

**Full-Length Paper****An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Findings from the Quantitative Phase**

Elizabeth A. Berman

Tufts University, Medford, MA, USA

*Formerly Library Associate Professor, University of Vermont

Abstract

This article reports on the second quantitative phase of an exploratory sequential mixed methods research design focused on researcher data management practices and related institutional support and services. The study aims to understand data management activities and challenges of faculty at the University of Vermont (UVM), a higher research activity Research University, in order to develop appropriate research data services (RDS). Data was collected via a survey, built on themes from the initial qualitative data analysis from the first phase of this study. The survey was distributed to a nonrandom census sample of full-time UVM faculty and researchers (P=1,190); from this population, a total of 319 participants completed the survey for a 26.8% response rate. The survey collected information on five dimensions of data management: data management activities; data management plans; data management challenges; data management support; and attitudes and behaviors towards data management planning. Frequencies, cross tabulations, and chi-square tests of independence were calculated using demographic variables including gender, rank, college, and discipline. Results from the analysis provide a snapshot of research data management activities at UVM, including types of data collected, use of metadata, short- and long-term storage of data, and data sharing practices. The survey identified key challenges to data management, including data description (metadata) and sharing data with others; this latter challenge is particular impacted by confidentiality issues and lack of time, personnel, and infrastructure to make data available. Faculty also provided insight to RDS that they think UVM should support, as well as RDS they were personally interested in. Data from this study will be integrated with data from the first qualitative phase of the research project and analyzed for meta-inferences to help determine future research data services at UVM.

Correspondence: Elizabeth A. Berman: elizabeth.berman@tufts.edu**Keywords:** data management, mixed methods research, quantitative research, research data services, academic libraries, survey**Rights and Permissions:** Copyright Berman © 2017All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Introduction

The need for data curation, “the active and ongoing management of data through its life cycle of interest and usefulness to scholarship, science, and education” (Council on Library and Information Resources 2016, para. 1), has become a major issue in scholarly communication: “Data curation activities enable data discovery and retrieval, maintain its quality, add value, and provide for reuse over time” (para. 1). Since 2003, the National Institutes of Health (NIH) have required investigators requesting \$500,000 or more in direct costs in any year of a grant to share their data with the scientific community (National Institutes of Health 2003). In 2011, the National Science Foundation (NSF) began to require that researchers submit a data management plan (DMP) with their grant applications; the purpose of the DMP was to account for the long-term preservation of and access to scientific research data produced through government funding. In 2013, the White House Office of Science & Technology Policy (OSTP) issued a directive that requires granting agencies to develop a plan to make both the data and published articles of federally funded research available to the public at no cost. Since that memorandum, federal agencies have been developing their own plans and policies to account for public access to federally funded research; the Association of Research Libraries (ARL) website (2016) is maintaining links to these agency plans.

Beyond federal research mandates, data in and of itself is increasingly being acknowledged as a scholarly product, a crucial part of academic discourse that has the potential to impact future research (Williford and Henry 2012). This is particularly true in interdisciplinary and transdisciplinary domains such as environmental studies where researchers are “dependent upon access, discovery, and interoperability of data sets drawn from a variety of sources” (Scaramozzino, Ramirez, and McGaughey 2012, 350). Data curation also extends into the arts and humanities; Flanders and Muñoz write, “a key aspect of humanities data curation is thus to ensure that the representations of objects of study in the humanities functions effectively as data: that they are processable by machines and interoperable such that they are durably processable across systems and collections while still retaining provenance and complex layers of meaning” (2014, para. 3).

This increased recognition of the importance of preserving and maintaining digital data has had a direct impact on higher-education institutions that are working to provide data curation services, or “the active management and appraisal of digital information over its entire life cycle” (Pennock 2007, para. 2). A number of researchers have conducted needs assessments or environmental scans of their institutions in order to understand their research data landscape. One popular method for conducting these scans has been to utilize quantitative methods, an approach that collects and analyzes numerical data from a sample population in order to examine the relationship among variables to test theories and generalize to a broader population (Creswell 2014; Singleton and Straits 2010). In particular, multiple studies have been published using survey instruments to collect data from a diverse sample (Table 1). These studies are generally framed around the Data Lifecycle Model (DDI Alliance Structural Reform Group 2004): collecting research data; describing, analyzing, and short-term storage of data; and access to and long-term preservation of data (Figure 1).

Table 1: Comparison of methods used in data management studies

Method	Author(s)	Institution	Sample Size
Survey	Akers and Doty (2013)	Emory University	13 questions 330 respondents
	D'Ignazio and Qin (2008)	SUNY College of Environmental Science & Forestry Syracuse University	--- 111 respondents
	Diekema, Wesolek, and Walters (2014)	multi-institution	16 questions 196 respondents
	Parham, Bodnar, and Fuchs (2012)	Georgia Institute of Technology	--- 63 respondents
	Scaramozzino, Ramírez, and McGaughey (2012)	California Polytechnic State University, San Luis Obispo	18 questions 82 respondents
	Steinhart <i>et al</i> (2012)	Cornell University	43 questions 86 respondents
	Tenopir <i>et al</i> (2011)	multi-institution	23 questions 1,329 respondents
	Weller and Monroe-Gulick (2014)	University of Kansas	--- 415 respondents
	Whitmire, Boock, and Sutton (2015)	Oregon State University	29 questions 443 respondents

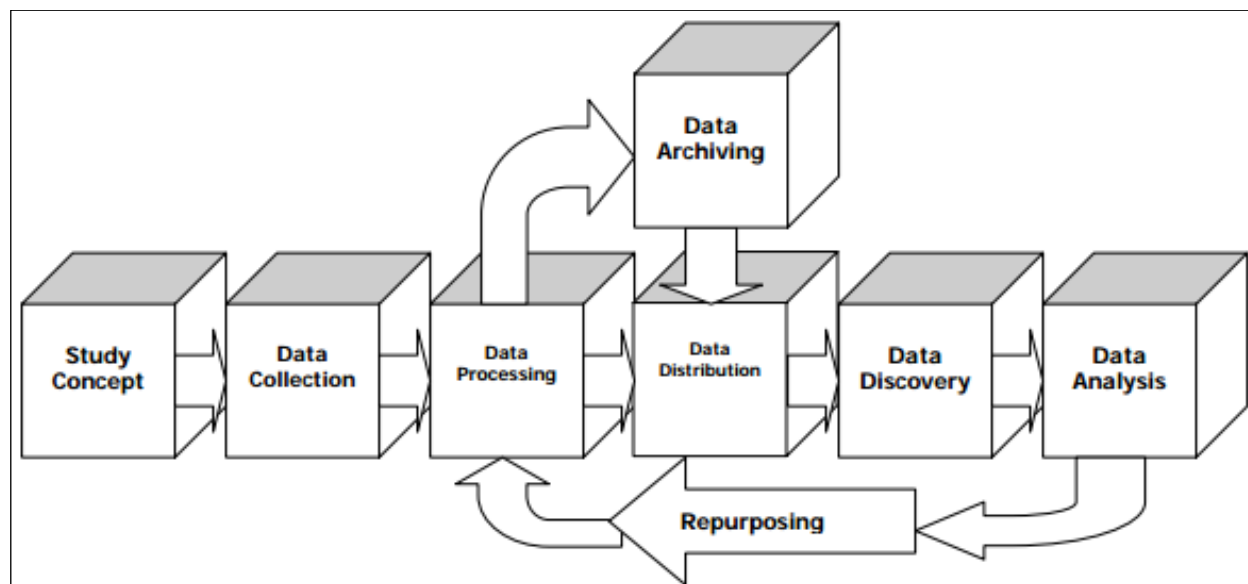


Figure 1: Data Lifecycle Model (DDI Alliance Structural Reform Group 2004)

A number of these studies explicitly focus on researchers in the science and technology fields, where discussions about data management have been accelerated due to NIH and NSF funding mandates. Cornell University's Research Data Management Service Group surveyed NSF Principal Investigators (PIs) "in order to understand how well-prepared researchers are to meet the new NSF data management planning requirement, to build our own understanding of the potential impact on campus services, and to identify service gaps" (Steinhart et al. 2012, 64). Diekema, Weslock, and Walters (2014) investigated whether science and engineering researchers had the skills to effectively manage data and whether the institution had the necessary infrastructure to support data management activities. To answer these research questions, the authors surveyed three groups of interest: STEM faculty, sponsored program officers, and academic librarians affiliated with institutional repositories.

