

1-1-2014

The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories

Cheryl D. Walters
Utah State University

Recommended Citation

Walters, Cheryl D., "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories" (2014). *Library Faculty & Staff Publications*. Paper 158.
http://digitalcommons.usu.edu/lib_pubs/158

This Article is brought to you for free and open access by the Merrill-Cazier Library at DigitalCommons@USU. It has been accepted for inclusion in Library Faculty & Staff Publications by an authorized administrator of DigitalCommons@USU. For more information, please contact becky.thoms@usu.edu.



The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on
Faculty, Sponsored Programs, and Institutional Repositories

Anne Diekema, Utah State University

Andrew Wesolek, Clemson University

Cheryl Walters, Utah State University

Abstract

The scholarly communication landscape is rapidly changing and nowhere is this more evident than in the field of data management. Mandates by major funding agencies, further expanded by executive order and pending legislation in 2013, require many research grant applicants to provide data management plans for preserving and making their research data openly available. However, do faculty researchers have the requisite skill sets and are their institutions providing the necessary infrastructure to comply with these mandates? To answer these questions, three groups were surveyed in 2012: research and teaching faculty, sponsored programs office staff, and institutional repository librarians. Survey results indicate that while faculty desire to share their data, they often lack the skills to do this effectively. Similarly, while repository managers and sponsored programs offices often provide the necessary infrastructure and knowledge, these resources are not being promoted effectively to faculty. The study offers important insights about services academic libraries can provide to support faculty in their data management efforts: providing tools for sharing research data; assisting with describing, finding, or accessing research data; providing information on copyright and ownership issues associated with data sets; and assisting with writing data management plans.

Keywords: data management, institutional repositories, scholarly communication, data sets

The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on
Faculty, Sponsored Programs, and Institutional Repositories

Introduction

In response to the Office of Management and Budget's revised Circular A-110 (Fischer, 2013), U.S. federal funding agencies added new data management and data sharing requirements to their grant applications. The National Science Foundation (NSF), the National Institutes of Health (NIH), and the National Endowment for the Humanities (NEH) began requiring the submission of a data management plan as a component in all or certain kinds of grant proposals. For example, the NIH Grants Policy Statement reads "“Timely release and sharing” is defined as no later than the acceptance for publication of the main findings from the final data set. All investigator-initiated applications with direct costs of \$500,000 or more (excluding consortium F&A costs) in any single year are expected to address data-sharing in their application"

(http://grants.nih.gov/grants/policy/nihgps_2012/nihgps_ch8.htm#_Toc271264947).

Moreover, data sharing pertains to more than just data sets; research tools developed under grant-funded research are expected to be shared also:

Investigators conducting biomedical research frequently develop unique research resources. NIH considers the sharing of such unique research resources (also called research tools) an important means to enhance the value of NIH-sponsored research. Restricting the availability of unique resources can impede the advancement of further research. Therefore, when these resources are developed with NIH funds and the associated research findings have been accepted for publication, or after they have been provided to NIH, it is important that they be made readily available for research purposes to qualified individuals within the scientific community.

While specific directions might differ per NSF directorate, the NSF Award and

Administration Guide has similar language: “Investigators are expected to share with

other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing”

(http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4).

These newly-required data management plans not only ensure the long-term preservation of data sets and research tools generated by agency-funded research, but also open them for sharing with other researchers. Once shared, this data, which the circular defines as “materials necessary to validate research findings, excluding preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, and communications”

(Fischer, 2013) can be repurposed and reused, generating new research or facilitating teaching. To those interested in the open dissemination of data and research, this was welcome news indeed.

However beneficial in the long run, these requirements create challenges for grant applicants who may or may not have existing skill sets or existing support infrastructures to comply with such mandates. Do universities and libraries have infrastructures in the form of systems or educational programs in place to manage this data and support faculty adherence to funder requirements? To answer this question, this study conducted in 2012/2013 investigates the impact of funder mandates on three different groups: traditional teaching or research faculty (henceforth referred to simply as “faculty”), sponsored programs offices, and academic librarians. Grant requirements clearly affect people applying for grants and on university campuses, those grant applicants generally

are faculty engaged in research. Sponsored programs offices typically are the units on campus responsible for reviewing grants before researchers submit them to funding agencies; they offer varying degrees of assistance to grant applicants. Likewise, many academic librarians assist researchers in a variety of ways, from writing the grant to hosting in a repository the articles and data sets resulting from grant-funded research.

This study seeks to answer the following research questions:

1. What has been the nature and extent of the effect on universities and researchers of new granting agency requirements for inclusion of data management plans in grant applications?
2. How are researchers storing and providing access to new data generated by their grant projects?
3. Are faculty using existing data stored in repositories for their own new research?
4. Are faculty incorporating use of research data sets into their teaching?

The paper is organized as follows: The Background section situates this study in the existing literature. The Methods and Data section describe the study's research questions and survey approach. The Results section and Discussion section respectively describe and discuss the study. Finally, the Conclusion section addresses the impact of data management and sharing on faculty, librarians, and sponsored programs personnel.

Background

E-science, data management and the role of librarians

Mandates from funding agencies to store, manage, and share research data have emerged largely in response to e-science. E-science is a relatively new data-intensive science typically carried out over distributed networks using grid computing capabilities to generate and analyze extremely large data sets. Data sets resulting from sensors, satellite surveys, and computing simulations are now generated at a fast rate (Hey & Hey, 2006), resulting in what is sometimes referred to as a data deluge (Marcum & George, 2012). E-science can be highly distributed in a collaborative sense and members of research teams may reside anywhere in the world (NSF, 2011). This distributed and intensive generation and use of born-digital data sets requires the storage, careful management, curation, and preservation of this data so it can be shared with the research team and made accessible to the research community at large (Liotta et al., 2005).

Extensive data management going beyond simple local storage falls outside the traditional realm of scientists who have seen changes in both their information environments and data requirements. Faculty in a 2008 study cited a definite need for guidelines and direction on data management (Delserrone, 2008). Qin and D'Ignazio (2010) studied how data management practices and the use of data have changed for scientists. They found both the extent of data generation and use frequency of external data sets were highly discipline dependent. Further, they found that while researchers found metadata useful for discovering and using external data sets generated by others, they were not likely to put much effort into adding metadata to their own data sets in order to enhance their accessibility by other researchers. Decisions on metadata were often made locally, ignoring guidelines from national research bodies.

While faculty appear to lack both the expertise and desire to ensure long-term access and preservation of their data sets, librarians and archivists are increasingly knowledgeable or willing to learn how to manage, describe, and preserve data sets. Many university libraries now have digital repositories to store research output and make it accessible via the web. Although repository content at present consists primarily of journal articles, many see the addition of data sets as a logical fit (e.g. Hey & Hey, 2006; Alvaro, Brooks, Ham, Poegel, & Rosencrans, 2011; Ramírez, 2011, Mullins, 2009).

Successful handling of research data will require librarians and archives to learn about three important aspects of data management: the data life cycle, technical aspects (storage, indexing, retrieval), and social and policy issues (Qin & D'Ignazio, 2010).

Scientific communication no longer ends with the publication of a journal article. Increasingly, data sets and databases are part of the scientific conversation and librarians need to consider how to make this new type of data accessible for data analysis by other researchers (Gold, 2007a, 2007b). Data sets pose technical challenges due to their size and changing nature, placing high demands on data storage facilities. Data sets sometimes are dynamic, resulting from sensors (e.g. weather recording devices) automatically collecting data and possibly requiring special interfaces between digital repositories and these data collection devices (Luce, 2012). Data sets may be distributed across various national and international institutions. Making this distributed data accessible requires not only new approaches for storage but also descriptors that work across disciplines and borders. Data sets also often require specialized metadata fields such as provenance, data parameters, and conditions of use. Lastly, social and policy issues associated with data sets have also changed. For example, since data sets now

need to be stored and maintained beyond the duration of research projects, new policies need to be in place determining who is financially responsible for their maintenance (Macdonald & Martinez-Uribe, 2009, 2010).

