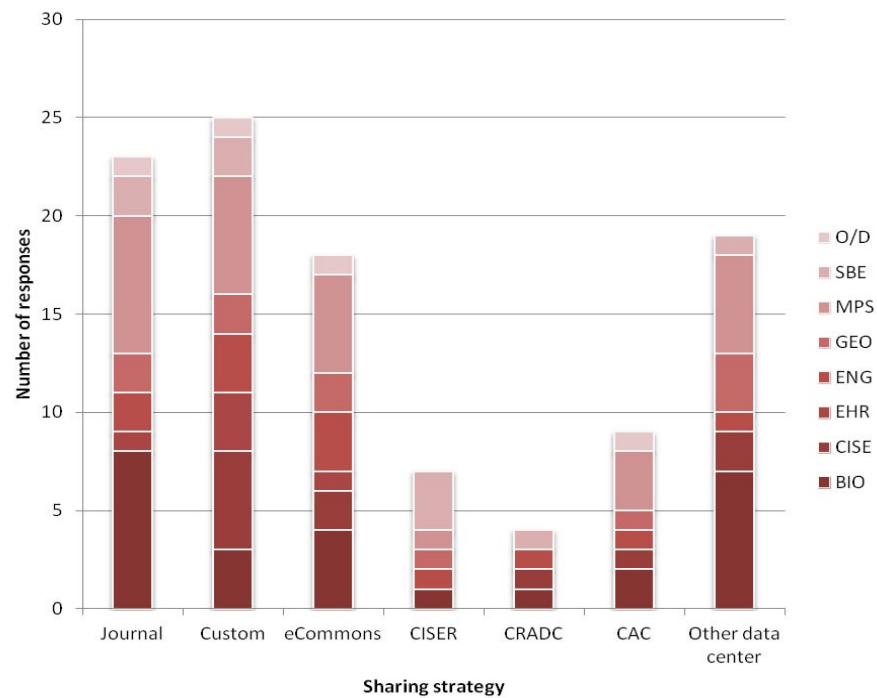
tasks themselves, and the other indicating they would outsource the tasks. All researchers planning to implement custom solutions intended to do so themselves, and not to outsource. We should also note that the response “I’m not sure” was selected as or more frequently than “Yes” or “No” for each of these questions, indicating a fair amount of uncertainty as to how to handle data sharing. Figure 8b shows how researchers with data sets of different sizes said they intend to share them.

It is worth noting here the small number of

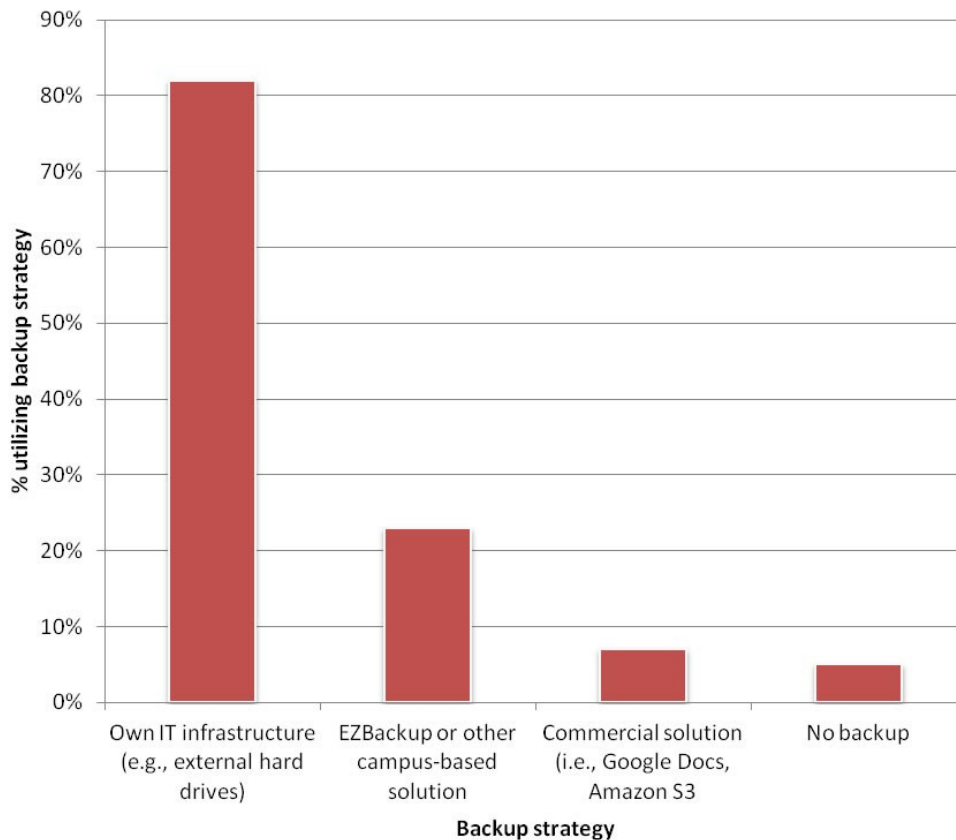
**Figure 8b:** Strategy for making data accessible, by size of data.