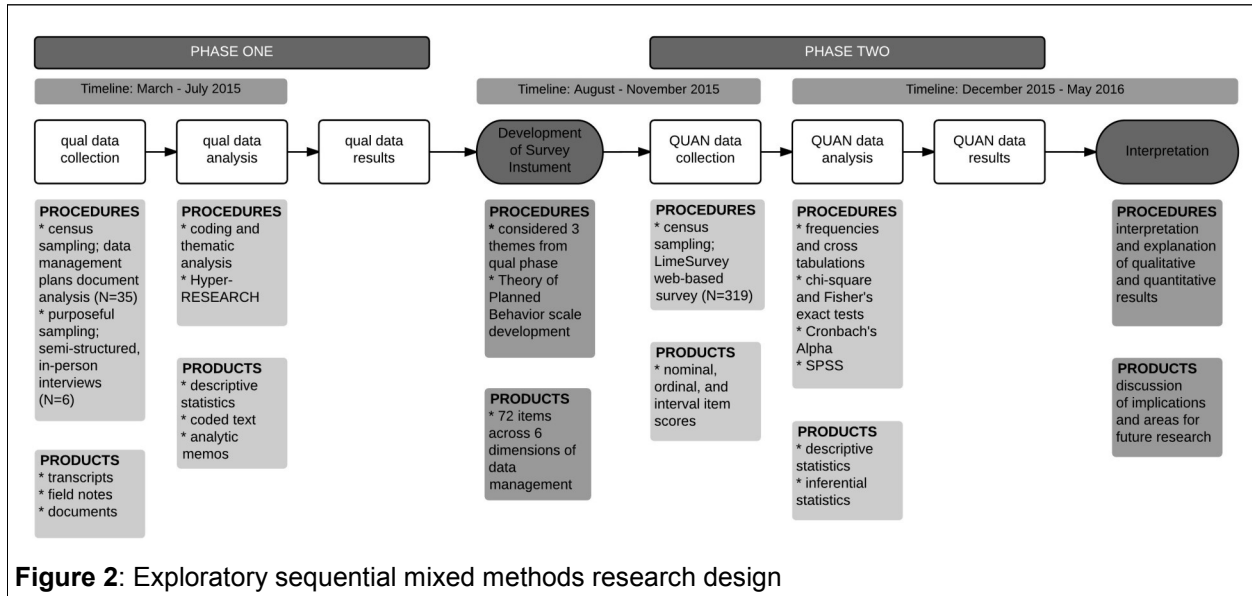
Other researchers are taking a broader approach, surveying the entire faculty population to understand similarities and differences in disciplinary management of digital data. Parham, Bodnar, and Fuchs (2012) designed a survey to better understand data resource output in order to "discover the types of data assets created and held by researchers, how the data are managed, stored, shared, and reused, and researchers' attitudes toward data creation, sharing, and preservation" (10). Scaramozzino, Ramirez, and McGaughey (2012) surveyed teacher-scholar faculty at California Polytechnic State University, San Luis Obispo, to address issues of data preservation, data sharing, and education needs of researchers managing data. Akers and Doty (2013) and Whitmire, Boock, and Sutton (2015) used surveys to understand varying approaches to data management in order to develop appropriate research data services.

These studies are informative to the research behaviors of faculty, but their focus on institutional populations limits their generalizability to all research faculty. McLure et al (2014) emphasize that "local studies can inform libraries and librarians about the behaviors, needs, interests, and concerns of researchers at individual institutions" (158). Guided by the literature, this study is crucial to unpacking and understanding specific approaches to data management, as well as data management needs and challenges, at the University of Vermont.

Purpose Statement

This article reports on the second phase of an exploratory sequential mixed methods research (MMR) design aimed at understanding data management behaviors and data management planning attitudes of faculty at the University of Vermont (UVM). The strength of mixed methods research is that it draws on the strengths of both qualitative and quantitative research, providing a more holistic understanding of a problem or phenomenon. The exploratory sequential mixed methods design, characterized by an initial phase of qualitative data collection and analysis, followed by a phase of quantitative data collection and analysis (Figure 2), was selected in order to develop better instruments to measure data management activities at UVM, including behaviors and attitudes toward data management planning (Creswell 2014).

For the quantitative phase of this study, a survey instrument was developed based on the qualitative analysis of the first phase of the study in order to establish a broad understanding of the campus data management environment (Berman 2017). The survey measured the following dimensions: data management activities; data management plans; data management



challenges; data management support; attitudes and behaviors towards data management planning; and demographics. This survey was deployed to all current UVM faculty and researchers in an attempt to reveal key distinctions among different populations of researchers and generalize the findings from the phase one qualitative research, which only focused on successful National Science Foundation (NSF) grantees (Berman 2017).

The second phase of this MMR research was guided by four research questions. The first two parallel the research questions from the qualitative phase, while questions three and four were developed explicitly from the qualitative data analysis (Berman 2017):

RQ1: How do faculty at UVM manage their research data, in particular how do they share and preserve data in the long-term?

RQ2: What challenges or barriers do UVM faculty face in effectively managing their research data?

RQ3: What institutional data management support or services are UVM faculty interested in?

RQ4: How do researchers' attitudes and beliefs towards the data management planning process influence their data management behaviors, in particular how do they intend to share and preserve their data?

The primary objective of this phase of the research study is to understand researchers' current data management behaviors and challenges within and across all disciplines. The results of this phase will be integrated with the results of the first phase to guide the development of research data services at UVM. As a result, the analysis of RQ4 will not be addressed in this publication as it proposes the development of a bipolar adjective scale to assess attitudes and beliefs towards the data management planning process in order to measure intention of implementing data management plans.

Methods

Population

The target population for this quantitative study was all full-time faculty at the University of Vermont. UVM is a higher-research activity Research University with a humanities and social sciences-dominant graduate instructional program (The Carnegie Classification of Institutions of Higher Education 2017). In 2015-2016, UVM enrolled 10,081 undergraduate students, 1,360 graduate students, and 457 medical students (University of Vermont 2017). Working with the Office of Institutional Research, a list was generated of 1,190 full-time instructional and research faculty as of October 1, 2015. Using nonrandom census sampling, the entire population was invited to participate in the survey via a personalized email invitation.

Survey Instrument Development

Surveys provide a means to standardize measurement of a phenomenon, ensuring that consistent information is obtained across all respondents (Fowler 2014). Utilizing design-level data linking (Creswell and Plano Clark 2011; Fetters, Curry, and Creswell 2013), themes from the analysis of the qualitative data were used to drive development of the survey instrument; in particular, the language used and themes addressed by interview participants and in data management plans formed the foundation for writing questions (Berman 2017). Questions related to attitudes and behaviors used the theory of planned behavior (Ajzen 1991; Ajzen 2005; Ajzen and Fishbein 2000) as a model of how researcher attitudes and beliefs guide intention and behavior towards data management. Survey development was also informed by prior research (in particular Akers and Doty 2013; Scaramozzino, Ramírez, and McGaughey 2012; Whitmire, Boock, and Sutton 2015).

The survey included 46 questions (Q1-Q46) and 72 items covering five dimensions: data management activities (Q4-Q16); data management plans (Q17-Q23); data management challenges (Q33); data management support (Q34-Q41); and attitudes and behaviors towards data management planning (Q24-Q32). Q1 was used to screen out participants who do not collect, generate, or use data for their research, while Q3 screened out participants who do not engage in management of digital data. These participants were branched to the demographics section. Demographic data (Q42-Q46) was requested from all survey participants and included college, department, rank, number of years at UVM, and gender. The full instrument can be found in the Appendix.

Survey Administration

The survey was created using UVM's LimeSurvey software license, which allowed for electronic distribution and collection of data. Following the advice of Dillman, Smyth, and Christian (2008), the layout provided intuitive navigation through the survey instrument, the questions were uncluttered and easy to read, and the response tasks were simple, with predominantly closed-question options. The survey was pre-tested by six faculty researchers in four disciplines to ensure that the questions were well understood and that the answers were meaningful (Madans et al. 2011; Presser et al. 2004). Based on feedback from the pre-test, survey questions and instrument design were modified. A final survey instrument was

submitted to the UVM Research Protections Office and received a Protocol Exemption Certification.

All full-time UVM faculty and researchers were invited to participate in the study via a personalized email that included a brief description of the purpose of the survey and a unique link to the survey. To encourage participation, the survey invited participants to enter their names into a raffle for six \$50 Amazon.com gift certificates at the completion of the survey. The survey was open from October 20, 2015 through November 11, 2015, with two reminder emails sent on October 29, 2015, and November 9, 2015. Data were downloaded from LimeSurvey and analyzed in SPSS version 22.

Results

Quantitative Survey Respondents

Of the 1,190 UVM faculty who were invited to participate in the survey, 345 participants started the survey and 319 participants completed the survey for a 26.8% response rate. This response rate is within the range of online response rates (20.0% to 47.0%) identified by Nulty (2008), and is comparable to response rates from similar published research (D'Ignazio and Qin 2008; Whitmire, Boock, and Sutton 2015). While appropriate measures were taken to reduce sources of bias, the relatively low response rate increases the potential for non-response bias, where respondents differ in meaningful ways from non-respondents (Singleton and Straits 2010). Descriptive statistics of respondent demographics can be found in Table 2.

Table 2: Descriptive statistics of participants in phase two

Value	Observed Frequency	Observed Proportion	Expected Frequency	Expected Proportion	Percentage Deviation	Standardized Residuals
College¹						
BSAD	5	0.020	6.47	0.026	-22.72%	-0.58
CALS	14	0.056	12.45	0.050	+12.45%	+0.44
CAS	85	0.341	65.74	0.264	+29.30%	+2.38
CEMS	18	0.072	16.19	0.065	+11.18%	+0.45
CESS	19	0.076	13.45	0.054	+41.26%	+1.51
CNHS	12	0.124	11.21	0.045	+7.05%	+0.24
COM	86	0.345	115.29	0.463	-25.41%	-2.73
RSNER	10	0.040	8.22	0.033	+21.65%	+0.62
TOTAL	249					
Rank						
Full professor	59	0.234	65.27	0.259	-9.61%	-1.05
Associate professor	81	0.321	71.11	0.306	+5.04%	+0.13
Assistant professor	80	0.317	63.0	0.250	+26.98%	+1.83
Senior lecturer	10	0.040	14.87	0.059	-32.75%	+0.92
Lecturer	13	0.052	13.61	0.054	-4.48%	-0.29
Other	9	0.036	18.14	0.072	-50.39%	-2.26
TOTAL	252					
Gender						
Female	129	0.512	104.33	0.414	+23.65%	+2.42
Male	123	0.488	147.67	0.586	-16.71%	-2.03
TOTAL	252					

1 BSAD = Business Administration; CALS = Agriculture & Life Science; CAS = Arts & Science; CEMS = Engineering & Mathematical Sciences; CESS = Education & Social Services; CNHS = Nursing & Health Sciences; COM = Medicine; RSENR = Environment & Natural Resources.