Data Management Services

Many universities and libraries have infrastructures in the form of systems or educational programs for managing data sets (Soehner, Steeves, & Ward, 2010; Carlson, Ramsey, & Kotterman, 2010). They offer a variety of services to assist researchers in meeting funder mandates by helping them manage, store, share, use, and cite data. Some institutions offer assistance with writing a data management plan for grant applications; this writing assistance can come from the library or Sponsored Programs office. Many institutions provide centralized data storage for archiving, collaboratively working on, or sharing data. Institutional repositories permanently manage, disseminate, and preserve data generated by university research activities. Some universities offer temporary storage while research projects are underway (e.g. DataStaR at Cornell University), providing collaborative open spaces where team members can communicate with each other and eventually share project results with the scientific community (Steinhart, 2010). Another new service is the provision of data literacy programs that teach researchers about the life cycle of data and how to find, use, and cite data sets (Gore, 2011). Other institutions are assisting researchers with high performance computing structures such as grid computing to crunch computational problems and analyze complex data sets. Experienced in metadata standards and information classification, librarians are helping researchers discover, access, and use the work of others via descriptive metadata and

logical arrangement (Gore, 2011; Witt, 2012). Purdue University Research Repository or PURR (<https://purr.purdue.edu/>) offers a virtual space where researchers can get help creating a data management plan, upload research to a project which is sharable with collaborators, or publish a data set with a Datacite Digital Object Identifier (DOI). Libraries appear well positioned to bridge the gap between disciplines, institutions, and campus departments in supporting faculty adherence to mandate requirements (Jones, 2008; Gore, 2011; Dietrich, Adamus, Miner, & Steinhart, 2012; Hswe & Holt, 2011).

Methods and Data

This study investigates the extent to which faculty are supported by librarians and their sponsored program offices with regards to the new data management requirements. Data were collected via three separate online surveys (Babbie, 1973) tailored and sent in mid-2012 to the three groups of interest: STEM faculty, sponsored programs offices (SPO), and academic librarians presumed to be affiliated with an institutional repository (IR). Each survey targets one of the groups and asks research questions pertaining to that particular group. All surveys, administered through SurveyMonkey, are intentionally brief to minimize the effort of the respondents. The sponsored program (SPO) survey has six questions, the academic librarian/institutional repository (IR) survey eleven, and the faculty survey sixteen. The questions are mostly multiple choice, with an open ended option to allow participants to elaborate on their responses. The survey questions for all three surveys are included in Appendix A.

Participants for all three categories are drawn from universities with a Carnegie classification of either high research (RU-H) or very high research (RU-VH) or land

grant universities irrespective of their Carnegie classification. The research Carnegie classification and land grant designation ensures that participants come from universities generating significant grant activity and therefore are more likely to be affected by the NSF and NIH data management plan mandates that we are investigating. One hundred eighty three institutions fit this institutional profile.

Individual participants in all three categories (faculty, office of sponsored program employees, and institutional repository managers) were identified by university web pages and were invited to participate via email. Difficulties in selecting survey participants included having too many potential faculty identified on websites, a lack of consistent identification of institutional repository managers, and difficulty selecting the Sponsored Program staff person most able and willing to answer the survey. To deal with the faculty selection issues, a web-based random number generator tool (<http://www.random.org/>), selected four faculty from each university's STEM (Science, Technology, Engineering and Mathematics) departmental websites resulting in a list of 3,150 faculty. This selection method prevented the repeated selection of the first four listed people on faculty websites. For the 177 Sponsored Programs Offices identified via web searches, the authors selected the person they deemed most likely to be in charge of reviewing grant applications. Of the 183 institutions investigated, 63 did not have an institutional repository and so were not contacted to fill out the IR survey. Of the remaining 120 institutions, careful web searches identified people most likely to be managing institutional repositories. IR personnel were identified and contacted. In the event that an IR manager was unidentifiable, digital initiatives staff were contacted.

Survey invitation emails with a link to the anonymous online survey were sent out

in June of 2012; follow-up emails were sent out in July (IR, OSP) or November (faculty) 2012. Responses were no longer accepted after January 18, 2013. To encourage participation, the surveys invited participants to enter their names into a drawing for two \$20 gift cards (for the first 50 respondents) and five \$10 gift cards. Names for the gift card drawings were collected separately from survey data to ensure full anonymity.

Limitations of the Study

This study examines the broad landscape of reactions to data management mandates. As such, it does not address the technical aspects of a data repository's ability to adequately handle specific data file types, nor does it address specific metadata solutions to describe the dynamic nature of some datasets.

Technological issues with the surveys allowed some respondents to continue their survey even when asked to exit on the basis of a certain response. In addition, some questions did not allow the selection of multiple options as intended. The technological survey issues might have led to contamination of the responses. The survey questions were not perfectly aligned across the three surveys preventing cross-group comparisons.

The response rates of all three surveys is low, especially for the faculty survey. The Sponsored Programs survey yielded an 18% response rate (177 total; 32 returned); the faculty survey yielded an 8% response rate (3,150 total; 253 returned); and the IR survey yielded a 28% response rate (120 total; 34 returned). The unusual sampling method and the generally low response rates might lead to non-response bias of this study. A lack of external validity for the sample limits how results can be generalized. The possibility of sampling error reduces the confidence that these results accurately and fully

represent the actual populations from which these samples were taken.

Despite these limitations, the study yields interesting and useful findings.

Findings

Table 1 summarizes study results in five logical groupings while narrative sections below discuss survey responses in detail.

| |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data management |
| <ul style="list-style-type: none"> · Faculty generally are unfamiliar with institutional data management requirements · Faculty generally have not changed their data management practices because of the mandates · Many sponsored program offices now provide data management plan assistance in response to mandates |
| Sharing of data |
| <ul style="list-style-type: none"> · Most faculty are willing to share research data · Some faculty have conditions for sharing: will share after publication and/or require proper attribution · Some data cannot be shared because it is proprietary data and/or confidential. |
| Storage of data sets |
| <ul style="list-style-type: none"> · Most faculty either generate and/or use data sets · Faculty are somewhat unaware of data repositories · Sponsored programs are more aware of institutional data repositories than faculty · Faculty generally store on on local computers or external drives · Faculty who already store data locally are slowly shifting to storing in repositories · About 1 in three faculty store data in library-based, agency-sponsored, or commercial repositories · A majority of institutional repositories handle data set storage |
| Use of data sets |
| <ul style="list-style-type: none"> · Faculty <i>use</i> data sets stored in repositories more often than they themselves <i>deposit</i> data in those repositories. · The extent to which Faculty teach students to use data sets varies greatly, ranging from intensive, to mere mention of data sets, to no mention at all |
| Institutional support for data management |

| |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> · Faculty would like assistance in meeting data management requirements · Desired assistance includes tools for sharing data, assistance accessing data sets, data visualization, workshops on data management, help with data set copyright and intellectual property issues · Only some institutions offer data management support, mostly through sponsored program office and university library (storage) · The role university library and information technology departments play in data management has increased as result of mandates · A majority of universities have an institutional repository · The most common types of repository platforms hosted by universities are Digital Commons, DSpace, Fedora, or homegrown platforms |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Table 1. Summary of study findings

Faculty Respondents

The 196 faculty respondents identify their academic titles as follows: sixteen (8.2%) are non-tenure track (lecturer, research associate, adjunct professor, etc.), 28 (14.3%) are assistant professors, 48 (24.5%) are associate professors, and 90 (45.9%) are full professors. Five respondents (2.7%) identify themselves as professor emeritus and 9 (4.6 %) as “other,” which includes various ranks of research professor, a department head, some research assistants, a research scientist, and a research instrumentation director.