**Figure 8c:** Strategy for making data accessible, by directorate.



**Figure 9a:** Responses to the question: “What is your current method of backing up the data you have produced or intend to produce for this project? Check all that apply.”

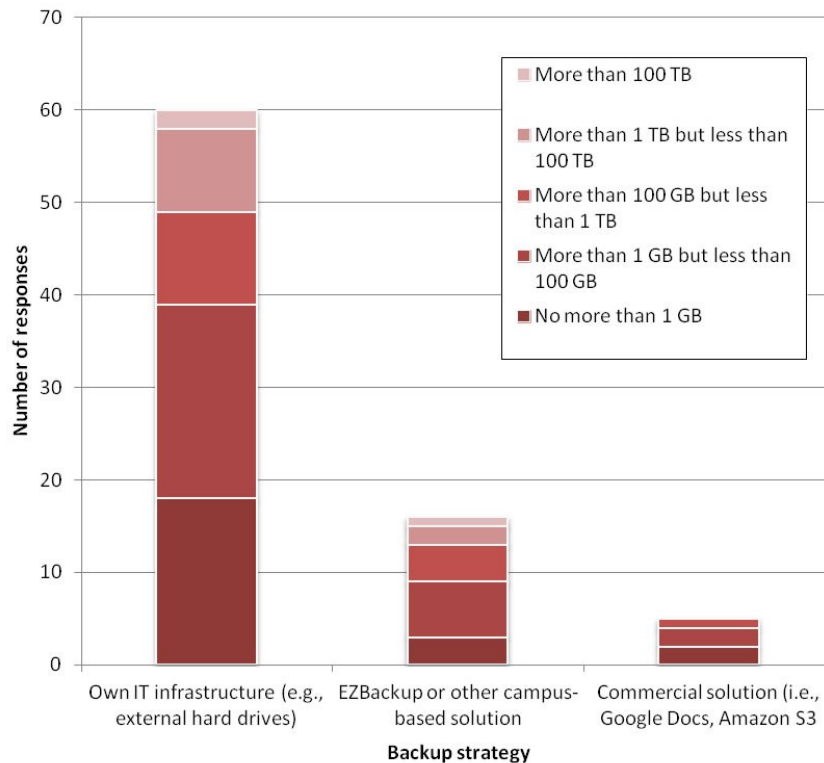


cases where the selected strategy is a mismatch for the size of data: researchers intending to share more than 100TB of data indicated they may do so by means of nearly every possible strategy offered, regardless of whether the strategy is an appropriate one for very large data sets. For example, eCommons recently revised its size limits for data sets upwards to 1GB per object and 10GB per project per year, but it remains effectively off-limits for researchers with larger quantities of data (Cornell University Library, n.d.). A similar mismatch is apparent when we look at plans for data sharing by directorate: The CISER data archive is intended for social science data, yet respondents from the physical and biological sciences included it as one of their potential strategies for sharing data (Figure 8c). These

mismatches will require service providers to manage researchers' expectations about service capabilities, and redirect them to more appropriate services.