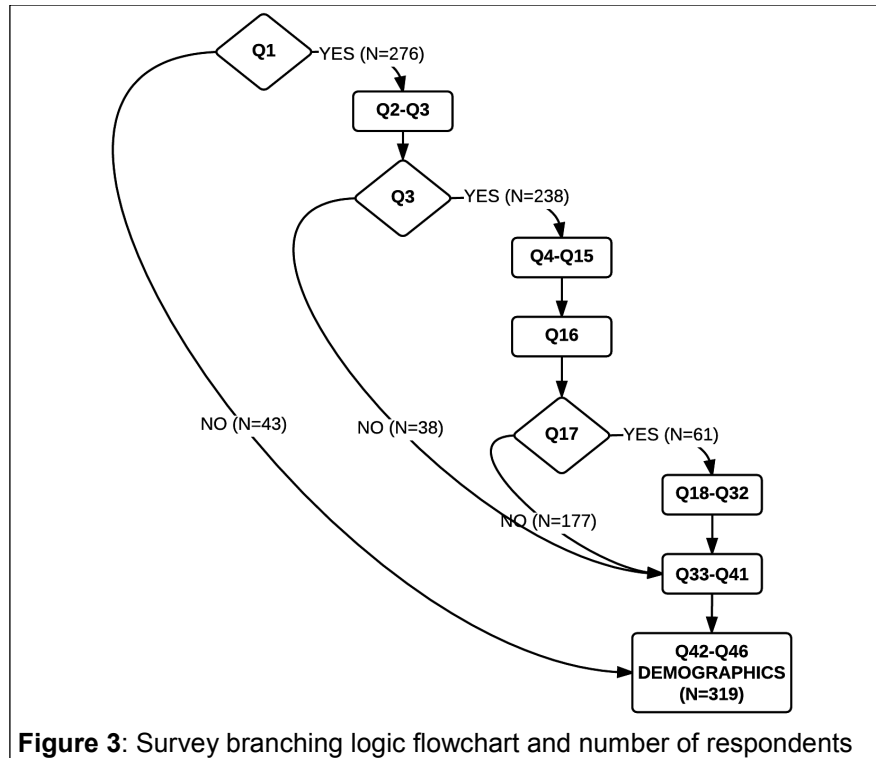
Due to the wide range of disciplines within the College of Arts and Sciences, faculty were also sorted into disciplinary categories for analysis: Arts & Humanities (A&H), Social Sciences & Business (SS&B), and Science, Technology, Engineering & Mathematics (STEM) (Table 3).

Table 3: Disciplinary alignment of survey respondents

Arts & Humanities (A&H) (N=38)	Social Sciences & Business (SS&B) (N=38)	STEM (N=162)
Art & Art History (CAS) Asian Languages & Literature (CAS) Classics (CAS) English (CAS) German & Russian (CAS) History (CAS) Music & Dance (CAS) Philosophy (CAS) Religion (CAS) Romance Languages & Linguistics (CAS) Theater (CAS)	Anthropology (CAS) Business (BSAD) Community Development & Applied Economics (CALs) Economics (CAS) Education (CESS) Geography (CAS) Leadership & Development Science (CESS) Political Science (CAS) Psychological Sciences (CAS) Social Work (CESS) Sociology (CAS)	Animal Science (CALs) Biochemistry (CALs) Biology (CAS) Chemistry (CAS) Computer Science (CEMS) Engineering (CEMS) Geology (CAS) Mathematics & Statistics (CEMS) Medicine (COM) Microbiology & Molecular Genetics (CALs) Natural Resources & Environment (RSENR) Nursing & Health Sciences (CNHS) Nutrition & Food Science (CALs) Physics (CAS) Plant & Soil Science (CALs) Plant Biology (CALs)

Because of the wide representation of researchers within the population of study, not all survey questions were applicable to all respondents. Screening questions and branching logic were employed to ensure participants were asked to respond only to relevant questions; depending on responses, participants could be asked to answer 6 questions (N=43), 16 questions (N=38), 30 questions (N=177), or 46 questions (N=61) (Figure 3). Because there were no required questions, response rates for each question varied.

Since the survey was distributed to the entire population, and not a random sample of the population, survey responses may be skewed towards researchers with a greater stake in data management activities. A chi-square goodness of fit test was calculated to determine if the sample proportions of UVM faculty college, rank, and gender were in the same proportions of those reported for the UVM faculty population. The test was conducted using $\alpha = 0.05$. As shown in Table 2, there was a statistically significant difference between the sample and the population for college ($n = 249$, $X^2 = 16.55$, $df = 7$, $p = 0.0205$), rank ($n = 252$, $X^2 = 11.61$, $df = 5$, $p = 0.0405$), and gender ($n = 252$, $X^2 = 9.56$, $df = 1$, $p = 0.002$). Faculty from the College of Arts and Sciences and the College of Education and Social Services were notably



over-sampled, while faculty from Grossman School of Business and the College of Medicine were under-sampled. As a result, the sample was not representative of the population, which may limit generalizability of the results to the campus.

Quantitative Data Analysis

RQ1. Data Management Activities

Survey questions were structured around data management activities based on the Data Lifecycle Model (DDI Alliance Structural Reform Group 2004) and the themes covered in the phase one qualitative research (Berman 2017). Questions included: types of data collected (Q2); data file size (Q4); generation and use of metadata (Q5); short-term (5 years or less) data storage (Q6); long-term (more than 5 years) data storage and preservation (Q8); data retention (Q9); data sharing practices (Q13) and limitations (Q14).

On average, respondents produced and collected 4.42 types of digital data, with a standard deviation of 2.49; full results of data types, by discipline, can be seen in Figure 4. Table 4 shows frequencies for data management activity variables, including metadata generation, digital data size, short-term data storage, long-term data storage and preservation, retention of digital data, and data sharing methods. Of respondents that do create metadata (N=50), seven indicated that they use known metadata standards, while the remaining 43 use a standard they devised. Seventeen survey respondents indicated they deposited data into repositories, notably GenBank, Protein Data Bank (PDB), the Long-Term Ecological Research Network (LTER), and the Gene Expression Omnibus (GEO). Analysis of these data management variables (Q4-Q9) and gender, rank, college, and discipline, produced no statistically significant differences.

Table 4: Data management activities variables

*Respondents were allowed to select multiple responses.

Value	Frequency	Percent
Q6. Metadata Generation		
Yes	50	28.1
No	128	71.9
TOTAL	178	
Q7. Digital Data Size		
1GB or less	37	22.0
More than 1GB, less than 100GB	67	39.9
More than 100GB, less than 1TB	34	20.2
More than 1TB, less than 100TB	28	16.7
More than 100TB, less than 1PB	1	0.6
More than 1PB	1	0.6
TOTAL	168	
Q8. Short-Term Data Storage Locations*		
Desktop or laptop hard drive	177	79.4
External hard drive or media	128	58.2
UVM network server	181	81.2
Third-party cloud storage service	67	30.0
Hard drive of instrument that generates data	53	23.8
TOTAL	223	
Q10. Long-Term Data Storage and Preservation Location (Always/Often)*		
External hard drive or media	141	64.7
UVM network server	174	79.8
Third-party cloud storage service	37	17.0
Institutional data repository	34	15.6
Discipline-specific data repository	20	9.2
Third-party data repository	12	5.5
Data are destroyed	22	10.1
TOTAL	218	
Q11. Retention of Digital Data		
Less than 1 year	2	0.9
1-4 years	26	12.0
5-10 years	59	27.3
More than 10 years	22	10.2
Indefinitely	107	49.6
TOTAL	216	
Q12. Data Sharing Methods (Always/Often)*		
Publications or presentations	126	50.0
Email or large file transfer	89	35.3
External hard drive or media	42	16.7
Personal website	10	4.0
Research group or project website	25	9.9
Collaborative web space	33	13.1
Institutional data repository	23	9.1
Discipline-specific data repository	17	6.7
Third-party data repository	10	4.0
I don't share data	10	4.0
TOTAL	252	

