

Improving University Research Value: A Case Study

SAGE Open
July-September 2012: 1–13
© The Author(s) 2012
DOI: 10.1177/2158244012452576
<http://sgo.sagepub.com>


Kelley O'Reilly¹, Jeffrey Johnson²,
and Georgiann Sanborn³

Abstract

This article investigates the current data management practices of university researchers at an Intermountain West land-grant research university in the United States. Key findings suggest that researchers are primarily focused on the collection and housing of research data. However, additional research value exists within the other life cycle stages for research data—specifically in the stages of delivery and maintenance. These stages are where most new demands and requirements exist for data management plans and policies that are conditional for external grant funding; therefore, these findings expose a “gap” in current research practice. These findings should be of interest to academics and practitioners alike as findings highlight key management gaps in the life cycle of research data. This study also suggests a course of action for academic institutions to coalesce campus-wide assets to assist researchers in improving research value.

Keywords

university research, research data management, data science, research funding

Introduction

Research grant funding rules and policies are changing world-wide to include the adoption of performance-based research funding models: the use of bibliometric means to assess research quality, and requirements affecting the data management practices of university researchers (Nicholls & Cargill, 2011). In 2007, 22 U.S. government agencies, including the National Science Foundation (NSF); *National Aeronautics and Space Administration* (NASA); the Departments of Energy, Agriculture, and Health and Human Services; and other government branches, including the Office of Science and Technology Policy, formed an interdepartmental group called the Interagency Working Group on Digital Data (IWGDD). The group's purpose is

to develop and promote the implementation of a strategic plan for the Federal government to cultivate an open interoperable framework to ensure reliable preservation and effective access to digital data for research, development, and education in science, technology, and engineering” (IWGDD, 2007, p. 3)

The aim is for the whole of U.S. science to achieve data access and sharing similar to what is currently enjoyed by genome researchers via GenBank, or astronomers via the National Virtual Observatory.

As means of example, one of the functions of NSF (2011) is “to provide a central clearinghouse for the collection,

interpretation, and analysis of data on scientific and engineering resources.” In line with this, an NSF-wide requirement, implemented in January 2011, requires researchers to submit a data management plan with each grant proposal submitted. Similarly, the National Institutes of Health (NIH, 2003) requires that any grant request for US\$500,000 or more specifically address a data-sharing plan as part of the proposal for consideration.

The United States is not alone in working to implement new strategy, standards, and controls concerning publicly funded research data. In Canada, Australia, the United Kingdom (e.g., Digital Curation Centre [DCC], Joint Information Systems Committee [JISC], and UKOLN, 2009), and Europe, developments in data management (or data science) are occurring (Fear, 2011; Jones, 2012; Swan & Brown, 2008). These data developments are rooted in the notion that data created from research are valuable resources that can be used and reused for future scientific and educational purposes, seek to “maximize research funding,” and signal a new era focused on research quality rather than research quantity (Nicholls & Cargill, 2011).

¹Western Michigan University, Kalamazoo, USA

²Utah State University, Logan, USA

³Database Consultant, Kalamazoo, MI, USA

Corresponding Author:

Kelley O'Reilly, Department of Marketing, Haworth College of Business,
Western Michigan University, 3141 Schneider Hall, Kalamazoo,
MI 49008-5430, USA
Email: kelly.oreilly@wmich.edu

Although these are recent efforts, in the United States, significant changes regarding access to research data were beginning to occur as early as 1999 when Congress proposed an amendment to the Office of Management and Budget (OMB). The amendment, Circular A-110, governs administrative requirements for grants with institutions of higher education, hospitals, and other nonprofit organizations ensuring that data produced under an award will be made available to the public through the procedures established under the Freedom of Information Act (OMB, 1999). Interestingly, the Freedom of Information Act permits access to data first produced in a project supported *partially or entirely* by a federal grant or cooperative agreement. Although unrestricted access to data is only available when in support of a federal regulation or other action that has the force and effect of law, A-110 does not distinguish between cited studies that may be “marginally significant” and data from critical research findings, thereby potentially creating significant “burden to investigators, institutions, and Freedom of Information Act Offices if they are to be responsive” (Miller & Baldwin, 2001, p. 824).

The efforts and vision of the IWGDD illustrate the data management challenges currently faced by researchers if they desire to secure public funding. These external regulatory pressures insinuate new evolving standards and protocols for compliance and data management capabilities that may be lacking for many in the U.S. research community. In addition, a focus on data management practices indicates that grant-funded research must *create value* beyond journal publications through the access and preservation of research data. It is this notion of research value that this exploratory study investigates. Namely, this study postulates that an inextricable nexus exists between data management and the creation of long-term research value. Yet certain gaps exist in the current approach researchers take concerning data management, consequently creating problematic impacts to research value.

Based on data collected via an online survey as well as secondary data, this study investigates the current data management practices of university researchers at an Intermountain West land-grant research university in the United States. This exploratory study highlights key management gaps in the life cycle of research data and suggests a course of action for the academic institutions to coalesce campus-wide assets to assist researchers in improving research value through increased funding and improved research data management execution.

Background

Although industry performs the largest proportion of R&D within the United States, higher education institutions are the dominant player for publicly funded research. Although public universities are, with a few exceptions, funded at the state level, the majority of research funding is from federal

sources (Organisation for Economic Co-operation and Development [OECD], 2008). In 2003, the distribution of research expenditures across sectors was higher education (55%), government (19%), business enterprise (16%), and private nonprofit (11%; Vincent-Lancrin, 2006). Of the research conducted by colleges and universities, totaling more than US\$30 billion in 2000, 58.2% was funded by the federal government, 7.3% by state and local governments, 7.2% by industry, 19.7% from institutional funds, and the remainder from private funding (OECD, 2008).