We also asked researchers for information on their backup practices. More than 80% indicated they rely on their own infrastructure for backups; 23% use a campus service for backups; 7% use a commercial solution; and 5% reported not backing up their data at all (Figure 9a). The only pattern we observed when we examined backup practices by size of data collection is that the small number of researchers who reported having more than 1TB of data to backup do not use commercial solutions for the task, and all of them reported backing up their data (Figure 9b).

**Figure 9b:** Backup strategy by size of data collection.



Not surprisingly, the comments associated with these questions support our finding that this is confusing terrain for researchers and that expectations are unclear. One comment in particular – “My data is in my papers,” implies that no further sharing is necessary, and reinforces the confusion over what constitutes as data. Researchers may consider tables and figures, which are representations of the underlying data, to be the same thing as the data upon which these representations are based. Others noted that it either is or is not standard practice in their discipline to submit data along with manuscripts for publication, that journals don’t necessarily require supporting data, and the difficulty in deciding which data should accompany a manuscript.

Some researchers commented that they are already using or plan to use campus-based solutions for data sharing. These include an enterprise wiki hosted by central IT services,

working with the Center for Advanced Computing, and solutions developed within their own labs. At the same time, for each specific campus-based solution identified in the survey (eCommons, CISER, the Cornell Restricted Access Data Center or CRADC ([http://ciser.cornell.edu/CRADC/What is CRADC.shtml](http://ciser.cornell.edu/CRADC/What%20is%20CRADC.shtml)), and CAC), multiple researchers commented that they were unfamiliar with that particular service, and some commented that they would prefer to use infrastructure specific to their discipline. Discipline-based repositories named by respondents include the Inter-University Consortium for Political and Social Research (ICPSR), GenBank and other National Center for Biotechnology Information (NCBI) resources, arXiv, and others. Researchers also mentioned discipline-agnostic services such as Amazon, Google Docs, and Drop-Box.

While researchers do indicate interest in us-

ing external services for providing access to data, the potential impact on institutional services is substantial. We expect use of institutional repositories for data sets to increase, as well as the use of central IT services (such as storage and web hosting) and more specialized data services.

### *Tracking research outputs*

Researchers can reasonably be expected to indicate the availability of research outputs in interim and final project reports, as well as in subsequent grant proposals where they are asked to report on the results of prior support. To help assess whether infrastructure to support this function would be useful, we asked if researchers keep track of research outputs and their availability. A majority of researchers (69%) responded that they do (Table 1), although respondents' comments indicate that the question was confusing. In comments, researchers reported that they do track publications, although perhaps not in a highly organized way. Responses to a follow-up question asking whether researchers would be interested in supplying information about their data to demonstrate compliance with the NSF policy showed significant interest in such a service: 49% of respondents reported they would use it; 41% were not sure (Table 1).

### **Conclusions**

The overarching finding from our survey is that there is a great deal of uncertainty among PIs about what the new NSF requirement means and how to meet it, and that researchers welcome offers of assistance – both with data management planning, and with specific components of data management NSF asks them to address in their plans (68% respond yes to the latter, Table 1). In fact, for survey questions where “I’m not sure” was a possible response, at least 20% of respondents chose that answer for all of the questions in Table 1. This uncertainty was further borne out in the comments

researchers made on multiple survey topics, and by responses to a question asking whether researchers want guidance on any aspect of data management planning (68% reported that they do). This is an interesting finding given that the NSF’s policy leaves much of the detail to “communities of interest” and peer review, as well as program management (NSF 2010c); it seems reasonable to assume that representatives of various “communities of interest” as well as prospective peer reviewers participated in the survey.