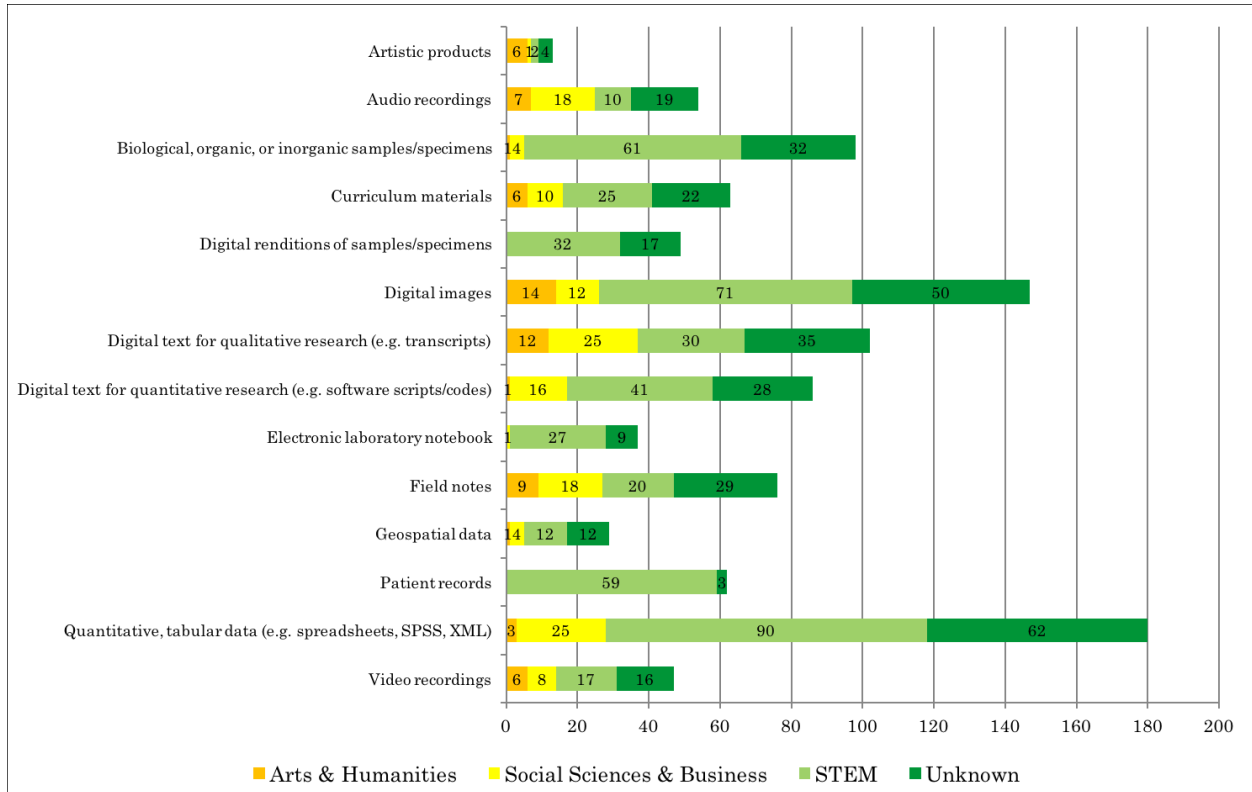


Figure 4: Q2. Which of the following best describe the types of data you have produced, or anticipate producing, as part of your research? Please choose all that apply. (N=276)

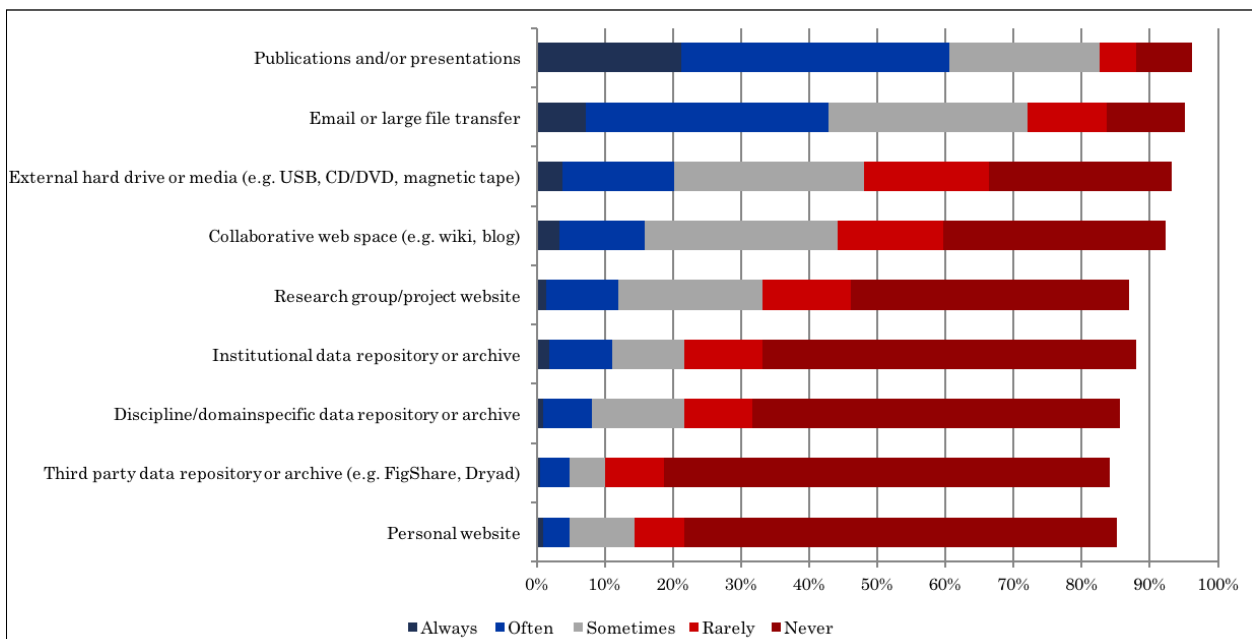


Figure 5: Q13. How often do you share your digital data with others (outside your research team) using the following methods (always, often, sometimes, rarely, never)? (N=208)

Figure 5 represents data sharing mechanisms (Q13), while Figure 6 shows limitations to sharing data (Q14). While the differences are not statistically significant, STEM faculty were three times more likely than other faculty to “always” or “often” share their data via discipline-specific or institutional data repositories. Each discipline faced different factors that impacted data sharing: for A&H, the top limitations were intellectual property concerns and the lack of time to make data available; for SS&B, the overwhelming concern was the ability to maintain confidentiality of research participants; while for STEM the lack of time, personnel, and tools/infrastructure to make data available were most limiting.

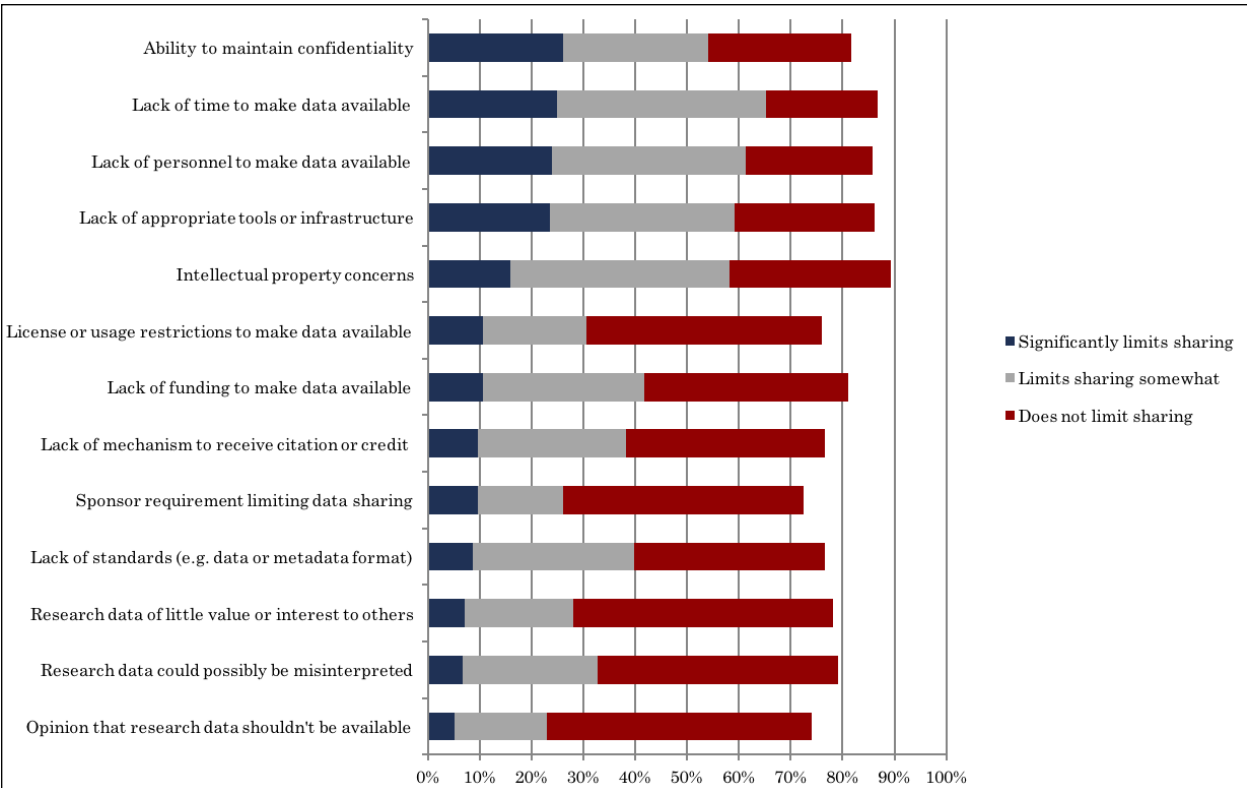


Figure 6: Q14. Please indicate how much each of the following factors limits the sharing of your research data (outside of your research team). (N=199)

Of total respondents, 109 (34.2%) received federal grants or contracts (Q15) and 61 (19.1%) have been required to submit at least one data management plan (DMP) (Q17). Of those who have submitted DMPs, 32 (52.5%) have submitted three or more, and 38 (62.3%) have had at least one DMP be part of a successful grant application. DMPs were most frequently submitted to the National Science Foundation and the National Institutes of Health, but other agencies included the U.S. Department of Energy, the U.S. Department of Agriculture, the U.S. Department of Education, the U.S. Department of Defense, NASA, and the National Institute of Justice.