Without effective data management planning, practices, and skills, many researchers may find themselves unable to secure funding, or to extend research funding for important work in progress. In addition, without effective data management, the potential for indecipherable data, data integrity issues, or lost knowledge is increased, creating implications for the policy decisions and actions taken based on research findings (Carlsen, 2006; Henty, Weaver, Bradbury, & Porter, 2008; Hopwood, 2008; Jones, Ball, & Ekmekcioglu, 2008; Lyon, 2007; Ma & Pearson, 2005; Mullins, 2006; Swan & Brown, 2008). Although previous research has examined the extent and contribution of academic research funding by external grants (e.g., see Hall, 2004; Harman, 2001, 2002, 2003), little research has been conducted on the data management practices of investigators conducting the research and how these practices align with the recent developments by IWGDD and other funding agencies (as an exception, see Fear, 2011). For these reasons, a better understanding of the current data management practices of university-based researchers is needed.

For purposes of this study, we define data management as an overarching term that refers to *all aspects of creating, housing, delivering, maintaining, and retiring data*. This definition is adapted from *Information Management Magazine*, a leading publication for business intelligence, analytics, integration, and data warehousing (Information Management, 2011). Data management typically addresses the creation of data architecture and is inclusive of the infrastructure, personnel, processes, and other requirements for identifying, consolidating, and optimizing data assets for efficiency and usefulness. However, the stages of creating, housing, delivering, maintaining, and retiring data represent the natural life cycle of data as shown in Figure 1, whereas the administration of the data architecture, infrastructure, personnel, processes, and other requirements represents the actual data management function that should include raw data sets as well as data that have been analyzed and normalized in a working data set. In this case study, we are focused on the administration of the data management function juxtaposed with the natural data life cycle.

Study Purpose and Research Questions

With an online survey and secondary data, this exploratory work investigates the current data management practices of



Figure 1. Simplified view of data management life cycle stages

university researchers at an Intermountain West land-grant research university in the United States. The study examines researchers' current data management practices and juxtaposes these current behaviors with a holistic view of the natural data life cycle in an attempt to identify gaps and improve long-term research value that aligns with the changing regulatory and funding environments.

Literature Review

The Changing Regulatory and Funding Environments

The United States has no single agency dominant with respect to research (OECD, 2008). Nor are there explicitly stated goals for publicly funded research. Although the President and Congress certainly provide direction (typically through major budget allocations), multiple goals emerge. Considering that each administration has its own agenda, two conditions have a high probability of impact on researchers and funding: (a) a significant funding shift under the Obama administration as a result of (b) a global economic financial crisis.

Adjusting to new "top-down" research priorities as set by each administration, as well as proposed changes to data management and access, may be challenging for U.S. researchers. The OECD (2008) suggests that "*There is a need to achieve better national level co-ordination in research policy and practice*" (p. 16). In the United Kingdom, we can glimpse what a national research policy may entail. Liz Lyon (2007), in her influential Consultancy Report for the JISC, suggested that

funding organizations should openly publish, implement, and enforce, a Data Management, Preservation, and Sharing Policy . . . [and that] each funded research project should submit a structured Data Management Plan for peer-review as an integral part of the application for public funding in the UK. (p. 6)

In line with the work on digital data curation and preservation in the United Kingdom, in the United States, the IWGDD has explicit goals for data management, which include the steps to document, organize, protect, and access research data (Lyon, 2007). In addition, NSF and the U.S. NIH require researchers to share their data (Deleserone, 2008).

Amid this environment of regulatory and policy changes, we find ourselves in a global financial crisis. As budgets tighten and certain funding dries up, the academic research community struggles to compete for public funding. Over the past few years, "the agency [NSF] boosted the size of

grants but held the number of awards steady and had to reject an increasing number of applications" (Mervis, 2007, p. 880). With more competition for research dollars, how will compliance with data standards and new evolving data management protocols affect researchers? Likewise, "This change from research quantity to research quality will require some different approaches to managing research activity" (Nicholls & Cargill, 2011, p. 214).

For instance, new performance-based research funding models and bibliometric systems that determine research funding levels based on an evaluation of research quality (e.g., based on citations rather than quantum), should reward research that generates greater impact as opposed to motivating researchers to "gin out articles" that may be in lower impact journals. However, although research quality is a priority for governmental offices that want to create greater accountability and more of a "pay for performance model," it does not necessarily align with new data management regulations that direct researchers to share data and provide open digital access. These imposed changes to data management requirements for the preservation, access, and sharing of publicly funded research data present immediate functional data challenges for researchers.

Functional Data Challenges

It has been suggested in the literature that institutions and funders have had little chance to keep pace with data-related matters, resulting in lagging structures, processes, and policies (Fear, 2011; Jones, 2012; Swan & Brown, 2008). In line with this research, additional studies point to three common data management issues: data storage, the lack of data policy, and issues involving legacy data. Data storage is often inadequate, with researchers resorting to suboptimal data storage methods, resulting in data that are often unreliable and short lived (Carlsen, 2006; Henty et al., 2008; Jones et al., 2008).

A lack of formal policies for creating and managing data is also common with some findings reporting data policies that are often idiosyncratic and largely defined by the individual researchers (Carlsen, 2006; Henty et al., 2008; Jones et al., 2008; Lyon, 2007). Similarly, concerning legacy data, a lack of data controls, such as access restrictions and edit rights, which increase the chance for data corruption and data integrity issues, is an often-cited concern (Carlsen, 2006; Henty et al., 2008; Jones et al., 2008; Ma & Pearson, 2005; Mullins, 2006).