The survey respondents who identify their departments (194) come from many different departments (see Appendix B for a detailed breakdown). Among the largest departments represented are Chemistry (40 or 20.6%), Physics (35 or 18.0%), Computer Science (23 or 11.8%), and Geology (21 or 10.8%). These department names are self-reported (i.e. not selected from a provided list) and subsequently are grouped together by the authors for reporting purposes. Likewise the faculty survey respondents who identify their colleges (193) also represent a variety of college types (see Appendix C for a detailed breakdown). These self-reported college names show diverse naming conventions that make it challenging for the authors to group like colleges. About a third of the

respondents (81 or 45.8%) come from Colleges of Arts (Liberal Arts, Letters and Science) and 31 (17.5%) come from Colleges of Science. The remaining respondents come from variously named and organized science and engineering departments (65, 36.7%).

Effects of Data Management and Sharing Mandates

When asked in the faculty survey if universities have data management requirements, most of the faculty (183 or 91.5%) respond that their institutions either do not have specific requirements of researchers with regards to data management (123 or 61.5%) or that they are unaware of such requirements (60 or 30.0%). A small minority (17 or 8.5%) state that either plans are in place or under development, or that the university just follows contractual obligations with funding agencies. In a few cases institutions just specify how long faculty needed to hold on to data for legal reasons.

The surveys ask faculty and sponsored programs personnel¹ if there have been any changes in their work due to the new mandates by large funding agencies such as NSF and NIH with regards to research data. The new mandates have not changed data management practices for the majority of faculty respondents (82.8%, 159). This is due mainly because many of these faculty members already store and share data (56.8%, 109) and simply will continue to do so. Some faculty members store data but have never shared it and aren't planning on sharing now (21.4%, 41); an even smaller number neither store nor share data and aren't going to change that (4.7%, 9). Other respondents (25.52%, 49)² say they have changed their data management practices due to the mandates and are

¹ IR Librarians were not specifically asked whether any change due to mandates has taken place.

² Note that respondents were allowed to check multiple answers in this question so

now storing data in either a private or public place. Others who already were storing their data are now storing their data in a repository rather than privately. Of the 21 faculty members who comment about whether the mandates had changed their data management practices, six (28.57%) say they are working on changing their data management practices and expect to be mandate-compliant in the future. One respondent shares only upon request; another shares data as part of a publication or once the information has been published, while a third shares only among collaborators. The remaining faculty members have varied responses: they have nothing much by way of data to share, think mandates are “stupid”, have no change in sharing practices, or say that they like sharing. Some of the individual responses by faculty illustrate that faculty do not always see the sharing of data as useful or appropriate. One faculty member thinks that keeping track of publications is difficult enough and “sifting through raw data of other researchers would be beyond human abilities.” Another thinks “nobody would know what to do with my data if they could access it.” And finally “it would be more useful if funding agencies would be more thoughtful about which researchers and projects they require to have a data management plan.” This researcher thinks that such requirements are little more than an administrative burden.

Nineteen (59.4%) of the Sponsored Programs Office respondents say that although they are aware of these new requirements, they have not affected what their office does. However, thirteen (40.6%) say that the new mandates have affected them. Of these thirteen, five have created new policies as a result of the new requirements, five now offer a new service to help grant applicants write the required data management plan,

the percentages do not add up to 100.

and one helps researchers find places to store their research data sets. Other changes cited involve providing guidelines on how to write data management plans on the OSP website, adding an extra step to the proposal review process to check the data management plan in proposals, and sending faculty reminders about including the data management plan. One university has added additional resources for managing and storing data while another refers faculty to the library for data archiving. Since the passage of data management and access mandates, Sponsored Programs officers at least seem more aware of the existence of data storage options on campus. The new mandates did not appear to have an effect on the number of grants that were submitted (96.7%, 29).

Sharing of Research Data

A majority of faculty report a willingness to share data. A majority of respondents (74.9%, 143) say they are comfortable sharing data versus 25.1% (48) who say that they are not.³ Forty eight faculty members comment on why they do not like to share data. Some of the main reasons for not sharing are: 1) the data have not yet been published, 2) preparing data for sharing takes effort and data might not be attributed back to the original researcher, 3) data are proprietary, and 4) data are confidential.

The most commonly cited reason for not sharing data is that the data have not yet been published, leaving open the possibility that sharing might happen after publication. Sharing unpublished data might lead to others publishing the research first -- researchers compete in publishing papers and coming out first with new research is extremely important. As one respondent puts it “Unpublished, unrefereed, unverified data should be

³ Faculty answering that they did not generate or use data sets were instructed to exit the survey, but some continued answering questions.

private.” A close second reason for not sharing data is the lack of “rigorous policies regarding reciprocity, priority acknowledge[me]nts and collaboration” with regards to that data. There is a shared concern that faculty would not be cited when other researchers used their data sets and that sharing did not count for anything in their annual evaluation process. Further, when they share their data, researchers lose control over their interpretation and use. “Others would just mine my data for their own purposes, not all of which might be good purposes.” Raw data cannot be shared as is and making the data usable for others would require a substantial effort for which faculty members are not compensated and generally don’t have time. Some faculty respondents refuse to share because their data were highly proprietary or because they felt sharing data would breach the confidentiality agreement with their study participants as required by their Institutional Review Boards.

The majority of comments (79 or 55.2%) from faculty who are willing to share their data express that sharing simply makes sense. For example, many faculty say that sharing data is part of the scientific process -- it allows others to check and build upon results thereby furthering science. Sharing also makes sense because the research is paid for by federal (i.e. public) funds and should therefore be available to the public. Others say that sharing is “the right thing to do” and that sharing is often reciprocal. The rest of the comments vary. Some (12 or 8.4%) support a more calculated view of sharing wherein the sharer is benefited (“information wants to be free, and widely shared data gets widely cited”). Others (32 or 22.4%) support sharing only under certain conditions such as “provided correct recognition is attributed and abusive practices [sic] are suppressed” or “after primary research is published. A smaller number (4 or 2.8%)

mention that sharing is simply required by the funding agency. Sixteen (11.2%) offer miscellaneous answers that do not fit any category.

Storage of data sets

Four out of five faculty create or use data sets. Most faculty surveyed (80.9%; 204) either generate data sets in their research or use them for research or teaching. Forty-nine faculty (19.4%) do not generate or use data sets. Despite the finding that most faculty generate and/or use data sets, many (43%; 92) can not identify specific data repositories for storing data sets. This lack of awareness about repositories is not surprising given that most faculty 83.1% (172) say they store their data on a local computer or external hard drive.

Faculty overwhelmingly indicate that they are themselves responsible for managing stored data (74.4%, 151). When examining this responsibility in detail however a somewhat different picture emerges. Thirty-four percent (69) indicate that the project's principal investigator is responsible for managing the data, 28.6% (26) indicate that this task is carried out by a graduate student, 12.8% (26) rely on a lab manager, 2% (4) rely on a librarian and 3.4% (7) do not know. 14.8% (30) rely on some other person, such as a "computer manager," "data center manager," or "systems administrator."