RQ2. Data Management Challenges

Addressing the challenges or barriers research faculty face in managing their data, survey questions focused on specific activities related to data management (Q33). Survey respondents rated how easy or difficult activities were, including: storing data short- and long-term, backing-up data, ensuring data are secure, describing data, analyzing data, and sharing data; results are shown in Figure 7. Cross tabulations were calculated for Q33 and gender, rank, college, and discipline. A chi-square test of independence was performed to examine the relationship between how difficult a respondent found specific data management activities and their discipline. For the creation of metadata, 15.6% (N=5) of faculty in the A&H found this “difficult” or “somewhat difficult,” compared to 36.0% (N=9) in SS&B and 43.4% (N=46) in STEM fields. Using $\alpha = 0.05$, these differences are statistically significant $\chi^2(2, N = 163) = 8.158, p = 0.017$.

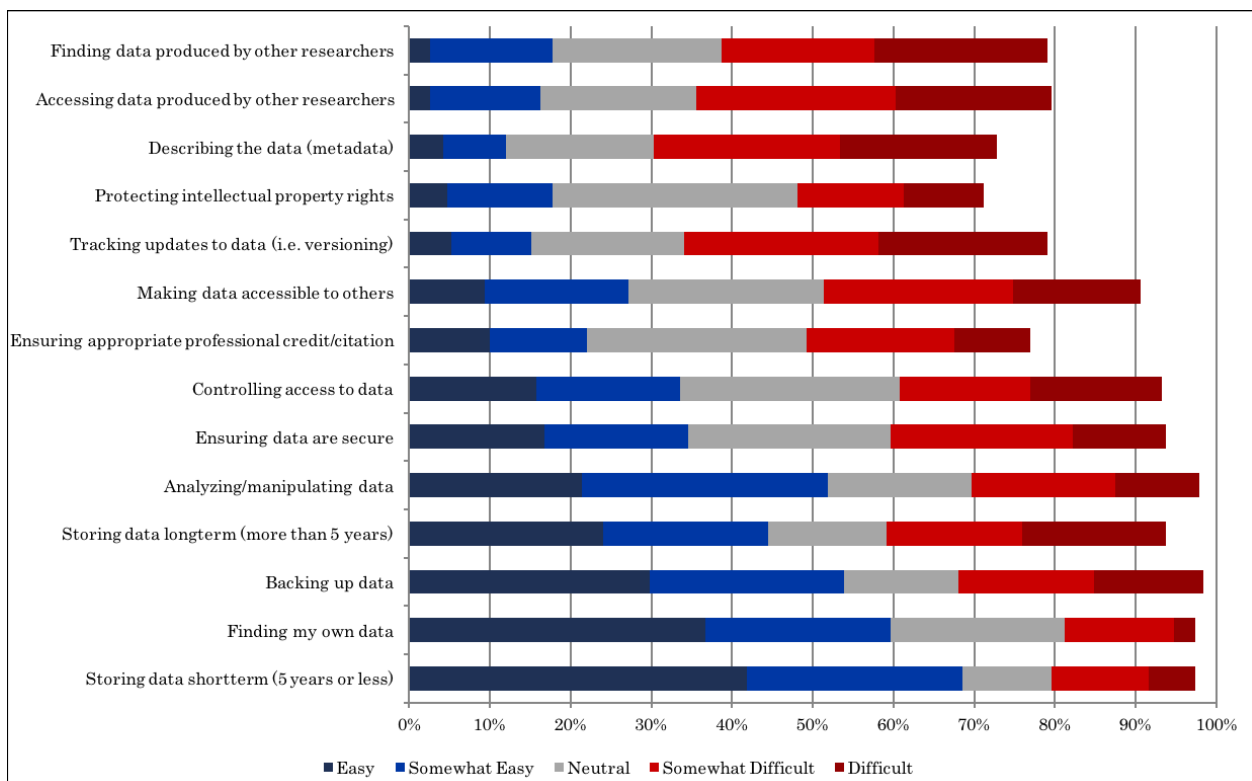
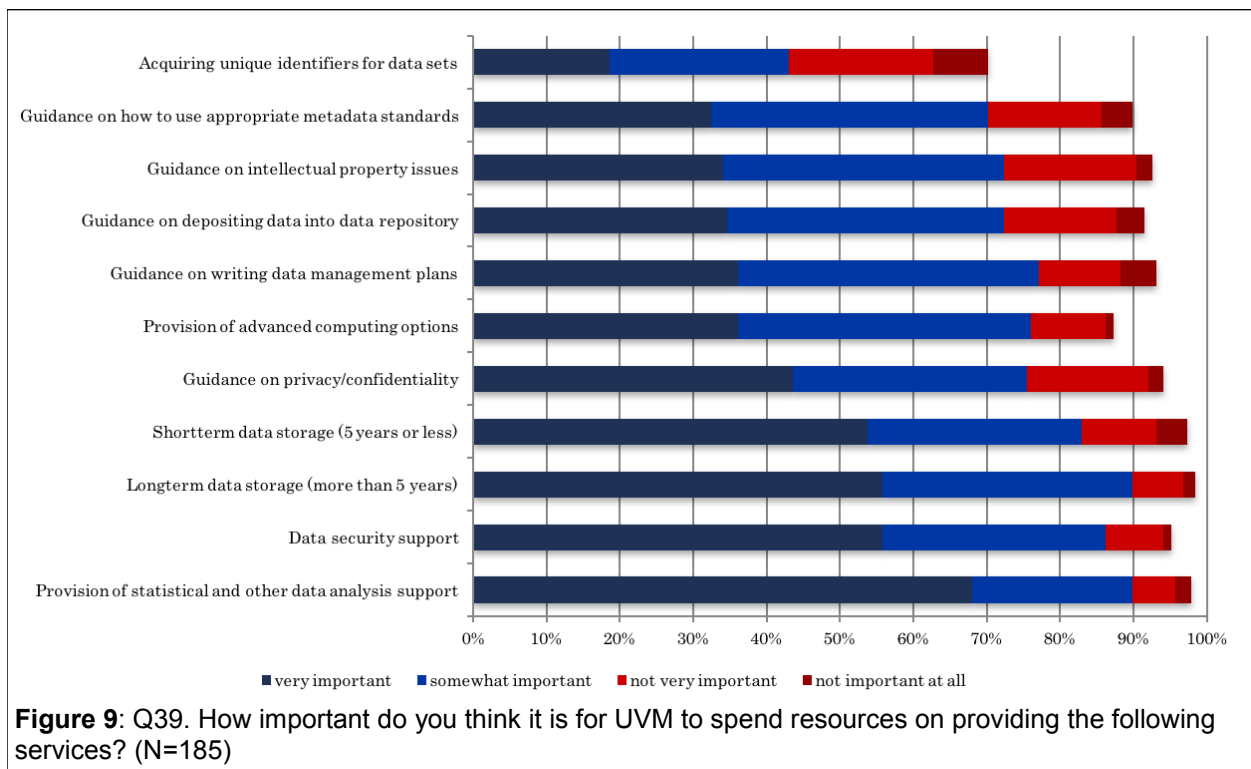
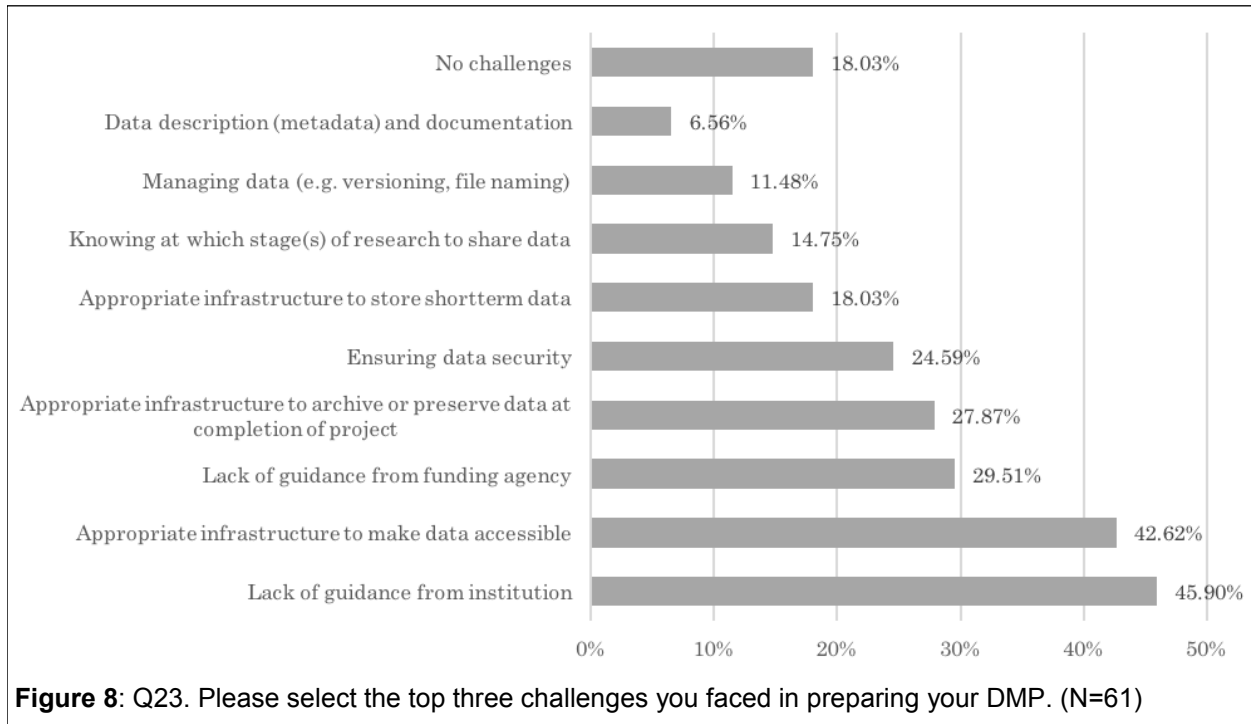