To address these issues, it is recommended that "Higher education institutions should implement an institutional Data

Management, Preservation, and Sharing Policy, which recommends data deposit in an appropriate open access data repository and/or data center where these exist” (Lyon, 2007, p. 6). In fact, it has been suggested that dealing with the “deluge of data may be among the greatest challenges in the 21st century” (Carlsen, 2006, p. 1).

Although the idea of research data being widely and openly shared is not new (Harman, 2001), the means and methods by which to do it are underdeveloped, fragmented, and lacking. So, although researchers may be willing to share data, there is some negativity among researchers about data management as simply a bureaucratic requirement imposed on their time (Carlson, Ramsey, & Kotterman, 2010; Henty et al., 2008). Therefore, it is suggested that research institutions, rather than researchers, play a key role in implementing effective data management systems for research data outputs (Lyon, 2007, 2012). Resolving the future needs of the research community is highly dependent on research institutions taking data seriously (Swan & Brown, 2008).

In addition, the implementation of effective data management systems for research data throughout its entire life cycle requires involvement from the subject matter experts within the information systems (IS) communities on research campuses. Without the IS community taking an active role in finding solutions that ease the data burdens for researchers, the new regulations for data management combined with performance-based bibliometric systems of evaluation for research quality may dramatically reduce meaningful research output due to core skill gaps in data management by the research community.

Researcher Skill Gaps

Research conducted in Australia and the United Kingdom indicates that a significant number of researchers want and need additional training in data skills (Henty et al., 2008; Swan & Brown, 2008). In addition, although many researchers do not have a research data management plan, they do recognize the need for one and seek training in areas related to data management planning (Henty et al., 2008). The literature supports the notion that researchers can be at a loss as to how they should maintain data in the long term, and this lack of data expertise is exacerbated by inadequate data documentation (Jones et al., 2008). Additional risks for researchers have emerged with new standards by the OECD (2008) that refer to “bad data management” as data-related research misconduct. This form of misconduct contemplates situations such as not preserving primary data, bad data management practices and data storage, and withholding data from the scientific community as scientific misconduct regardless of whether the data-related misconduct was intentional or not.

Understanding that it is common for researchers to be responsible for their own data management, best practice guidance is needed to help researchers fulfill their responsibilities as “data creators and authors” (Lyon, 2007, p. 59). It

is recognized that managing data for a small study, with data collected at a single moment by one researcher is significantly easier than data management for complex, collaborative research teams that may be working on longitudinal studies. However, considering that many funding agencies are following the lead of the NSF by funding fewer researchers, but with greater average grant amounts, it is likely that the data management protocols, needs, and issues will also be larger and more complex over time.

Therefore, by understanding the current state of researcher data management practices in the United States, this exploratory case study will be accretive to the literature base and identify important knowledge and training gaps needed for public-funded researchers—particularly those working on large-scale studies that involve research teams. In addition, the descriptive survey outputs and analysis protocols can be used by other higher education institutions as a basis for discussion with their own staff and faculties regarding specific data management knowledge and needs that may exist.

Method

Case Study Approach

In an effort to highlight the current data management practices of university researchers, the case study methodology (Yin, 2003) was chosen. Case study research is useful when investigating a contemporary phenomenon within its real-life context involving multiple sources of data (Yin, 2003). Case studies are often not a methodological choice, *per se*, but rather a choice regarding what is to be studied (Stake, 2005). In this instance, the study’s focus is that of academic researchers and their data management practices at one Intermountain West university in the United States. It is believed that the case study approach yields the best descriptive output in alignment with the study’s goals. Quantitative data were analyzed via simple correlations, regression, and via descriptive statistical summaries.

In addition to the quantitative data, limited qualitative data were captured from open-ended survey questions. These data were analyzed and coded in accordance with traditional qualitative methods (Glesne, 2006). For purposes of this study, coding techniques and paradigms relied on Strauss’s systematic schema for coding qualitative data (Strauss & Corbin, 1990). This coding schema method was adopted because its systematic nature makes it easy for interested readers to see the codes that emerged from the data. Throughout the coding and analysis processes, grounded theory tenets were followed (Glaser & Strauss, 1967; O’Reilly, Paper, & Marx, 2012).

Study Design

Based on the gap in the literature regarding data management practices at U.S. research institutions, an online survey was conducted in 2009 at one Intermountain West university.

In addition, a data set was supplied by the Research Office of the university and provided the contact information for all principal investigators who were conducting current research studies that were funded, at least in part by public institutions. The entire population of researchers meeting the study criteria at this university is 426. All principal investigators were contacted by email to provide the Letter of Informed Consent and request participation in the online survey. During this initial contact, researchers were also given the opportunity to “opt out” of the survey. In sum, 28 researchers elected to opt out of the survey and 13 researchers had an email address that was undeliverable, leaving 385 researchers receiving the online survey instrument. After the survey was administered, another 18 researchers elected to “opt out” and another seven email addresses were undeliverable. Therefore, 360 researchers were potential respondents for the survey, with 135 respondents (37.5%).

Sample and Selection

This Intermountain West university is internationally recognized for its intellectual and technological leadership in land, water, space, and life enhancement. As the state's land-grant and space-grant institution, the university has 850 faculty members who provide education for more than 23,000 undergraduate and graduate students, including 10,000 in its continuing education sites located throughout the state. The university occupies 7,000 acres, 400 of which are on campus, with more than 200 buildings, 63 of which are devoted to academics. The university also has three branch campuses and extension offices in all the state's counties. Student-centered, hands-on learning opportunities are plentiful as this university attracted more than US\$186 million in research revenue in 2008.