Seventy-seven (37.1%) faculty respondents store data in a library-based, agency-sponsored, or commercial repository. A slight majority, however, (57%, 122) use other means of storage such as web servers, personal websites, or published complete data sets. The survey also asks how many faculty are aware of two specific archives: ArXiv and ICPSR (Inter-university Consortium for Political and Social Research). ArXiv is by far

the most familiar to faculty (46.3%; 57); ICPSR is familiar to only 2.4% (3). This lack of awareness with regard to ICPSR is not surprising, however, given that this archive is social science in nature while faculty survey respondents are from STEM departments. When asked to identify “Other” specific repositories, faculty mention about a hundred, citing some repositories multiple times (See Appendix D). Repositories mentioned multiple times are run by federal agencies such as NOAA, NASA, and Jet Propulsion Lab. For conciseness, some of the identified repositories are combined into groups in the table in Appendix D; for example, instead of individually listing each of the many NASA databases mentioned by name, these are grouped together under the single entry for “NASA data archives.”

Surprisingly, the majority of faculty (78.2%; 154) do not know if their parent university has a central repository for hosting research data. On the other hand, when Sponsored Programs staff are asked if their universities help researchers store their research data, 80% indicate that they do so, either through Central Information Technology (IT) or the Library and most of the Institutional Repository staff surveyed (60.6%, 20) indicate that their IRs handle data or data sets. While these numbers cannot be directly compared, they do seem to indicate that the majority of campuses offer repositories for research data and that many faculty are unaware of this resource.

Faculty Use of Other Researchers’ Data

When asked to describe their use of data repositories or data sets created by researchers, 56.1 % (78) of faculty indicate that they do use data sets in repositories, while only 31.7% (44) contribute data sets to them for storage. Fifty nine faculty

(42.4%) search data sets for usable information, but it is unclear how readily they are able to find this information. Forty three (31%) researchers are either not aware of data repositories in their fields or have not used them even if they are aware of them.

The survey also asks whether or not faculty incorporate data sets into their teaching. A majority (74.8%, 104) of faculty refer to or require their students to use data sets. Twenty eight percent (58) of faculty incorporate data sets in their teaching and show students how to find and use them, while 22.7% (47) only make students aware of the existence of data sets. Reasons given by faculty for not using data sets in their teaching include: they are not relevant, they are beyond the level of the class, or the faculty member does not know how to find data sets. A small number of faculty respondents (8.7%, 18) do not teach.

Institutional Support for Data Management

Seventy three percent (143) of faculty indicate that their respective universities are of no help in managing data. Of those that do offer some help, 12% (24) help their faculty write proposals, 13% (26) provide space to store data, and 5% (9) provide workshops on best practices for storing data sets. When faculty are asked whether their institutions provide a central repository for hosting data, 78.2% (154) do not know. Faculty also are asked what types of data management services they would find useful: a slight majority of faculty (58.1%; 97) indicate that they would like assistance meeting data sharing requirements; 37.7% (63) want help writing data management plans; 53.9% (90) indicate that tools for sharing research data would be helpful (53.9% (90); 50.9% (85) want assistance in finding or accessing other data; 37.7% (63) request data

visualization assistance; 29.9% (50) want workshops on data management; and 42.5% (71) want information on copyright and ownership issues associated with data sets.

Interestingly, though faculty indicate that they would like assistance meeting data management requirements, many report that their data practices have not been affected by the mandates.

Sponsored programs personnel are asked to describe the support their institutions offered for data management prior to the passage of data management mandates. Close to half (44.8%, 13) indicate that their university provided support for the data storage needs of faculty through individual campus departments, 34.5% (10) through central IT, and 17.2% (5) through the library. No respondents indicate that the sponsored programs office filled this role. Only two respondents indicate that the university did not provide data storage services while eight (27.6%) Sponsored Programs officers indicate they do not know if their university helped researchers with storing data.

With the passage of data management and access mandates, the role of central IT and the library has increased, according to respondents to the Sponsored Programs survey, while the role of individual departments has remained about the same. When asked to indicate the various places their universities now help researchers store data sets, some respondents indicate more than one place: 43.3% (13) of respondents indicate that their central IT now provides storage for data sets, 36.7% (11) indicate that the library fills this role, and 46.7% (14) indicate that individual departments fill this role. Only 3 Sponsored Programs officers say they do not know if the university now helps researchers store their data. Since the passage of data management and access mandates, Sponsored Programs officers at least seem more aware of the existence of data storage on campus.

According to institutional repository managers, most universities offer data set storage services, and usually these services are provided by campus institutional repositories. At the university level, 82% (28) provide services for hosting and preserving data sets, 15% (5) do not, and 6% (2) are not sure. And usually institutional repositories support storing faculty data sets: 61% (20) do, 33% (11) do not include data, and 6% (2) do not know. Of those that do provide such services, 19% (6) rely on Digital Commons, 44% (14) on Dspace, 16% (5) on Fedora, and 13% (4) on a home-grown platform. Three respondents do not know the type of platform used. The file sizes these repositories can accommodate varies widely. 36% (10) repositories can accommodate any size, 2 can house up to 2 gigabytes, 2 can house up to 150 megabytes, and the remainder offer a variety of responses such as 3-4 gigabytes, the availability of alternative initiatives on campus designed to accommodate large files, etc.

Discussion

Research data management and data sharing are hot topics generating proposed legislation as well as plans for a federated system of university repositories and a clearinghouse formed from existing online services. In February 2013, the *Fair Access to Science and Technology Research Act of 2013* (FASTR) was introduced simultaneously in both the House and Senate. This bill requires federal agencies with R&D budgets exceeding \$100 million annually to develop plans to ensure the “free online public access to such final peer-reviewed manuscripts or published versions” of research and data funded by federal money. While FASTR is being debated in Congress, the Executive Office of the President’s Office of Science and Technology Policy (2013) has already

taken the initiative of directing all funding agencies with greater than \$100 million in research and development expenditures to develop public access plans for federally funded research—both data and publications—within six months. This is an important step because an executive mandate is not held up by legislative debate and is effective immediately. However, passage of FASTR is also key because unlike executive orders, congressional legislation cannot be overturned by a future administration.

Meanwhile, universities and their libraries are not sitting back inactive. The Association of American Universities, Association of Public Land-grant Universities, and Association of Research Libraries (2013) jointly are planning a SHared Access Research Ecosystem (SHARE) which would be a federated system of university repositories to archive, preserve and provide open access to federally funded research results. SHARE would use the existing infrastructure of university institutional repositories to store research data and comply with the White House directive for providing public access to funded research results. Publishers have proposed another solution called the Clearinghouse for the Open Research of the United States (CHORUS) (Sporkin, 2013). CHORUS would be built using existing online services such as CrossRef and FundRef which respectively provide links between research and between research and funding.

Whatever may happen, librarians and university repository managers have the existing skillsets to play an important role in storing and making research data accessible. However, libraries must advocate their existing services and infrastructure to their faculty users, while making known the fact that they are in an appropriate position to ensure compliance with these types of mandates. If they do not other entities will. Particularly

now with this sea change of government support for opening up access to federally-funded research data, faculty need clear answers and convenient options for data management and accessibility. Conditions favor increased Library advocacy for existing and new data management infrastructure and services in view of the rapidly changing legislative landscape regarding data management and its open availability. The following section answers the research questions that guided this study.

1. What has been the nature and extent of the effect on universities and researchers of new granting agency requirements that researchers include data management plans in their grant applications? The greatest impact of new granting agency requirements for inclusion of data management plans is two-fold: (1) sponsored programs offices are providing new services to help researchers with writing data management plans and finding places to store their data and (2) many researchers are switching data storage from private to public storage. Forty percent of sponsored program respondents indicate they have created new policies and/or services in response to the mandates. And while a majority of faculty respondents indicate that they were already storing their data sets before the mandates, about a quarter of them say they are moving their data from local to publicly accessible storage in response to the mandates.