Figure 7: Q33. How easy or difficult is each of the following activities with regard to managing your UVM research data? (N=191)

A subset of survey questions focused specifically on guidance for (Q22) and challenges faced in creating data management plans (DMPs) (Q23). Of the 61 respondents who submitted a DMP, the majority (68.9%) did not receive guidance; those that did receive some form of assistance most frequently relied on the funding agency’s website. Researchers who have been required to submit at least one DMP were asked to rank the top three challenges they faced in preparing them; results are shown in Figure 8. While not statistically significant, survey respondents who have received a grant with an associated DMP were more likely to have no challenges with preparing DMPs.



RQ3. Institutional Support for Data Management

The survey asked respondents to rate how important it is for UVM to spend resources on specific research data services (Q39). The most popular answers for “very important” were: provision of statistical and other data analysis support (69.6%), data security support (58.7%), long-term data storage and preservation (56.8%), and short-term data storage (55.2%). Full responses can be seen in Figure 9. Cross tabulations were calculated for Q39 and gender, rank, college, and discipline. No statistically significant difference were found between Q39 and gender, rank, or college. However, statistically significant interactions were found between Q39 and discipline using a chi-square test of independence. Cramer’s V effect size was also calculated to understand the strength of the association. Results can be found in Table 5.

Table 5: Percentage of respondents who think it’s very important that UVM supports specific data services (Q36), by discipline

*p is significant at <0.05

	A&H	SS&B	STEM	Value	Sig.	Phi
Provision of advanced computing ^a	5 14.3%	4 12.9%	42 34.1%	9.186	0.010*	0.222
Provision of statistical and other data analysis support ^b	5 13.9%	24 66.7%	65 51.2%	22.309	0.000*	0.335
Short-term data storage (5 years or less) ^b	9 24.3%	15 42.9%	50 39.4%	3.362	0.186	0.130
Long-term data storage and preservation (more than 5 years) ^b	8 21.6%	17 48.6%	51 40.2%	6.109	0.047*	0.175
Acquiring unique identifiers for data sets ^c	0 0.0%	2 9.1%	21 20.0%	8.525	0.014*	0.232
Data security support ^d	6 17.1%	18 52.9%	51 40.5%	9.949	0.007*	0.226
Guidance on depositing data into discipline-specific data repository ^a	3 8.3%	9 30.0%	36 29.3%	6.841	0.033*	0.190
Guidance on how to use appropriate metadata standards ^e	5 13.5%	7 21.9%	32 26.9%	2.868	0.238	0.124
Guidance on writing DMPs ^f	4 10.8%	8 25.0%	34 27.6%	4.446	0.108	0.152
Guidance on intellectual property issues ^g	4 10.8%	9 27.3%	34 27.6%	4.558	0.102	0.154
Guidance on privacy and confidentiality ^d	4 10.8%	13 39.4%	42 33.6%	8.598	0.014*	0.210

a N=189; b N=199; c N=159; d N=195; e N=188; f N=192; g N=193

Survey respondents were also asked to rate their interest in data management support activities (Q40). The most popular answers were: provision of data management plan templates and tools (51.6%), data storage and preservation (50.0%), and an informational website with best practices and campus resources (46.9%). Full responses can be found in Figure 10. Analysis of this variable and gender, rank, college, and discipline produced no statistically significant differences, but the top ranked activities differed by discipline (Table 6).



Table 6: Highest rated data management support activities (Q40), by discipline

OVERALL	A&H	SS&B	STEM
DMP templates and tools	Data storage and preservation	DMP templates and tools	Data storage and preservation
Data storage and preservation	DMP consultation	Informational website	DMP templates and tools
Informational website	DMP templates and tools	Data storage and preservation	Informational website
Institutional data repository	Informational website	Institutional data repository	Institutional data repository
DMP consultation	Institutional data repository	Assistance meeting data sharing requirements	DMP consultation

Discussion

RQ1. How do faculty at UVM manage their research data, in particular how do they share and preserve data in the long-term?

In order to understand how faculty at UVM manage their research data, it was first important to understand the nature of the data they were working with. The survey results demonstrate that UVM faculty researchers don't collect just one or two types of data, but multiple data types depending on the research. Weller and Monroe-Gulick (2014) write: "The overlapping use of different research methodologies by single researchers forces the reconsideration of the typical view of academic researchers as a specialist in a specific type of research method. Instead, researchers are approaching their primary subject of study using a range of research methods" (478). It is helpful to understand this broader picture of who is generating what types of data in order to understand specific needs and/or challenges.

Despite being in the proclaimed 'era of Big Data' (Kitchin 2013), the majority of survey respondents (61.9%) collect small data, or data less than 100 GB; only a small percentage (17.9%) collect more than 1 terabyte (TB) of data for a single research project. This supports survey results indicating that respondents generally find it easy storing data in the short-term. UVM faculty use multiple locations to store their data in the short-term, with 86.1% of the survey respondents utilizing redundant systems for back-up. Similar to other studies (Diekema, Wesolek, and Walters 2014; Akers and Doty 2013; Whitmire, Boock, and Sutton 2015), network servers, computer hard drives, and external media were the most popular locations for both these activities.

The majority (87.1%) of digital data is being stored for five years or more in accordance with UVM policy (University of Vermont Sponsored Project Administration 2017), but issues of storage and preservation become more troublesome as researchers move from short- to long-term. Long-term storage locations echo short-term storage locations – namely campus network servers and external hard drives – but this can be problematic: "Trust in the department or university server for long-term storage may be misplaced, particularly if no formal agreements or practices are in place to curate data over time" (Scaramozzino, Ramírez, and McGaughey 2012, 361). Only a small percentage of data are going into institutional data repositories (15.6%), discipline-specific data repositories (9.2%), and third-party data repositories (e.g. FigShare) (5.5%), which are specifically designed for data preservation.

Metadata is essential to the identification, structuring, organization, and retrieval of data (Si et al. 2013). "Data sets that have metadata that conforms to a standard will be more interoperable with other data sets, more discoverable (by machines and by humans), and are likely to be more thoroughly documented compared to those that have an ad hoc schema" (Whitmire, Boock, and Sutton 2015, 394). 72.9% of survey respondents do not create metadata for their data, and of those who do generate metadata, an alarmingly small percentage self-identified as using a standardized metadata schema (N=7). These results, while more pronounced, are similar to other studies (Akers and Doty 2013; Diekema, Wesolek, and Walters 2014; Scaramozzino, Ramírez, and McGaughey 2012; Steinhart et al. 2012; Tenopir et al. 2011; Whitmire, Boock, and Sutton 2015; Qin and D'Ignazio 2010), suggesting a much larger issue among researchers. "There is a lack of awareness about the importance of metadata among the scientific community – at least in practice – which is a serious problem as

their involvement is quite crucial in dealing with problems regarding data management” (Tenopir et al. 2011, 20).

The survey data shows that respondents are willing to share their data with others, with only 4.8% indicating that they “always” or “often” don’t share data. Data sharing happens both through direct methods, a response to a specific request for data, and indirect methods, which provide unmediated access to data (e.g. data repositories). The most popular mechanism for sharing data is through publications or presentations, with 16.8% of respondents exclusively sharing data via this method. This approach to data sharing is not ideal in that the data shared through formal scholarship results in access to summarized and analyzed data, which is only a representation of the underlying primary data and is not the data itself. This indicates that “considerable confusion exists as to what ‘counts’ as data, even among researchers who are likely among their discipline’s experts” (Steinhart et al. 2012, 67).

RQ2. What challenges or barriers do UVM faculty face in effectively managing their research data?

When asked to reflect on their data management practices, activities that survey respondents found “easy” or “somewhat easy” included: storing short-term data (68.9%), backing up data (53.9%), and analyzing and manipulating data (51.8%). The top two activities that respondents found “difficult” or “somewhat difficult” were creating metadata to describe data (42.4%) and making data accessible to others (39.3%).

In terms of metadata, it is important to note that approximately one-fourth (27.2%) of respondents did not rate this activity, which suggests a larger issue of unfamiliarity with the concept of metadata; this is supported by the fact that only seven survey respondents indicated use of a standardized metadata schema. These results are supported by Scaramozzino, Ramírez, and McGaughey (2012), who found that only 20% of faculty at Cal Poly were aware of criteria for the creation of descriptive information for data. Interestingly, for the subset of respondents who have submitted at least one DMP, very few researchers indicated that they were challenged by metadata creation (6.56%). This may indicate that the explicit request for metadata in DMPs has heightened researcher awareness for the need to properly utilize standard data description, but it does not sufficiently explain the low usage of metadata standards overall.