Survey Instrument

The survey instrument used was adapted from the work of Henty et al. (2008) and Harman (2003, 2004) in which they explored the current data management practices (Henty) and attitudes of university researchers in Australia (Harman). Data management practices by researcher were explored via survey items related to the key stages of data life cycle management and included (a) formal data management planning; (b) physical storage and access of data files; (c) basic security provisions for data such as password protection, data backup, and remote access; and (d) access to data management expertise.

Other Data

In addition to the survey data, data were supplied by the University Research Office detailing the total funding amounts secured by each researcher in aggregate, as well as the number of current external grants. Finally, we collected

data on each respondent's experience as demonstrated by his or her academic rank.

Data Types and Coding

The data collected in the study consisted of two data sets: the original researcher file and the survey responses collected from those researchers. Both of these flat (nonnormalized) files were uploaded into a Microsoft SQL Server database and normalized into a relational database schema to improve the ability to query the data in various ways. In addition, numeric coding and scoring schemes were applied to the response file to render it more amenable to statistical analysis.

The original researcher file was a comma-separated text file containing information about the researchers who were surveyed, including their department and college affiliation and their current funded studies. This file originally contained 1,327 records, one record per study, with each researcher and each department/college possibly appearing on more than one record. These data were uploaded into SQL Server and normalized into three tables: department (including college), study, and investigator, with 426 distinct researchers in the investigator table.

At the conclusion of the survey data collection period, 135 of the 426 targeted researchers had responded. The data set that was produced by the online survey tool was a comma-separated text file consisting of 135 records, one for each respondent, with the question text in the columns of the first (header) record of the file and the actual text of the responses in the data columns for each record. Metadata tables were created and populated to define the survey questions and the possible responses for each question. A numeric code was defined for each possible response to each question and stored in the response metadata table, as was a scaled numeric score for each possible response to each question. When the survey response data file was uploaded into the SQL Server database, each text response was replaced with the appropriate numeric code, and the survey response data were stored in a normalized table as well as a flat table. Multiple cross-checks were run against the frequency tables, and the detail data returned by the survey tool to ensure data integrity were maintained throughout the normalization and transformation processes.

A backup database was created in the SQL Server instance, and individual tables were backed up to preserve each known good configuration prior to any transformation. All data extraction, transformation, and/or load (ETL) processes were performed using saved Data Transformation Services (DTS) jobs or saved Transact-SQL scripts to ensure reproducible results. In addition, the database was backed up periodically to protect the data against catastrophic hardware failure.

At the end of the data upload effort, the survey results were available in a normalized relational database, which is

Table 1. Survey Respondents by College/Department

College/department	Total researchers		Survey respondents	
	Number of researchers	Total (%)	Number of respondents	Total (%)
Administration-provost	2	0.47	0	0.00
Administration-VP administration affairs	2	0.47	0	0.00
Administration-VP research	2	0.47	1	0.28
College of Agriculture	68	15.96	26	7.22
College of Business	3	0.70	2	0.56
College of Education and Human Services	79	18.54	27	7.50
College of Engineering	69	16.20	18	5.00
College of Humanities, Arts, and Social Sciences	29	6.81	8	2.22
College of Natural Resources	50	11.74	18	5.00
College of Science	66	15.49	23	6.39
Cooperative Extension	37	8.69	10	2.78
Learning Resources Program	6	1.41	1	0.28
Regional Campuses and Distance Education	1	0.23	0	0.00
School of Graduate Studies	3	0.70	0	0.00
Student Administration and Services	9	2.11	1	0.28
	426	100.00	135	37.50 ^a

Note: VP = vice president.

^aBased on $n = 360$ receiving survey.

easy to query and provides many views of the survey data. Virtually, any level of granularity or aggregation of the data was possible. Transact-SQL views were created to provide the requested data sets for analysis. These views were opened directly by SPSS (version 17.0) for statistical analysis.

Key Findings

Survey Administration and Response Rates

The online survey instrument was initially sent at the end of February 2009 and remained open for response until the last day of March 2009. It should be noted that during this collection period, the university required a 1-week unpaid furlough for all staff and faculty due to budget issues. The university's spring break for students also occurred at this same time. Effects of the mandatory furlough on response rates are unknown but is worthy of mention due to the fact that this university had never before implemented a measure of this kind. Of the 360 investigators who received the survey, 135 responded for a 37.5% response rate. Of the 135 respondents, 117 respondents fully completed the survey and 18 respondents partially responded. There were no differences detected between the nonresponders, partial responders, and full responders in terms of individual study funding, or number of studies underway. This was ascertained by comparing the original researcher file supplied by the University Research Office and the survey respondent data set. The breakdown of respondents by college is detailed in Table 1.

Table 2. Data Collection Time Period

Over what period of time was/ is the data being collected?	Response frequency (%)	Response count
0-6 months	2.5	3
7-12 months	13.1	16
1-2 years	22.1	27
3-5 years	41.8	51
6 or more years	19.7	24
Do not know	0.8	1
	Total	122

Research Project Characteristics

Overall, 83.7% of the respondents are being funded by governmental agencies for their current research projects. The average project receives just over US\$111,000 per year in funding, but due to most researchers working on multiple grants, the funding on average *per researcher* is more than US\$346,000 annually. In general, research data are being collected over time with frequency counts and percentages as shown in Table 2.

Although the longitudinal nature of data collection can complicate the data management issues that researchers face, the types of data being collected can also magnify the data management issues. As shown in Table 3, the number, types, and complexity of data being collected are vast.

This variance of data types and complexity raises questions regarding data storage type and the location for data storage. Table 4 details the type and storage devices used.