2. How are researchers storing and providing access to new data generated by their grant projects? Most faculty surveyed (3 out of 4) are willing to share their data. They feel sharing data is part of the scientific process, allowing others to check their work and build upon it. They create data sets, generally manage their own data, and store data sets locally. While some studies have found that 10% of requests for academic research data are denied (Campbell, E. G., Weissman, J. S., Causino, N., & Blumenthal,

D. (2000), others write that academics are practiced at communicating discoveries and sharing materials (Evans, 2010). What is clear from our study is that at least conceptually, faculty like the idea of sharing data. Nearly a quarter of faculty, however, do not wish to share their data for a wide variety of reasons such as concern about pending publications, confidentiality, proprietary issues, lack of proper attribution, and the additional work involved in making data accessible to others.

3. Are faculty researchers using existing data stored in repositories for their own new research and teaching? A slight majority of the faculty surveyed use the data sets of other researchers in their own research and most refer students to data sets for research purposes. Slightly less than half of faculty use data sets in their teaching, mostly by showing students how to find and use them or, at minimum, simply making students aware that data sets exist. Generally, use of data sets in teaching is limited to upper-division or graduate classes.

While the survey yields a diverse and interesting list of specific data repositories (see Appendix D) from faculty that use them, it also indicates that many faculty are not aware of the data repositories in their field, tend not to use them, or have trouble naming them. When they do use repositories, the use is more often to find and use the data sets of others rather than to contribute their own data sets.

4. Where in universities are research data sets being stored; what entities are storing data and what entities are paying for this storage? Most faculty store their data locally on their hard drive or an external drive and appear to be unaware of whether their university or library offers storage facilities. However, results of this study indicate that some faculty are using public storage. Migrating data from local storage to institutional

repositories and preserving data for later use allows for replicability of results and potentially increases visibility of research. A study by Connell (2011) of institutional repository use at Ohio State University reveals data as the third most used type of material after journal articles and undergraduate theses. To play a key role in the data storage movement librarians and digital repository managers need to make their presence and services known to their faculty colleagues. For example, to increase faculty participation in the institutional repositories, librarians can point to research that demonstrates that articles that are freely available online, either in open-access journals or as pre-prints in institutional repositories receive more citations than articles that are harder to come by (e.g. Kurtz, Eichhorn, Accomazzi, Grant, Demleitner, & Henneken et al., 2005; Hajjem, Harnad, & Gingras, 2005).

In fact, a number of new opportunities are emerging for academic librarians (Newton, 2010). This study identifies several faculty data management needs that librarians can fill: creating or making available tools for sharing research data; assistance with finding or accessing research data; providing information on copyright and ownership issues associated with data sets, and helping faculty write data management plans. Further, to help faculty in finding the data sets of other researchers, librarians can help develop and implement metadata standards for data sets and help faculty effectively describe contributed data sets so that they are shareable, findable, and usable. And to mitigate the complaint by faculty that their data are not properly attributed, librarians can ensure that each stored data set is accompanied by a recommended citation that users can copy and paste to facilitate proper attribution. These additional services, brought on by the funding agency mandates, extend the traditional role of libraries and demand a more

active collaboration between librarians and faculty members with regards to research data. With this extended role academic librarians are uniquely positioned to become even more important partners in the scholarly communication process.

Conclusion and Future Research

This research suggests that faculty are indeed interested in sharing their data and using the data of others. They may not always agree with the data management mandates, but tend to see value in the purpose. This study reveals, however, that while faculty are interested in managing and sharing data, they do not think that they have the necessary infrastructure or skillset for accomplishing these tasks. As many as 80% of surveyed faculty do not know if their parent institutions have repositories to manage data. Moreover, a majority of faculty and sponsored programs officers reported that data management mandates have not impacted their workflows. It seems, then, that while respondents recognized the need to share data, there may be some confusion and perhaps resulting paralysis in how to proceed.

This study reveals an important knowledge gap that libraries appear to be well positioned to fill. In many cases faculty desires for infrastructure or skills identified in this study are already being provided by libraries, but the faculty, and to a lesser extent, sponsored programs officers, are not aware of these services. Librarians and IR professionals have an important role to play in helping faculty comply with funder mandates, but they need to do a better job of making faculty aware of existing data management infrastructure and related services. While librarians may see the library as well positioned to take on these new roles, faculty, at this stage, do not. The faculty who continue to store their data on personal computers and external hard drives need

convincing that the library is a logical partner in the storage and description of their research data.

Further research is needed to explore effective ways for librarians and faculty to collaborate in managing, disseminating, and preserving research data. A logical starting place for this further work is an existing, successful faculty-librarian collaboration -- information literacy instruction. Librarians partner with faculty by sharing their specialized knowledge and expertise in information literacy using collaboration models in varying levels of intensity (Pritchard, 2010). These collaboration models might be adapted for data management. For example, a high-level of collaboration found in information literacy instruction is embedding librarians in university classes. Transferring this successful model to data management might involve embedding librarians into research centers or individual research projects where they can directly share their expertise and eventually create institutionalized partnerships.

References

- Alvaro, E., Brooks, H., Ham, M., Poegel, S., & Rosencrans, S. (2011). E-Science Librarianship: Field Undefined. *Issues in Science and Technology Librarianship*, (66), 28-43. Available online: <http://www.istl.org/11-summer/article1.html>
- Association of American Universities, Association of Public Land-grant Universities, & Association of Research Libraries (2013). SHared Access Research Ecosystem (SHARE) (Draft, June 7, 2013). Available online: <http://www.arl.org/storage/documents/publications/share-proposal-07june13.pdf>
- Babbie, E. R. (1973). *Survey research methods*. Belmont, Calif.: Wadsworth Pub. Co.
- Campbell, E. G., Weissman, J. S., Causino, N., & Blumenthal, D. (2000). Data withholding in academic medicine: characteristics of faculty denied access to research results and biomaterials. *Research Policy*, 29(2), 303-312. Available online: <http://www.sciencedirect.com/science/article/pii/S0048733399000682>
- Carlson, J., Ramsey, A. E., & Kotterman, J. D. (2010). Using an institutional repository to address local-scale needs: a case study at Purdue University. *Library Hi Tech*, 28(1), 152-173. Available online: <http://www.emeraldinsight.com/journals.htm?articleid=1846230&show=abstract>
- Connell, T. H. (2011). The Use of Institutional Repositories: The Ohio State University Experience. [Case Study]. *College & Research Libraries*, 72, 253-274.
- Delserone, L. M. (2008). At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries. *Library Trends*, 57(2), 202-210. Available online:

<https://www.ideals.illinois.edu/bitstream/handle/2142/10670/delserone.pdf?sequence=2>

Dietrich, D., Adamus, T., Miner, A., & Steinhart, G. (2012). De-Mystifying the Data

Management Requirements of Research Funders. *Issues in Science and Technology*

Librarianship, (70). Available online: <http://www.isitl.org/12-summer/refereed1.html>

Evans, J. A. (2010). Industry collaboration, scientific sharing, and the dissemination of

knowledge. *Social Studies of Science*, 40(5), 757-791. Available online:

<http://sss.sagepub.com/content/40/5/757.full.pdf+html>

Faily, S., & Fléchais, I. (2010). Designing and aligning e-Science security culture with design.

Information Management & Computer Security, 18(5), 339-349 Available online:

<http://www.emeraldinsight.com/journals.htm?articleid=1896385>

Fair Access to Science and Technology Research Act of 2013. H.R. 708 and S. 350, 113th Cong.