One noteworthy finding: survey respondents found it “difficult” or “somewhat difficult” to both find (40.3%) and access (44.0%) data produced by other researchers. In thinking about accessibility in terms of metadata, Diekema, Wesolek, and Walters (2014) found that, while researchers utilize metadata to find data created by other researchers, “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” (323). This presents an interesting paradox: Faculty benefit from the utilization of standardized metadata, but do not directly address this issue when sharing their own data.

The top four factors that “significantly” limited the sharing of research data were: the ability to maintain confidentiality (25.6%), the lack of time to make data available (24.6%), the lack of personnel to make data available (23.6%), and the lack of appropriate tools or infrastructure to make data available (23.1%). For researchers who have submitted at least one DMP, the lack

of appropriate infrastructure was noted as a significant challenge by 42.6% of the respondents. While NSF guidelines allow for the costs associated with data management to be included in proposal budgets, Scaramozzino, Ramírez, and McGaughey (2012) found that faculty rarely accounted for these costs in their grant applications, while Steinhart et al (2012) reported that the size of the grants did not increase to cover data management expenses even when the costs were included. Several studies found that the perceived effort required to share data was a notable limitation (Campbell et al. 2002; Foster and Gibbons 2005; Tenopir et al. 2011). In particular, this lack of time, personnel, tools, and infrastructure to effectively share data could be positively impacted through greater direct support of data management on the UVM campus, reducing the burden on individual researchers.

RQ3. What institutional data management support or services are UVM faculty interested in?

The results of this survey suggest several areas in which UVM could strategically be allocating resources to support data management activities. A high percentage of respondents were interested in data management plan templates and tools (51.7%) and an informational website with best practices/campus resources (46.9%), both of which would provide indirect support for the management of research data and address the explicit needs of faculty submitting DMPs. More in-depth supports, including data management workshops and data management consultations, were surprisingly not perceived as important areas for UVM to support (27.1% and 33.9%, respectively). These needs differ than those identified at other institutions (Diekema, Wesolek, and Walters 2014; Parham, Bodnar, and Fuchs 2012; Akers and Doty 2013; Steinhart et al. 2012; Weller and Monroe-Gulick 2014), suggesting differences in institutional context and reinforcing the need for local environmental scans to understand researcher practices.

Surprisingly, 69.6% of survey respondents found provision of statistical and data analysis support to be “very important;” this finding was unexpected due to lack of coverage in the existing literature. UVM offers a free Statistical Consulting Clinic for faculty and students, which provides a range of services across all stages of research; however questions arise about whether this service is known to researchers, or whether this service meets all researchers’ needs. Additionally, data security support (58.7%), long-term data storage and preservation (56.8%), and short-term data storage (55.2%) were all seen as important activities for UVM to actively support. These results are noteworthy in that they reflect activities that faculty generally don’t find difficult; one possible interpretation is that respondents are demonstrating the need for UVM to maintain these services. What becomes unclear from the results is faculty’s interpretation of ‘long-term data storage and preservation.’ The results suggest that faculty may simply see this as an extension of short-term data storage – the simplest form of keeping data – as opposed to data preservation, which takes into account factors such as ongoing maintenance and data obsolescence. Multiple understandings of ‘long-term data storage and preservation’ is suggested by survey results: only 36.4% of researchers found the creation of an institutional data repository as very important, despite its ability to help facilitate the preservation – long-term storage – of research data. In fact, studies have found that the availability of data repositories – institutional, organization, or disciplinary – have been an important factor influencing data sharing behavior (Choudhury 2008; Cragin et al. 2010; Witt 2008).

Conversely, guidance for the creation of metadata received very little support in the survey, despite the paucity of standardized metadata formats in use and its identification as a top challenge for researchers. Only 32.1% of survey respondents felt it was “very important” for UVM to provide guidance on appropriate use of metadata and 13.5% were interested in metadata support. This inconsistency again supports the notion that researchers do not understand the need for and importance of metadata for long-term data preservation and data sharing, and suggests that metadata education represents a significant area for outreach, even if researchers are not self-identifying it as a need.

The chi-square test of independence for Q39, “How important do you think it is for UVM to spend resources on providing the following services?,” demonstrated several statistically significant differences between disciplines: STEM faculty found provision of advanced computing and acquiring unique identifiers (e.g. DOIs) more important than the other disciplines, while SS&B faculty found data security support more important than other disciplines. Both STEM and SS&B faculty found several additional activities more important than researchers in the A&H, including: provision of statistical and other data analysis support, long-term data storage and preservation, guidance on depositing data into discipline-specific data repositories, and guidance on privacy and confidentiality. These results emphasize the differences between research in the sciences/social sciences and the humanities. It also may represent a limitation of the survey instrument itself. Wording of the first screening question (Q1) asked: “Data is any recorded material necessary to validate your research. This can be numeric data, textual data, images, audio or video files, artifacts, etc. Do you collect, generate, or use data in your research?” The exclusive use of the word ‘research’ – as opposed to ‘research and scholarship’ – may have negatively biased A&H faculty from seeing themselves in this study or engaging with the survey, therefore underrepresenting the activities they feel it’s important for UVM to support.

Conclusion

The purpose of this article was to report on the second phase of an exploratory sequential mixed methods research study aimed at understanding researchers’ data management behaviors, including barriers or challenges they face, with the intention of developing appropriate research data services and support at the University of Vermont. The goal in using a survey was to easily collect data to characterize the data management practices of UVM faculty across all disciplines. While the sample was not representative of the population, thus limiting the generalizability of the results to the broader UVM campus, the data obtained are informative for the pragmatic aims of this research study.

Disciplinary differences in data management behavior have been noted in previous literature (e.g. Akers and Doty 2013; Witt et al. 2009; Jahnke, Asher, and Keralis 2012), although most of these studies did not test for statistical significance between groups. Analysis across multiple demographic factors, including gender, rank, and discipline showed differences in behaviors, challenges, and interests, but the majority of these differences were not significant. In part, this may be attributable to the variety of qualitative and quantitative data that researchers collect that transcend discipline or epistemological orientation. Regardless of the significance, or lack thereof, of these differences, it is clear that any future data management services provided by UVM will need to address a variety of needs.

Although not tested for statistical significance because of the small sample size, one factor that differentiated data management behaviors among faculty was whether or not a faculty member had been required to submit a data management plan. The faculty who had submitted DMPs were more attuned to the official language of the data management lifecycle, including a greater awareness of metadata, data sharing, and data preservation. Some studies have found that federal data sharing policies have created peer-pressure for researchers to share and curate data (McCullough, McGeary, and Harrison 2008; Piwowar and Chapman 2010). This study was conducted four years after the first researchers were required to submit formal DMPs to the National Science Foundation; it is reasonable to assume that as more faculty are exposed to the requirements in DMPs, there will be a greater sense of awareness about the entire data curation process.

This study was designed as mixed methods research, combining the qualitative and the quantitative data to provide a more holistic understanding of data management practices, challenges, and opportunities at UVM. In order to adhere to a rigorous mixed methods research design, and to develop appropriate research data services at the University of Vermont, it is necessary to integrate and analyze the combined results of both the survey and the phase one qualitative research (Berman 2017). This interpretation-level linking of data (Fetters, Curry, and Creswell 2013; Creswell and Plano Clark 2011) will allow the formulation of meta-inferences, or “an overall conclusion, explanation or understanding developed through and integration of the inferences obtained from the qualitative and quantitative strands of a mixed method study” (Teddlie and Tashakkori 2008, 300). It is in these meta-inferences, in which qualitative and quantitative data will be compared and contrasted, and will ultimately address the purpose of the study: to understand researchers’ current data management behaviors, challenges, and preferences, in order to develop research data services at the University of Vermont.

Supplemental Content

Appendix

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2017.1098> under “Additional Files”.

Acknowledgements

The author thanks UVM faculty colleagues Bernice Garnett, assistant professor, Department of Education, Jason Parker, assistant research professor, Department of Plant & Soil Science, and Alan Howard, Director of the Statistical Consulting Clinic, for their advice and feedback at multiple stages of this study.

Disclosure

The author reports no conflict of interest.

References

Ajzen, Icek. 1991. “The Theory of Planned Behavior.” *Organizational Behavior and Human Decision Processes* 50(2): 179-211. [http://www.dx.doi.org/10.1016/0749-5978\(91\)90020-T](http://www.dx.doi.org/10.1016/0749-5978(91)90020-T)

———. 2005. *Attitudes, Personality and Behavior*. 2nd ed. Berkshire, UK: Open University Press.

Ajzen, Icek, and Martin Fishbein. 2000. "Attitudes and the Attitude-Behavior Relation: Reasoned and Automatic Processes." In *European Review of Social Psychology*, edited by W. Stroebe and M. Hewstone, 1-33. Chichester, UK: Wiley.