Researchers’ comments expressed their frustration that while NSF guidelines allow for costs associated with data management to be included in proposal budgets, the overall size of awards is not increasing to accommodate this new expense. This may contribute to their reluctance, when asked, to utilize a for-fee service for metadata creation (although some are willing to use a service if it’s free of charge). Another difficult challenge for institutional service providers is developing charge-back models that allow researchers to pay up front for costs incurred beyond the end of a research grant. Researchers were also concerned that meeting the requirement will take too much time away from research.

Taken together these results suggest some important challenges for institutions attempting to meet the data management requirements of their researchers, and for funders that are moving toward implementing new requirements. From the comments in the survey, we infer (not surprisingly) that researchers who already share data are reasonably comfortable with and capable of managing this task. Those for whom sharing data is new may be frustrated and are uncertain how to approach it. Researchers lack sufficient information about the services available to them, or do not fully understand the capabilities and limitations of those services. Service providers will need to manage expectations of local services, guide researchers to the services that best meet



their needs, and offer guidance in best practices to meet funders' requirements. Cornell's early efforts in meeting these new requirements have centered largely on education and outreach on data management planning, and one-on-one consultations for grant proposal writers and researchers with specific data management needs at other stages of the research process. In terms of infrastructure, work is underway to expand Cornell's VIVO (<http://vivo.cornell.edu/about>) application to support the basic description of data sets, as a tool to support the tracking of research outputs, and we are in the process of evaluating the need for additional infrastructure to support researchers' data management needs.

## References

Block, Bill, Eric Chen, Jim Cordes, Dianne Dietrich, Dean Krafft, Stefan Kramer, Dave Lifka, Janet McCue, and Gail Steinhart. 2010. Meeting Funders' Data Policies: Blueprint for a Research Data Management Service Group (RDMSG). Retrieved February 15, 2012, from <http://hdl.handle.net/1813/28570>.

Cornell University Library. n.d. eCommons Policies: Data Deposit Policy. Retrieved October 23, 2011, from <http://ecommons.library.cornell.edu/policy.html#data>.

Interagency Working Group on Digital Data. 2009. Harnessing the Power of Digital Data for Science and Society: Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. Retrieved July 7, 2011, from [http://www.nitrd.gov/about/Harnessing\\_Power\\_Web.pdf](http://www.nitrd.gov/about/Harnessing_Power_Web.pdf).

National Science Foundation. 2010a. Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans – Press Release 10-077. Retrieved September 4, 2010, from [http://www.nsf.gov/news/news\\_summ.jsp?](http://www.nsf.gov/news/news_summ.jsp?)

[cntn\\_id=116928&org=NSF&from=news](http://www.nsf.gov/news/cntn_id=116928&org=NSF&from=news).

National Science Foundation. 2010b. Dissemination and Sharing of Research Results. Retrieved October 6, 2010, from <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.

National Science Foundation. 2010c. Data Management & Sharing Frequently Asked Questions (FAQs). Retrieved July 5, 2011, from <http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp>.

National Science Foundation. 2011. Proposal and Award Policies and Procedures Guide. Retrieved July 7, 2011, from <http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/>.

Open Government Partnership. 2011. National Action Plan for the United States of America. Retrieved October 13, 2011 from [http://www.whitehouse.gov/sites/default/files/us\\_national\\_action\\_plan\\_final\\_2.pdf](http://www.whitehouse.gov/sites/default/files/us_national_action_plan_final_2.pdf).

Steinhart, Gail, Eric Chen, Florio Arguillas, Dianne Dietrich, Stefan Kramer. 2011. Research Data Management Service Group Survey of NSF Principal Investigators at Cornell University. Retrieved November 14, 2011 from <http://hdl.handle.net/1813/25624>.

*Disclosure:* The authors report no conflicts of interest.

\* *Present Address:* German Institute for Economic Research (DIW Berlin), Berlin, Germany

All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a Creative Commons Attribution 4.0 International License.

<http://creativecommons.org/licenses/by/4.0>

ISSN 2161-3974