(2013). Available online: <http://thomas.loc.gov/cgi-bin/query/z?c113:H.R.708>:

Fischer, E. A. (2013). Public Access to Data from Federally Funded Research: Provisions in

OMB Circular A-110. Washington, DC: Library of Congress. Congressional Research

Service. Available online: <https://www.fas.org/sgp/crs/secretary/R42983.pdf>

Gold, A. (2007a). Cyberinfrastructure, Data, and Libraries, Part 1: A

Cyberinfrastructure Primer for Librarians. *D-Lib Magazine*, 13(9/10). Available online:

<http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>

Gold, A. (2007b). Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and

the Data Challenge: Roles and Actions for Libraries. *D-Lib Magazine*,

13(9/10). Available online: <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>

Gore, S. A. (2011). e-Science and Data Management Resources on the Web. *Medical Reference*

Services Quarterly, 30(2), 167-177. Available online:

<http://www.ncbi.nlm.nih.gov/pubmed/21534116>

- Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten Year Cross-disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin*, 28(4), 39-47.
- Hey, T., & Hey, J. (2006). e-Science and its implications for the library community. *Library Hi Tech*, 24(4), 515-528. Available online:
<http://www.emeraldinsight.com/journals.htm?articleid=1583887>
- Hswe, P., & Holt, A. (2011). Joining in the Enterprise of Response in the Wake of the NSF Data Management Planning Requirement. *Research Library Issues* (274), 12-17. Available online: <http://publications.arl.org/rli274/12>
- Jones, E. (2008). E-Science Talking Points for ARL Deans and Directors. Washington, DC: Association of Research Libraries. Available online:
<http://www.arl.org/storage/documents/publications/e-science-talking-points.pdf>
- Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., et al. (2005). The effect of use and access on citations. *Information Processing & Management*, 41(6), 1395-1402.
- Liotta, L. A., Lowenthal, M., Mehta, A., Conrads, T. P., Veenstra, T. D., Fishman, D. A., & Petricoin, E. F. (2005). Importance of Communication Between Producers and Consumers of Publicly Available Experimental Data. *Journal of the National Cancer Institute*, 97(4), 310-314. Available online:
<http://jnci.oxfordjournals.org/content/97/4/310.abstract>
- Luce, R. E. (2012). Grand Challenges and New Roles for the Twenty-First-Century Research Library in and Era of E-Science. In D. B. Marcum & G.

- George (Eds.), *The Data Deluge: Can Libraries Cope with e-Science?* (pp. 3-15). Santa Barbara, Calif.: Libraries Unlimited.
- Macdonald, S., & Martinez-Uribe, L. (2010). Collaboration to Data Curation: Harnessing Institutional Expertise. *New Review of Academic Librarianship*, 16(S1). Available online: <http://www.tandfonline.com/doi/abs/10.1080/13614533.2010.505823#.Ugqv6mRFxJE>
- Macdonald, S., & Martinez-Uribe, L. (2009). User Engagement in Research Data Curation. In Agosti, M, Borbinha, J, Kapidakis, S., Papatheodorou, C., & Tsakonas, G. (Eds.), *Research and Advanced Technology for Digital Libraries* (pp. 309-314). New York: Springer. Available online: http://link.springer.com/chapter/10.1007%2F978-3-642-04346-8_30
- Marcum, D. B., & George, G. (Eds.). (2012). *The Data Deluge: Can Libraries Cope with e-Science?* Santa Barbara, Calif.: Libraries Unlimited.
- Mullins, J. L. (2009). Bringing Librarianship to E-Science. *College & Research Libraries*, 70(3), 212-213. Available online: <http://crl.acrl.org/content/70/3/212.full.pdf+html>
- National Science Foundation. (2011). Changing the Conduct of Science in the Information Age: Summary Report of Workshop Held on November 12, 2010 National Science Foundation. Arlington, VA: National Science Foundation. Available online: <http://www.nsf.gov/pubs/2011/oise11003/oise11003.pdf>
- Newton, M. P., Miller, C. C., & Bracke, M. S. (2010). Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force. *Collection Management*, 36(1), 53-67. Available online: <http://www.tandfonline.com/doi/abs/10.1080/01462679.2011.530546#.Ugq2R2RFxJE>
- Pritchard, P. A. (2010). The Embedded Science Librarian: Partner in Curriculum Design

and Delivery. *Journal of Library Administration*, 50(4), 373-396. doi:

10.1080/01930821003667054

Qin, J., & D'Ignazio, J. (2010). The Central Role of Metadata in a Science Data Literacy Course.

Journal of Library Metadata, 10(2/3), 188-204. Available online:

<http://www.tandfonline.com/doi/abs/10.1080/19386389.2010.506379#.Ugq3e2RFxJE>

Ramírez, M. L. (2011). Whose role is it anyway?: A library practitioner's appraisal of the digital

data deluge. *Bulletin of the American Society for Information Science*, 37(5), 21-23.

Available online: http://works.bepress.com/marisa_ramirez/20/

Soehner, C., Steeves, C., Ward, J. (2010). E-Science and Data Support Services: A Study of

ARL Member Institutions. Washington, D.C.: Association of Research Libraries.

Retrieved from <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf>

Sporkin, Andi (2013). *Understanding CHORUS*. Available online:

<http://www.publishers.org/press/107/>

Steinhart, G. (2010). *DataStaR: A Data Staging Repository to Support Sharing*

and Publication of Research Data. Paper presented at the International

Association of Scientific and Technological University Libraries, 31st

Annual Conference. Available online:

<http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1012&context=iatul2010>

United States. Executive Office of the President. Office of Science and Technology

Policy. (2013) *Memorandum For the Heads of Executive Departments and*

Agencies: Increasing Access to the Results of Federally Funded Scientific

Research. Available online:

http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Witt, M. (2012). Co-designing, Co-developing, and Co-implementing an Institutional Data Repository Service. *Journal of Library Administration*, 52(2), 172-188. Available online: *<http://www.tandfonline.com/doi/abs/10.1080/01930826.2012.655607#.UgrGDBaRP8s>*

Appendix A

Survey Questions

Faculty Survey:

1. Do you generate data sets in your research or use existing data sets in your research or teaching?
2. Are you aware of any data repositories where people can access data sets created by researchers?
3. If you answered YES to the previous question, which of the following data repositories are you familiar with (please check all that apply)?
4. Please describe your use of data repositories containing data sets created by researchers (please check all that apply)
5. Do you incorporate existing data sets into your teaching?
6. Please describe where you store your research data (please check all that apply)
7. Who is responsible for managing the stored data? (please check all that apply)
8. Have data management mandates from funding agencies changed the way you manage, store, and share your research data? Please check all that apply.
9. Does your institution ask researchers to meet specific requirements related to research data management (e.g. data must be archived in a permanent repository, data must be available to the public)?
10. Does your institution provide help to researchers with managing data? If so, please describe?
11. Does your institution provide a central repository for hosting research data?
12. Which of the following services do you/would you find useful? (please check all that apply)
13. Are you comfortable sharing your data sets with other researchers?
14. What is your academic title?
15. What is the (generic) name of your college or school (e.g. College of Science, College of Engineering)?
16. What is the generic name of your department (e.g. Animal Science)?

Institutional Repository Survey:

1. Does your university provide services for hosting and preserving data sets generated by grant-funded research?
2. If "Yes" to Question One, what platform is used to host data sets?
3. Does your university offer an institutional repository (IR) for faculty publications and research?
4. If your university has an IR, does it include research data or data sets?
5. If "Yes" to number three, what size files can your repository accommodate?
6. Does your university assist grant applicants in writing data management plans?
7. If someone at the library assists researchers with data management plans, that person is:

8. Do you think the library should be involved in data management planning? Please elaborate.
9. At what point in the process of data management planning should the library become involved?
10. Do you tailor data management plans to fit individual people or departments or does your institution take a one size fits all approach?
11. Assuming your library is involved in working with departments or researchers to develop data management plans, are your subject librarians/liaisons involved in these negotiations?