Table 3. Data Types

Please check all the types of data that you are collecting	Response frequency (%)	Response count
Data automatically generated from or by computer programs	25.6	31
Data collected from sensors or instruments	41.3	50
Experimental data	53.7	65
Data collected by human interviewers/observers	47.1	57
Laboratory notes	28.1	34
Images, scans, or X rays	20.7	25
Websites	14.9	18
Blogs or discussion threads	1.7	2
Email	9.9	12
Digital audio or video files	9.1	11
Documents and reports	39.7	48
Other types of data being collected	11.6	14
	Total	121

Table 4. Data Storage Devices and Storage Locations

Where is the study's final data physically stored (or planned to be stored)?	Response frequency (%)	Response count
Stored in nondigital form such as paper records, photographs, videotapes or audiotapes, and so on	8.2	10
On a computer workstation	36.1	44
On a thumb drive or other removable medium	10.7	13
On a server	29.5	36
Other (please specify)	15.6	19
	Total	122
In what format will the final data be stored?	Response frequency (%)	Response count
Paper records only	3.3	4
Word or other text-based program	17.2	21
Excel or other spreadsheet program	21.3	26
SPSS or other statistical package	11.5	14
Database, relational database (SQL Server, MySQL, Access, etc.)	23.0	28
Other (please specify)	23.8	29
	Total	122

Data Management Practices

Current data management practices of researchers were explored through a series of questions. These questions related to whether there was a data management plan in place, if there was data management expertise on the research project, and data collection, housing, delivery, maintenance, and retiring of data being addressed.

In response to the question related to whether a formal data management plan was in place for the study, 60.7% of researchers have a formal plan in place and 36.1% stated that no plan was in place. A few researchers indicated that they were unsure (3.3%) of a formal data management plan, and these respondents were treated the same as “no formal plan in place” for statistical purposes. For the most part, the majority of researchers have either data management expertise and/or data management support available to them as shown in Table 5.

Data access, security, and data administration practices were ascertained through several questions related to access, security, and provisions in place to ensure the protection of

research data. In general, 68.3% of researcher's password protects their research data files, and an almost corresponding number (68.1%) do not allow researchers outside of their research teams to access their data. In line with this, only 20.8% of researchers allow any remote access of their research data. Those that do allow remote access use remote log-in (10.8%), file transfer protocol (FTP), or other methods of file transfer (10%), or web interfaces (16.7%). From a data access perspective, the majority of researchers have multiple people on their respective research teams accessing the data with varying degrees of practice for the backup of data as shown in Table 6.

Data Ownership

Researchers were queried regarding data ownership. The most common ownership structure reported was that of the public institution that funded the study ($n = 45$), the researcher and/or the researchers' team as data owners ($n = 39$), the university ($n = 20$), and finally, data that was

Table 5. Data Management Support

Describe the data management support available to you and your study	Response frequency (%)	Response count
None available	41.3	50
I have expertise in data management	16.5	20
I am not an expert, but there is someone working on this study with expertise in data management	21.5	26
I have expertise and also have at least one other person working with me on this study with expertise in data management	20.7	25
	Total	121

Table 6. Data Access and Backup Practices

How many people other than you access the data from your location?	Response frequency (%)	Response count
None	24.2	29
1-5	62.5	75
6-10	7.5	9
11 or more	5.8	7
	Total	120

How often is your data backed up?	Response frequency (%)	Response count
Never	2.5	3
Rarely	17.5	21
Monthly	20.0	24
Weekly	25.0	30
Daily	35.0	42
	Total	120

declared to be within the public domain ($n = 10$). Although ownership was most common for the agency funding the researcher, researchers report that in most cases they are primarily responsible for data once a study is complete ($n = 84$), with the funding institution next common ($n = 20$), and last, the university ($n = 8$) as shown in Table 7. However, many questioned their ability to “do that” with most expressing a frustration regarding how they might deal with legacy data or how data should be “managed” after a study is published. As one researcher eloquently stated when asked who would manage the data after a study concluded, “That’s a huge problem. It [data] will most likely die a lingering death resulting from the lack of attention.”

About the Researcher

Finally, the survey concluded with some general questions related to a researcher’s publication record from study data, their academic rank, previous data management training, and a chance for researchers to identify their own skill gaps by indicating content areas where they would like to receive additional data management training. Overall, although 34.8% of respondents had no publications or articles under review from their research data, the mean publications per researcher were just fewer than two publications, and 20%

Table 7. Data Ownership

Who owns the research data in your study?	Response frequency (%)	Response count
Public institution funding study	39.5	45
Researcher or research project team	34.2	39
Researcher’s university	17.5	20
Data are in the public domain	8.7	10
	Total	114

of respondents report having five or more publications. Most notable in this area was the self-reported researcher skill gaps. As one research participant shared,

No matter how we put it, creating, maintaining, and manipulating a database is not a chore everybody can easily perform. Setting it [the database] up and making it available to other team members while keeping it secured is most likely beyond the capability of many researchers.

Additional details regarding researcher characteristics are shown in Table 8.

