



Full-Length Paper

An Exploratory Sequential Mixed Methods Approach to Understanding Researchers' Data Management Practices at UVM: Findings from the Qualitative Phase

Elizabeth A. Berman

Tufts University, Medford, MA, USA

*Formerly Library Associate Professor, University of Vermont

Abstract

The objective of this article is to report on the first qualitative phase of an exploratory sequential mixed methods research design focused on researcher data management practices and related institutional research data services. The aim of this study is to understand data management behaviors of faculty at the University of Vermont (UVM), a higher-research activity Research University, in order to guide the development of campus research data management services. The population of study was all faculty who received National Science Foundation (NSF) grants between 2011 and 2014 who were required to submit a data management plan (DMP); qualitative data was collected in two forms: (1) semi-structured interviews and (2) document analysis of data management plans. From a population of 47 researchers, six were included in the interview sample, representing a broad range of disciplines and NSF Directorates, and 35 data management plans were analyzed. Three major themes were identified through triangulation of qualitative data sources: data management activities, including data dissemination and data sharing; institutional research support and infrastructure barriers; and perceptions of data management plans and attitudes towards data management planning. The themes articulated in this article will be used to design a survey for the second quantitative phase of the study, which will aim to more broadly generalize data management activities at UVM across all disciplines.

Correspondence: Elizabeth A. Berman: elizabeth.berman@tufts.edu

Keywords: data management, mixed methods research, qualitative research, research data services, academic libraries, interview, document analysis

Rights and Permissions: Copyright Berman © 2017



All 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 Open Access (OA) movement, which began nearly two decades ago, is a response to traditional publishing practices that restrict access to academic research via expensive publisher subscription models. According to SPARC, “Open Access is the free, immediate, online availability of research articles coupled with the rights to use these articles fully in the digital environment. Open Access ensures that anyone can access and use these results – to turn ideas into industries and breakthroughs into better lives” (SPARC 2016a, para. 1). OA pushes against the dominant for-fee publishing paradigm by promoting unrestricted online access to peer-reviewed scholarly research either through open-access journals or through self-archiving in digital or institutional repositories. Following in the steps of Open Access, Open Data is “research data that is freely available on the internet permitting any user to download, copy, analyze, re-process, pass to software or use for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself” (SPARC 2016b, para. 1).

The purpose of open data allows any researcher to discover, use, and interpret the data to test credibility and reproducibility of research results, a key tenet in scholarly discourse, while also ensuring a higher level of data integrity to guard against research misconduct (Coates and Konkiel 2013; Tenopir et al. 2011; SPARC 2016b). Proponents argue that open data also helps accelerate scientific breakthroughs and innovation, promotes entrepreneurship, and enhances economic growth (Office of Science and Technology Policy 2013). Piwowar, Day, and Fridsma (2007) discovered that in addition to benefiting the scientific community, researchers who share data also benefit directly by increased citations of their work, enhancing their scholarly impact. A more practical purpose for openness is the simple fact that digital data is fragile, more prone to loss and corruption than physical data, and researchers need to be more thoughtful and proactive to ensure its preservation (Bracke 2011).

In 2003, the National Institutes of Health (NIH) issued a policy statement about data sharing: “NIH reaffirms its support for the concept of data sharing. We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health. The NIH endorses the sharing of final research data to serve these and other scientific goals” (National Institutes of Health 2003, para. 2). As a result of this policy, investigators submitting an NIH grant for \$500,000 or more are expected to include a plan for data sharing, or an explanation as to why data sharing is not possible.

In 2011, the NSF took open data a step further by requiring that researchers submit formal data management plans with their grant applications to account for how data will be made accessible and preserved in the long-term. This requirement affirms NSF’s commitment to advancing science and the interests of the public through dissemination and sharing of research data and research products (National Science Foundation 2014). Despite this requirement, the mandate comes with limited or vague information for researchers. For example, the Data Management for NSF Engineering Directorate Proposals and Awards (2011) states, “Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements,” and “The DMP should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data” (3). With lack of information comes a lack of technical or structural support from NSF. And while NSF

guidelines allow for the costs of data management to be included in budgets, studies have found that faculty rarely account for these costs (Scaramozzino, Ramírez, and McGaughey 2012), and when they do, data management expenses tend to be the first item cut from the budget (Steinhart et al. 2012; Bardin, Resnick, and Camina 2012). As a result, researchers have turned to their own institutions for guidance in supporting their data management activities.