Sponsored Programs Survey:

1. Have the recent requirements by funding agencies (NSF and NIH) that grant applications must include a data management plan affected your Sponsored Programs Office?
2. If you answered "Yes" to the preceding question, please indicate how your Sponsored Programs Office has been affected by new grant requirements for a data management plan. (Please check all that apply).
3. Before the new requirements, did your university help researchers store their research data? How? (Please check all that apply)
4. Does your university now help researchers store their research data? How? (Please check all that apply)
5. To what extent has the NSF or NIH mandate affected the number of grants submitted to you?
6. What tools, policies, or training has your university's Sponsored Programs Office developed to assist grant applicants with writing a data management plan? (Please check all that apply)

Appendix B
Faculty Respondents' Departments

| Department | # | % |
|--------------------------------------------------------|----------|----------|
| Chemistry | 40 | 20.6 |
| Physics | 35 | 18 |
| Computer Science | 23 | 11.9 |
| Geology | 21 | 10.8 |
| Physics and Astronomy | 13 | 6.7 |
| Chemistry & Biochemistry | 10 | 5.2 |
| Earth Sciences | 10 | 5.2 |
| Mathematics | 10 | 5.2 |
| Environmental Science | 3 | 1.5 |
| Biology | 2 | 1 |
| Computer Science and Engineering | 2 | 1 |
| Geology and Geophysics | 2 | 1 |
| Geosciences | 2 | 1 |
| Mathematics and Computer Science | 2 | 1 |
| Atmospheric Science | 1 | 0.5 |
| Biochemistry; Chemistry; Center for Structural Biology | 1 | 0.5 |
| Biological & Physical Sciences | 1 | 0.5 |
| Biological Sciences | 1 | 0.5 |
| Biomedical Sciences | 1 | 0.5 |
| Chemistry and Chemical Biology | 1 | 0.5 |
| Communication | 1 | 0.5 |
| Computational Sciences | 1 | 0.5 |
| Electrical and Computer Engineering | 1 | 0.5 |

| | | |
|---------------------------------------------------------|-----|------|
| Electrical Engineering and Computer Science | 1 | 0.5 |
| Geology and Planetary Science | 1 | 0.5 |
| Marine Geology and Geophysics | 1 | 0.5 |
| Marine Science | 1 | 0.5 |
| Materials Science and Engineering | 1 | 0.5 |
| Physics, Chemistry and Mathematics | 1 | 0.5 |
| Planning | 1 | 0.5 |
| Radiology | 1 | 0.5 |
| School of Physics, Astronomy and Computational Sciences | 1 | 0.5 |
| Statistics | 1 | 0.5 |
| | 194 | 99.6 |

Appendix C
Faculty Respondents' Colleges

| College | # | % |
|---------------------------------------------------------|----------|----------|
| Arts & Sciences | 61 | 34.5 |
| College of Science | 31 | 17.5 |
| Engineering | 22 | 12.4 |
| College of Liberal Arts & Sciences | 14 | 7.9 |
| College of Letters and Science | 6 | 3.4 |
| Liberal Arts | 6 | 3.4 |
| College of Mathematical and Natural Sciences | 4 | 2.3 |
| College of Natural Science(s) | 3 | 1.7 |
| College of Physical and Mathematical Science | 3 | 1.7 |
| College of Engineering and Physical Sciences | 2 | 1.1 |
| College of Engineering and Science | 2 | 1.1 |
| College of Engineering, Forestry, and Natural Sciences | 2 | 1.1 |
| College of Medicine | 2 | 1.1 |
| Natural Sciences | 2 | 1.1 |
| College of Agriculture and Life Sciences | 1 | 0.6 |
| College of Applied Science | 1 | 0.6 |
| College of Earth, Ocean, and Atmospheric Sciences | 1 | 0.6 |
| College of Engineering and Architecture | 1 | 0.6 |
| College of Engineering Technology and Physical Sciences | 1 | 0.6 |
| College of Environment | 1 | 0.6 |
| College of Environment & Life Sciences | 1 | 0.6 |
| College of Science & Letters | 1 | 0.6 |

| | | |
|------------------------------------------|-----|-------|
| College of Science and Engineering | 1 | 0.6 |
| College of Science and Liberal Arts | 1 | 0.6 |
| College of Sciences and Humanities | 1 | 0.6 |
| Engineering and Computing | 1 | 0.6 |
| Mathematics | 1 | 0.6 |
| Natural and Environmental Science | 1 | 0.6 |
| Research Institute | 1 | 0.6 |
| School of Geoscience | 1 | 0.6 |
| School of Marine and Atmospheric Science | 1 | 0.6 |
| | 177 | 100.5 |

Appendix D
Repositories listed by faculty in Faculty Survey Question 3

| Name of Database/Repository | Times Cited | URL* |
|------------------------------------------------------------------------------------------------------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------|
| American Chemical Society Supporting Information | 1 | http://www.cas.org |
| Antarctic Marine Geology Research Facility at Florida State Univ | 1 | http://www.arf.fsu.edu |
| ArXiv (e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics) | 57 | http://arxiv.org |
| Astrophysics Data System ADS "(more important than Arxiv") | 2 | http://www.adsabs.harvard.edu |
| BEG Core Repository (Bureau of Economic Geology) | 1 | http://www.beg.utexas.edu/index.php |
| BIND (Biomolecular Interaction Network Database) | 2 | http://www.ncbi.nlm.nih.gov/pmc/articles/PMC165503/ |
| Biological Magnetic Resonance Data Bank | 1 | http://www.bmrw.wisc.edu |
| BRENDA (The Comprehensive | 1 | http://www.brenda-enzymes.info |

| | | |
|----------------------------------------------------------------------|---|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Enzyme Information System) | | |
| CADC (Canadian Astronomy Data Centre) | 1 | http://www1.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/cadc/ |
| CADIA | 1 | unknown url |
| CCDC (Cambridge Crystallographic Data Centre) | 3 | http://www.ccdc.cam.ac.uk/pages/Home.aspx |
| CSD (Cambridge Structural Database) | 1 | http://www.ccdc.cam.ac.uk/Solutions/CSDSystem/Pages/CSD.aspx |
| CAPS (Center for Analysis and Prediction of Storms) Univ of Oklahoma | 1 | http://www.caps.ou.edu |
| CEDAR Madrigal databases | 1 | http://www.openmadrigal.org |
| ChemiFinder (American Chem Soc Data Base) | 1 | unknown url |
| CISL (Computational & Information Systems Lab) Research Data Archive | 1 | http://rda.ucar.edu/ |
| CMU Motion Capture Database (Motor?) | 1 | http://mocap.cs.cmu.edu |
| CT Colonography data -- maybe the Cancer Imaging Archive? | 1 | https://wiki.cancerimagingarchive.net/display/Public/CT+Colonograph |
| CZO (Critical Zone Observatories) Data | 1 | http://criticalzone.org/national/data/access-czo-data-1-national/ |
| DAAC (Distributed Active Archive) | 1 | http://nsidc.org/daac/about/ |