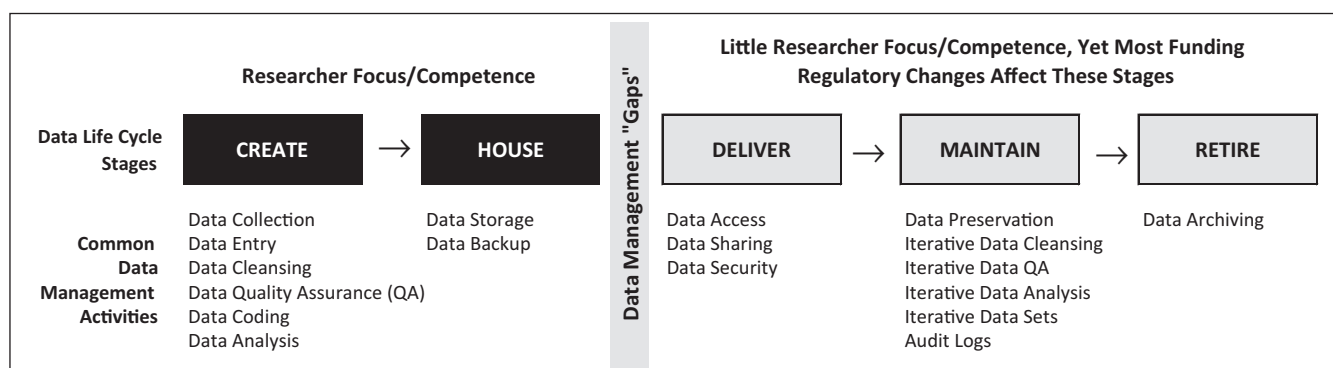
Discussion of Findings

Data Management Life Cycle Gaps

From the analysis, it appears that researchers are primarily focused on the collection and housing of research data. Researcher self-reported competence is highest during the data collection stage and most consider their efforts adequate for data housing—although certain questionable practices, such as no backup plans for data, were reported. Considering that performance-based and bibliometric systems for grant funding are becoming more common (particularly in Europe and Australasia), these two stages likely yield the greatest benefit to an individual researcher’s career, tenure, and promotion decisions, and certainly future external grant funding. However, additional research value exists within the other life cycle stages for research data—specifi-

Table 8. Researcher Characteristics

What is your current academic rank?	Response frequency (%)	Response count
Professor (or emeritus)	38.5	45
Associate professor (research or academic)	23.9	28
Assistant professor	24.8	29
Postdoctoral fellow	0.0	0
Staff	6.0	7
Lecturer	0.9	1
Other (please specify)	6.0	7
Total		117
Which most closely represents your level of training (either formal or on-the-job training) in data management	Response frequency (%)	Response count
Beginner: I have no data management training	33.3	39
Intermediate: I have had some training but need help	36.8	43
Advanced: I have extensive training and/or I am proficient	28.2	33
Certified: I am a certified database administrator, developer, or data management professional	0.0	0
Other (please specify)	1.7	2
Total		117
Please indicate if you are interested in receiving training or assistance in any of the following areas (check all that apply)	Response frequency (%)	Response count
Digitization advice, tools, and services	30.5	25
Creating a research data management plan at the beginning of a project	73.2	60
Creating a research data management plan after a project has finished	42.7	35
A data-transition plan (e.g., for personnel changes, departing postgraduate student assistants)	28.0	23
Data rescue for older digital materials such as older media or migration of data from legacy systems	35.4	29
Data-sharing plan	25.6	21
Other (please specify)	6.1	5
Total		82

**Figure 2.** Data management life cycle "gaps"

cally in the stages of delivery and maintenance. These stages are also where most new demands and requirements exist for data management plans and policies as conditions of external grant funding. These gaps, between researcher competence and focus and the new funding regulatory changes, are shown in Figure 2.

It seems plausible that institutions that are able to effectively leverage campus-wide assets to address research data life cycle and management issues should be the beneficiaries

of greater research dollars. We posit that the combination of research and IS assets should provide the best scenario for addressing these gaps. As suggested in the literature, institutions and researchers have had little chance to keep pace with data-related matters, resulting in lagging structures, processes, and policies (Fear, 2011; Jones, 2012; Swan & Brown, 2008). In line with this, respondents in this survey had varied experiences concerning data storage, data policies, and issues involving legacy data.

Table 9. Regression Model: Data Management Practices and Researcher Experience

Model summary									
Model	R	R ²	Adjusted R ²	SE of the estimate	Change statistics				
					R ² change	F change	df1	df2	Significant F change
1	.229 ^a	.052	.044	19.716	.052	6.357	1	115	.013

^aPredictors: (constant), researcher experience.

First, the functional data challenges faced by university researchers are often left in the hands (and minds) of individual researchers. For instance, whereas 60.7% of researchers report that a formal data management plan is in place for their data, 39% of researchers report backing up their data monthly or less frequently, and 41.3% report having no expertise themselves and no access to a data management expert to assist with the housing, delivery, maintenance, and retirement of research data over time. To address these functional challenges will require institutions rather than individual researchers to take action and make changes (Lyon, 2007, 2012). Without institutional involvement, it is doubtful if an individual researcher will be able to overcome the deluge of data and evolving reporting requirements necessary to remain competitive for governmental grants.

Second, concerning housing research data, a surprising number of researchers (46.8%) are storing data files on a single computer workstation or removable storage device, such as a thumb drive. As one participant mentioned, "For the majority of the research, we are simply storing and sharing data on individual PC's. We share data as needed over email." Although this is not overly problematic at face value, it is interesting to note that this researcher is referencing a large collaborative project with backup protocols of monthly or less frequently. Therefore, considering that the typical study is collecting data over time (vs. in a single data collection effort), the issues of data reliability are confounded by these imperfect and fragile storage methods, with potential impact for data integrity, loss, or corruption (Carlsen, 2006; Henty et al., 2008; Jones et al., 2008), particularly when the projects are team-based collaborative efforts.

Third, it is interesting to note that a regression analysis yielded a significant correlation between higher aggregate scores for researcher data management practices and a researchers' experience and training ($(1, 115) = 6.357, p = .013 < .05$). Thus, if institutions are inclined to support researchers in their craft, training should be one of the first areas for consideration, with input from researchers providing focus toward establishing a data management plan at the beginning and/or end of a project and for assistance with legacy data issues (Table 9). In fact, several researchers commented on this. For instance, "I think that research data management training would be invaluable," and "I'd like someone on campus to teach MS Access for professors," and "I think on-campus training is an excellent idea."