Akers, Katherine G., and Jennifer Doty. 2013. "Disciplinary Differences in Faculty Research Data Management Practices and Perspectives." *International Journal of Digital Curation* 8(2): 5-26.
<http://www.dx.doi.org/10.2218/ijdc.v8i2.263>

Association of Research Libraries. 2016. "White House Directive on Public Access to Federally Funded Research and Data." Accessed May 26. <http://www.arl.org/focus-areas/public-access-policies/federally-funded-research/2696-white-house-directive-on-public-access-to-federally-funded-research-and-data#agency-policies>

Berman, Elizabeth A. 2017. "An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Findings from the Qualitative Phase." *Journal of eScience Librarianship* 6(1): e1097. <http://www.dx.doi.org/10.7191/jeslib.2017.1097>

Campbell, Eric G, Brian R Clarridge, Manjusha Gokhale, Lauren Birenbaum, Stephen Hilgartner, Neil A Holtzman, and David Blumenthal. 2002. "Data Withholding in Academic Genetics: Evidence from a National Survey." *Journal of the American Medical Association* 287(4): 473-480. <http://www.dx.doi.org/10.1001/jama.287.4.473>

Choudhury, G Sayeed. 2008. "Case Study in Data Curation at Johns Hopkins University." *Library Trends* 57(2): 211-220. <http://www.dx.doi.org/10.1353/lib.0.0028>

Council on Library and Information Resources. 2016. "Data Curation." Accessed May 27.
<http://www.clir.org/initiatives-partnerships/data-curation>

Cragin, Melissa H., Carole L. Palmer, Jacob R. Carlson, and Michael Witt. 2010. "Data Sharing, Small Science and Institutional Repositories." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368(1926): 4023-4038. <http://www.dx.doi.org/10.1098/rsta.2010.0165>

Creswell, John W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks, CA: SAGE Publications.

Creswell, John W., and Vicki L. Plano Clark. 2011. *Designing and Conducting Mixed Methods Research*. SAGE Publications. 2nd ed. Thousand Oaks, CA: SAGE Publications.

D'Ignazio, John, and Jian Qin. 2008. "Faculty Data Management Practices: A Campus-Wide Census of STEM Departments." *Proceedings of the American Society for Information Science and Technology* 45(1) 1-6.
<http://www.dx.doi.org/10.1002/meet.2008.14504503139>

DDI Alliance Structural Reform Group. 2004. "DDI Combined Life Cycle Model." Created June 10.
<http://www.ddialliance.org/system/files/Concept-Model-WD.pdf>

Diekema, Anne R., Andrew Wesolek, and Cheryl D. Walters. 2014. "The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories." *Journal of Academic Librarianship* 40(3-4): 322-331. <http://www.dx.doi.org/10.1016/j.acalib.2014.04.010>

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian. 2008. *Internet, Mail, and Mixed Mode Surveys: The Tailored Design Method*. 3rd ed. New York: John Wiley.

Fetters, Michael D., Leslie A. Curry, and John W. Creswell. 2013. "Achieving Integration in Mixed Methods Designs - Principles and Practices." *Health Services Research* 48(6pt2): 2134-2156.
<http://www.dx.doi.org/10.1111/1475-6773.12117>

Flanders, Julia, and Trevor Munoz. 2014. "An Introduction to Humanities Data Curation." *Digital Humanities Data Curation*. <http://guide.dhcurator.org/contents/intro>

- Foster, Nancy Fried, and Susan Gibbons. 2005. "Understanding Faculty to Improve Content Recruitment for Institutional Repositories." *D-Lib Magazine* 11(1). <http://www.dx.doi.org/10.1045/january2005-foster>
- Fowler, Floyd J. 2014. *Survey Research Methods*. 5th ed. Thousand Oaks, CA: SAGE Publications.
- Jahnke, Lori, Andrew Asher, and Spencer D C Keralis. 2012. "The Problem of Data." *Council on Library and Information Resources*. <http://www.clir.org/pubs/reports/pub154>
- Kitchin, Rob. 2013. "Big Data and Human Geography: Opportunities, Challenges and Risks." *Dialogues in Human Geography* 3(3): 262-267. <http://www.dx.doi.org/10.1177/2043820613513388>
- Madans, Jennifer, Kristen Miller, Aaron Maitland, and Gordon B. Willis. 2011. *Question Evaluation Methods*. New York: John Wiley.
- McCullough, B. D., Kerry Anne McGeary, and Teresa D. Harrison. 2008. "Do Economics Journal Archives Promote Replicable Research?" *Canadian Journal of Economics* 41(4): 1406-1420. <http://www.dx.doi.org/10.1111/j.1540-5982.2008.00509.x>
- McLure, Merinda, Allison V. Level, Catherine L. Cranston, Beth Oehlerts, and Mike Culbertson. 2014. "Data Curation: A Study of Researcher Practices and Needs." *Portal: Libraries and the Academy* 14(2): 139-164. <http://www.dx.doi.org/10.1353/pla.2014.0009>
- National Institutes of Health. 2003. "Final NIH Statement on Sharing Research Data." Published February 26. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
- Nulty, Duncan D. 2008. "The Adequacy of Response Rates to Online and Paper Surveys: What Can Be Done?" *Assessment & Evaluation in Higher Education* 33(3): 301-314. <http://www.dx.doi.org/10.1080/02602930701293231>
- Office of Science and Technology Policy. 2013. "Increasing Access to the Results of Federally Funded Scientific Research." Published February 22. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
- Parham, Susan Wells, Jon Bodnar, and Sara Fuchs. 2012. "Supporting Tomorrow's Research: Assessing Faculty Data Curation Needs at Georgia Tech." *College & Research Libraries News* 73(1): 10-13. <http://crln.acrl.org/cgi/content/abstract/73/1/10>
- Pennock, Maureen. 2007. "Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information." *Library and Archives Journal* 1(January): 1-3. http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf
- Piwowar, Heather A., and Wendy W. Chapman. 2010. "Public Sharing of Research Datasets: A Pilot Study of Associations." *Journal of Informetrics* 4(2): 148-156. <http://www.dx.doi.org/10.1016/j.joi.2009.11.010>
- Presser, Stanley, Jennifer M. Rothgeb, Mitch P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. 2004. *Methods of Testing and Evaluating Survey Questionnaires*. Hoboken NJ: John Wiley.
- Qin, Jian, and John D'Ignazio. 2010. "The Central Role of Metadata in a Science Data Literacy Course." *Journal of Library Metadata* 10(2-3): 188-204. <http://www.dx.doi.org/10.1080/19386389.2010.506379>
- Scaramozzino, Jeanine Marie, Marisa L. Ramírez, and Karen J. McGaughey. 2012. "A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University." *College & Research Libraries* 73(4): 349-365. <http://www.dx.doi.org/10.5860/crl-255>
- Si, Li, Xiaozhe Zhuang, Wenming Xing, and Weining Guo. 2013. "The Cultivation of Scientific Data Specialists: Development of LIS Education Oriented to E-Science Service Requirements." *Library Hi Tech* 31(4): 700-724. <http://www.dx.doi.org/10.1108/LHT-06-2013-0070>
- Singleton, Royce A., and Bruce C. Straits. 2010. *Approaches to Social Research*. 5th ed. New York: Oxford University Press.

Steinhart, Gail, Eric Chen, Florio Arguillas, Dianne Dietrich, and Stefan Kramer. 2012. "Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements." *Journal of eScience Librarianship* 1(2): e1008. <http://www.dx.doi.org/10.7191/jeslib.2012.1008>

Teddlie, Charles, and Abbas Tashakkori. 2008. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Techniques in the Social and Behavioral Sciences*. Thousand Oaks, CA: SAGE Publications.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. "Data Sharing by Scientists: Practices and Perceptions." *PLoS ONE* 6(6): e21101. <http://www.dx.doi.org/10.1371/journal.pone.0021101>

The Carnegie Classification of Institutions of Higher Education. 2017. "Methodology." Accessed February 27. <http://carnegieclassifications.iu.edu/methodology/basic.php>

University of Vermont. 2017. "UVM Facts." Accessed February 22. https://www.uvm.edu/uvm_facts

University of Vermont Sponsored Project Administration. 2017. "SPA Home: University of Vermont. Data Acquisition, Management, Sharing and Ownership." Accessed February 27. <http://www.uvm.edu/spa/?Page=dataacquisition.html>

Weller, Travis, and Amalia Monroe-Gulick. 2014. "Understanding Methodological and Disciplinary Differences in the Data Practices of Academic Researchers." *Library Hi Tech* 32(3): 467-482. <http://www.dx.doi.org/10.1108/LHT-02-2014-0021>

Whitmire, Amanda L., Michael Boock, and Shan C. Sutton. 2015. "Variability in Academic Research Data Management Practices: Implications for Data Services Development from a Faculty Survey." *Program: Electronic Library and Information Systems* 49(4): 382-407. <http://www.dx.doi.org/10.1108/PROG-02-2015-0017>

Williford, Christa, and Charles Henry. 2012. "One Culture Computationally Intensive Research in the Humanities and Social Sciences." *Council on Library and Information Resources Publication* 151. <http://www.clir.org/pubs/reports/pub151/pub151.pdf>

Witt, Michael. 2008. "Institutional Repositories and Research Data Curation in a Distributed Environment." *Library Trends* 57(2): 191-201. <http://www.dx.doi.org/10.1353/lib.0.0029>

Witt, Michael, Jacob Carlson, D. Scott Brandt, and Melissa H. Cragin. 2009. "Constructing Data Curation Profiles." *International Journal of Digital Curation* 4(3): 93-103. <http://www.dx.doi.org/10.2218/ijdc.v4i3.117>