| | | |
|-----------------------------------------------------------------------------------------|---|---------------------------------------------------------------------------------------------------------------------|
| Center) National Snow & Ice Data Ctr | | |
| Dryad - international repository of data underlying scientific and medical publications | 1 | http://datadryad.org |
| Durham HepData | 2 | http://durpdg.dur.ac.uk/ |
| Earthchem | 2 | http://www.earthchem.org/data/contribute |
| ELRA -- maybe European Language Resources Association? | 1 | not sure; maybe http://catalogue.elra.info |
| Elsevier Supporting Information | 1 | http://www.elsevier.com |
| ESA (Ecological Society of America) data archives | 2 | http://data.esa.org/esa/style/skins/esa/index.jsp |
| ESRL (Earth System Research Laboratory) | 1 | http://www.esrl.noaa.gov/psd/data/ |
| EuPathDB (Eukaryotic Pathogen Database Resources) | 1 | http://eupathdb.org/eupathdb/ |
| Genome projects (<i>perhaps projects like 1000 Genomes, etc.</i>) | 1 | http://www.1000genomes.org/data etc. |
| GEO (Gene Expression Omnibus) | 1 | http://www.ncbi.nlm.nih.gov/geo/ |
| GEOCHRON "data from geochronology & thermochronology communities") | 1 | http://www.geochron.org/about.php |
| GeoMapApp at Columbia University | 1 | http://www.geomapapp.org |
| GeoRoc (Geochemistry of | 1 | http://georoc.mpch-mainz.gwdg.de/georoc/Entry.html |

| | | |
|---------------------------------------------------------------------|---|-------------------------------------------------------------------------------------------------------------------------|
| Rocks of the Oceans and Continents) | | |
| GERM (Geochemical Earth Reference Model) | 1 | http://earthref.org/GERM/ |
| Global Charcoal Database | 1 | http://www.ncdc.noaa.gov/paleo/impd/gcd.html |
| Hawaii Ocean Time Series Oceanographic Data | 1 | http://www.soest.hawaii.edu/HOT_WOCE/index.php |
| Human Metabolomic Data Bank | 1 | http://www.hmdb.ca/ |
| HyperLEDA | 1 | http://leda.univ-lyon1.fr/ |
| ICPSR | 3 | http://www.icpsr.umich.edu/icpsrweb/landing.jsp |
| IEDA (Integrated Earth Data Applications) | 1 | http://www.iedadata.org/ |
| International Virtual Observatory | 1 | http://www.ivoa.net/ |
| IRIS (Industrial Research and Development Information System) | 4 | http://www.nsf.gov/statistics/iris/ |
| IRSA (Infrared Science Archive) | 1 | http://www.irsa.com.ar/irsa/index_eni.htm |
| ISIS (International Species Information System) | 1 | http://www2.isis.org/AboutISIS/Pages/About-ISIS.aspx |
| Jet Propulsion Lab Physical Oceanography (PO), etc. | 2 | http://podaac.jpl.nasa.gov/dataaccess |
| KEEL (Knowledge Extraction based on Evolutionary Learning) data set | 1 | http://sci2s.ugr.es/keel/data sets.php |

| repository | | |
|------------------------------------------------------------|---|-----------------------------------------------------------------------------------------------|
| Linguistic Data Consortium | 1 | http://www ldc.upenn.edu/About/ |
| LTER (Long Term Ecological Research Network) NSF sponsored | 2 | http://www.lternet.edu/ |
| MagIC Database (magnetic data, UC San Diego) | 4 | http://earthref.org/MAGIC/ |
| MAST (Mikulski Archive for Space Telescopes) | 1 | http://archive.stsci.edu |
| Metlin (Metabolite and Tandem MS Database) | 1 | http://metlin.scripps.edu/index.php |
| MetPetDB (Database for Metamorphic Petrology) | 1 | http://metpetdb.rpi.edu/metpetweb/#home |
| MGDS (Marine Geoscience Data System) | 1 | http://www.marine-geo.org/partners.php |
| NASA Ames PAH IR Spectroscopy Database | 1 | http://www.astrochem.org/pahdb/ |
| NASA data archives (Kepler and many more) | 6 | http://data.nasa.gov |
| National Geophysical Data Center NGDC; NGDC-Snotel | 2 | http://www.ngdc.noaa.gov |
| National Nuclear Data Center NNDC | 1 | http://www.nndc.bnl.gov/ |
| National Virtual Observatory | 1 | http://www.us-vo.org/ |

| | | |
|------------------------------------------------------------------------------|---|---------------------------------------------------------------------------------------------------------------------|
| NAVDAT (North American Volcanic and Intrusive Rock Database) | 2 | http://www.navdat.org/ |
| NCBI (National Center for Biotechnology Information) databases | 4 | http://www.ncbi.nlm.nih.gov |
| NDBC (National Data Buoy Center) NOAA sponsored | 2 | http://www.ndbc.noaa.gov/ |
| NGDS (National Geothermal Data Systems) | 1 | http://geothermaldata.org |
| NIST (National Institute of Statistics & Technology Standard Reference Data) | 1 | http://www.nist.gov/srd/index.cfm |
| NNDC (National Nuclear Data Center) Brookhaven National Lab | 1 | http://www.nndc.bnl.gov |
| NOAA Tides & Currents | 1 | http://tidesandcurrents.noaa.gov |
| NOAA's International Multiproxy Paleofire Database | 1 | http://www.ncdc.noaa.gov/paleo/impd/paleofire.html |
| NODC/NOAA (National Oceanographic Data Center) | 4 | http://www.nodc.noaa.gov/access/index.html |
| Nomads (NOAA National Operational Model Archive & Distribution System) | 1 | http://ncdc.noaa.gov/ |
| NSF databases | 1 | http://www.nsf.gov/statistics/database.cfm |

| | | |
|-----------------------------------------------------------------------------|---|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| OpenTopography for LiDAR data | 1 | http://www.opentopography.org |
| Paleobiology Database | 1 | http://paleodb.org |
| PDG (Particle Data Group) | 1 | http://pdg.lbl.gov |
| PetDB, the Petrological Database | 3 | http://www.earthchem.org/petdb |
| Protein Data Bank (PDB) | 6 | http://www.wwpdb.org |
| PUBCHEM | 1 | http://pubchem.ncbi.nlm.nih.gov |
| PUBMED & other Nat Lib of Medicine databases | 2 | http://wwwcf2.nlm.nih.gov/nlm_eresources/eresources/search_databases |
| Royal Society of Chemistry Supporting Information | 1 | http://www.rsc.org/Publishing/CurrentAwareness/About_Databases.aspx |
| SAID (Scattering Analysis Interaction Dialin) at Center for Nuclear Studies | 1 | http://gwdac.phys.gwu.edu |
| SIMBAD Astronomical Database | 2 | http://simbad.u-strasbg.fr/simbad/ |
| Sloan Digital Sky Survey | 2 | http://www.sdss.org/ |
| SOFIA-USGS South Florida | 1 | http://sofia.usgs.gov/exchange/index.php |
| Talk Bank | 1 | http://talkbank.org/ |
| Tree of Life | 1 | http://itol.embl.de/index.shtml |
| UC Irvine Machine Learning Repository | 2 | http://archive.ics.uci.edu/ml/ |
| UNAVCO for | 1 | http://www.unavco.org/crosscutting/cc-data.html |

| | | |
|-----------------------------------------------------|---|-------------------------------------------------------------------------------------------------------------------------|
| geodetic and InSAR data | | |
| USGS-NWIS (National Water Information System) | 2 | http://qwwwebsiteservices.usgs.gov |
| VizieR (Astronomy catalogs and data tables archive) | 2 | http://vizier.cfa.harvard.edu/viz-bin/VizieR |
| WFCAM Science Archive | 1 | http://surveys.roe.ac.uk/wsa/ |
| Wiley Supporting Information | 1 | http://www.wiley.com/WileyCDA/Section/id-301546.html |
| World Data Centers (Paleo, etc.) -- many of these | 2 | |

** URLs provided by study authors, not faculty respondents*