This last area is of particular concern due to the researchers' perceived ownership of their data after study completion—a notion not in step with the idea of data value being realized through the use and reuse of data, or with the ownership interests of federal funding agencies (Lyon, 2007, 2012).

How the IS Community Can Build Value

Researchers exhibit great focus on the collection of data and seem to be well prepared to determine an appropriate housing plan for research data. However, not surprisingly, the delivery, maintenance, and retiring of data are beyond the typical researcher's skill set and their interest. According to Carlson et al. (2010),

open access and digital preservation do not seem to resonate as immediate, direct issues at a local-scale of concern for many researchers. These grand-scale challenges are often things that researchers should do for the good of the scholarly community or institution, rather than things that they have to do to satisfy their day-to-day needs or get their work done. (p. 155)

In these areas (delivery, maintenance, and retiring of research data), the university IS community excels and stands uniquely positioned to deliver greater value campus-wide.

It is widely accepted in tertiary education that a high-quality research culture underpins and enhances the teaching and learning that occurs in these settings. So naturally, support and assistance offered to research activities can improve the environment for all stakeholders. However, to fully leverage the IS community assets in addition to those of researchers requires more than a utopian view of value. From our analysis, it is clear that the "research perspective" is significantly different than the "IS perspective" regarding data management. In fact, the five-stage view of data management life cycle (Figure 1) is more of an IS perspective than research perspective. Researchers, at least those in this study, seem to view the research data life cycle in three stages: (a) collect data, (b) analyze and store data, and (c) publish data. Because researchers are rewarded through publication, it is natural that aspects relating to data management after journal acceptance are viewed as "bureaucratic red tape" rather than activities that add value to the research endeavor.

Combining the expertise and efforts of researchers and the university IS community might provide value and benefit to both stakeholder groups. However, based on the data from this study as well as conversations with IS and research-focused professionals, a melding must take place to unify the two constituent groups together in a way that magnifies their likeness rather than differences. One place to start is to establish good communication between researchers and the IS community to create a common lexicon for coordinating efforts. This will require the IS community to understand and be sensitive to the protective stance researchers are likely to assert, particularly concerning human subject research endeavors and corresponding data. This will also require the research community to consider themselves as data stewards rather than data owners. Finally, data management initiatives must be institutionally driven (e.g., by the Office of the Vice President for Research) rather than driven by a particular discipline or the requirements of a specific funding agency (Deleserone, 2008).

Only time will tell whether coalescing the research and IS assets on university campuses through institutionally driven data management assistance and support programs will provide evidence linking sound data management and life cycle procedures and processes to better research outcomes and value. However, at this point, it is an argument that many in the research community dispute. As one researcher stated,

It is not clear why data management practices should affect success in obtaining and managing a grant; the quality of the research is the major predictor in obtaining a grant . . . It is expected that investigators will protect data for privacy and security, but there is no formal assessment when evaluating a grant. There is no guarantee that relational database practices will result in better data management; at times, this practice may be overkill. In addition, data can be encrypted whether on server, on workstation, or on a USB stick, and such differences may not be critical. Such factors may not increase the probability of obtaining external funding.

From a researcher perspective, this comment makes intuitive sense. But, we posit that it also highlights a limitation of many researchers—the recognition that data are what underpins any research output. Regardless of the theory/theories used, the statistical tests employed, and grand-scale collection of data and artifacts, data are what ground the output. Without accurate, complete, and reliable data, our outputs become conceptual and limited. Therefore, if that premise can be accepted, then the manner and means by which research data are collected, manipulated, and managed must be inextricably linked to the quality of the research output. Likewise, the regulatory requirement for data sharing is not addressed by many researchers, and this aspect of access and

sharing may be a fundamental mechanism for generating long-term research value. When scrutinized under new performance-based or bibliometric funding schemas, effective data management and holistic oversight of research data life cycles will improve the likelihood of data sharing, long-term data use, and therefore, an improvement in the quantitative measures of research value—thus, providing greater future funding opportunities.

Limitations and Topics for Future Study

We have explored how researchers currently manage their research funding against a backdrop of significant changes to the requirements of funding agencies. However, these changes in research funding requirements have impacts beyond that of the researcher. These impacts create several opportunities for future work. For instance, should funding agencies, government, or industry be responsible for data management? Could university libraries best serve the role of data archivist in support of university researchers? How are Institutional Review Boards responding to the sharing requirements during their review and administration of research projects? Finally, will cloud computing have any impact on the data management practices of university researchers? Each of these areas, although beyond the scope of this work, creates interesting and noteworthy topics for future study.

Concluding Remarks

This study set out to understand the current data management practices at one U.S. research university. The literature supports the notion that researchers can be at a loss as to how they should maintain data in the long term (Jones, 2012; Jones et al., 2008). Considering the growing competition for research dollars in the current economy, those best prepared to demonstrate data proficiency as well as research and grant prowess will most likely be the greatest grant beneficiaries. The role that research institutions can play in supporting researchers through training seems intuitive and supported by the data in this study. Understanding the self-reported researcher skill gaps and importance of training gives institutions better insight as to the role they can play in aligning research and IS assets of university communities to collectively improve data creation, preservation, and its use and reuse. For it is our belief that an important competitive advantage exists for researchers and institutions that leverage better data management practices in an effort to increase external research grant funding.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

- Carlsen, S. (2006). Lost in a sea of science data: Librarians are called in to archive huge amounts of information, but cultural and financial barriers stand in the way. *Chronicle of Higher Education*, 52, A35.
- Carlson, J., Ramsey, A. E., & Kotterman, D. J. (2010). Using an institutional repository to address local-scale needs: A case study at Purdue University. *Library Hi Tech*, 28(1), 152-173.
- Deleserone, 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.
- Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk>
- Fear, K. (2011). You made it, you take care of it. Data management as personal information management. *International Journal of Digital Curation*, 6, 54-77.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Hawthorne, NY: Aldine De Gruyter.
- Glesne, C. (2006). *Becoming qualitative researchers* (3rd ed.). Boston, MA: Pearson Education.
- Hall, B. H. (2004, February). *University-industry research partnerships in the United States*. Kansai Conference Paper. Retrieved from http://elsa.berkeley.edu/~bhhall/papers/BHH04_Kansai.pdf
- Harman, G. (2001). University-industry research partnerships in Australia: Extent, benefits and risks. *Higher Education Research & Development*, 20, 243-264.
- Harman, G. (2002). Australian university-industry research links: Researcher involvement, outputs, personal benefits and "withholding" behavior. *Prometheus*, 20, 143-158.
- Harman, G. (2003). Australian academics and prospective academics: Adjustment to a more commercial environment. *Higher Education Management and Policy*, 15, 105-122.
- Henty, M., Weaver, B., Bradbury, S., & Porter, S. (2008). *Investigating data management practices in Australian universities. Australian Partnership for Sustainable Repositories (APSR)*. Retrieved from: http://www.apsr.edu.au/orca/investigating_data_management.pdf
- Hopwood, P. (2008). *Data governance: One size does not fit all*. Retrieved from <http://www.dmreview.com/issues>
- Information Management. (2011). *Data management channel*. Retrieved from http://www.information-management.com/channels/data_management.html
- Interagency Working Group on Digital Data. (2007). *US national science data group formed*. Retrieved from http://scilib.typepad.com/science_library_pad/2007/03/us_national_sci.html
- Joint Information Systems Committee. Available from <http://www.jisc.ac.uk>
- Jones, S. (2012). Developments in research funder data policy. *International Journal of Digital Curation*, 7, 114-125.
- Jones, S., Ball, A., & Ekmekcioglu, C. (2008). The data audit framework: A first step in the data management challenge. *International Journal of Digital Curation*, 3, 113-120.
- Lyon, L. (2007). Dealing with data: Roles, rights, responsibilities and relationships. Retrieved from <http://opus.bath.ac.uk/412/>
- Lyon, L. (2012). The informatics transform: Re-engineering libraries for the data decade. *International Journal of Digital Curation*, 7, 126-138.
- Ma, Q., & Pearson, M. L. (2005). ISO 17799: Best practices in information security management? *Communication of the Association for Information Systems*, 15, 577-591.
- Mervis, J. (2007). Grants management: NSF survey of applicants finds a system teetering on the brink. *Science*, 317, 880-881.
- Miller, H. G., & Baldwin, W. H. (2001). Research letters. A terse amendment produces board change in data access. *American Journal of Public Health*, 91, 824-825.
- Mullins, C. S. (2006). *The impact of regulatory compliance on data management*. Retrieved from <http://www.neonesoft.com/doc/pres/Compliance.pdf>
- National Institutes of Health. (2003). Retrieved from http://grants.nih.gov/grants/policy/data_sharing/
- National Science Foundation. (2011). *See National Science Foundation web site, data management & sharing frequently asked questions (FAQs)*. Retrieved from <http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp>
- Nicholls, M. G., & Cargill, B. J. (2011). Establishing best practice university research funding strategies using mixed-mode modeling. *Omega*, 39, 214-225.
- Office of Management and Budget. (1999). *Circular A-110*. Retrieved from <http://www.whitehouse.gov/omb/circulars/a110/a110.html>
- O'Reilly, K., Paper, D., & Marx, S. (2012). Demystifying grounded theory for business research. *Organizational Research Methods*, 15, 247-262. doi: 10.1177/1094428111434559.
- Organisation for Economic Co-operation and Development. (2008). *Governance of public research: Steering and funding of research institutions country report: United States*. Retrieved from <http://www.oecd.org/dataoecd/24/33/2507966.pdf>
- Stake, R. E. (2005). Qualitative case studies. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (3rd ed., pp. 443-466). Thousand Oaks, CA: SAGE.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: SAGE.
- Swan, A., & Brown, S. (2008). *The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs*. Report to the JISC. Retrieved from <http://www.jisc.ac.uk/publications/publications/dataskillscareersfinalreport.aspx>
- UKOLN. (2009). Retrieved from <http://www.ukoln.ac.uk/about>
- Vincent-Lancrin, S. (2006). What is changing in academic research? Trends and futures scenarios. *European Journal of Education*, 41, 169-202.

Yin, R. (2003). *Case study research: Design and methods*. London, England: SAGE.

Bios

Kelley O'Reilly is an assistant professor of sales and business marketing at Western Michigan University. She has more than 20 years of executive level experience in franchise retail and service businesses and her research interests are qualitative research methods, customer relationship management (CRM), sales, and service quality. Her work has been published in journals, such as *Organizational Research Methods*, *Information Resources Management Journal*, *Informing Science: The International Journal of an Emerging Transdiscipline*, *International Journal of Electronic*

Business Research, and *Journal of Information Technology Case and Application Research*.

Jeffrey Johnson is an associate professor in the Management Information Systems Department at the Jon M. Huntsman School of Business at Utah State University. His research interests include personal web use at work and employee Internet abuse, Internet game addiction, and web analytics.

Georgiann Sanborn is a certified Microsoft SQL Server Database developer and administrator. She has more than 20 years of experience designing relational databases and developing database-enabled applications.