In order to develop appropriate services, many institutions have been conducting needs assessments using qualitative research, a research design focused on exploring and gaining a deep understanding of the meaning individuals or groups ascribe to a phenomenon, setting, event, etc., through the collection of non-numeric data, including text, photographs, and audio (Creswell 2014; Berg and Lune 2012; Tracy 2013; Marshall and Rossman 2011). The purpose of these needs assessments is to better understand how researchers are currently managing data in order to design and build infrastructure and services to support data management activities (Walters 2009), including data curation, or “the active management and appraisal of digital information over its entire life cycle” (Pennock 2007, para. 2). Methods represented in the literature include interviews, focus groups, and document analysis (Table 1).

Table 1: Comparison of methods used in data management needs assessments.

*Multiple methods used in single study

Method	Author(s)	Institution	Sample Size
Interviews	Diekmann (2012)	The Ohio State University	14 participants
	Lage, Losoff, and Maness (2011)	University of Colorado Boulder	26 participants
	Marcus <i>et al</i> (2007)*	University of Minnesota	7 participants
	Peters and Dryden (2011)	University of Houston	10 participants
	Walters (2009)	Georgia Institute of Technology	5 participants
	Westra (2010)	University of Oregon	25 participants
	Williams (2013)	University of Illinois at Urbana-Champaign	7 participants
	Witt <i>et al</i> (2009)	Purdue University University of Illinois at Urbana-Champaign	19 participants
Focus Groups	Bardin, Resnick, and Camina (2012)	University of California Los Angeles	2 groups 8 participants
	Marcus <i>et al</i> (2007)*	University of Minnesota	18 groups 65 participants
	Mattern <i>et al</i> (2015)	University of Pittsburgh	2 groups 8 participants
	McLure <i>et al</i> (2014)	Colorado State University	5 groups 31 participants
Document Analysis	Mischo, Schlembach, and O'Donnell (2014)	University of Illinois at Urbana-Champaign	1,260 documents
	Parham <i>et al</i> (2016)	multi-institution	500 documents
	Parham and Doty (2012)	Georgia Institute of Technology	181 documents

A systematized qualitative approach, the Data Curation Profiles (DCP) interview protocol, was developed as part of a toolkit for researchers to use for “exploring, learning about and interacting with data producers and collecting information about datasets and collections” (Witt et al. 2009, 101). It also can be used, “to help gather information to make local data development policies and selection and deselection decisions” (102). Another structured approach to gathering qualitative data was the E-Science Institute, created by the Association of Research Libraries (ARL) and the Council on Library and Information Resources (CLIR) Digital Library Federation (DLF), which focused on assisting academic and research libraries develop a plan for eResearch support, or support for data-intensive scientific research using advanced computational methods. This program included a baseline assessment of institutional engagement with eScience, which included conducting interviews and focus groups with key stakeholders, both internal and external to the libraries (DuraSpace Organization 2016).

Other researchers have developed their own interview procedures, with similar aims but different outcomes. As a result of interviews with thirteen Principal Investigators (PIs) of NSF and NIH grants, Peters and Dryden (2011) proposed the creation of a Data Working Group, comprised of multiple stakeholders, to develop, promote, and communicate campus data support, in particular campus infrastructure needs. Diekmann (2012) conducted in-depth interviews with 14 faculty in the agricultural sciences in order to understand the non-institutional data management infrastructure available. Westra (2010) interviewed 25 scientists with the tangible outcome of identifying researchers to pilot data curation projects. Lage, Losoff, and Maness (2011) conducted interviews with STEM faculty in order to create personas, or “fictionalized aggregates” of potential researchers (918), to demonstrate a variety of factors that influence interest in data management support services. These included available support for data management activities (funding, personnel, time), storage issues, and attitudes towards data sharing.

Focus groups are an extension of the individual interview, in which a facilitator leads a group discussion on a particular topic or topics to unpack motivations, decisions, and priorities (Berg and Lune 2012). Marcus et al (2007) conducted a series of interviews and focus groups at the University of Minnesota, and concluded: “Researchers’ practices regarding data curation and preservation are idiosyncratic, haphazard, and in great need of attention” (10). As a result, the University of Minnesota Libraries strategically focused on supporting researchers to organize and preserve their data. Focus group participants in Mattern et al (2015) demonstrated, “a research process that was more personal and, in most cases, more imperfect than the research lifecycle models that academic libraries are increasingly using for RDS [research data services] development and communication” (421). McLure et al (2014) elicited from participants “self-perceived deficits in skill and knowledge” (156), which emphasized the need for a variety of data support services.

Document analyses, the “careful, detailed, systematic examination and interpretation of a particular body of material in an effort to identify patterns, themes, biases, and meanings” (Berg and Lune 2012, 349), were also represented in the literature through three studies evaluating the content of NSF data management plans (DMPs). Parham et al (2016) used a common rubric to assess the content of DMPs in order to understand how researchers from different fields and different institutions are managing, sharing, and archiving their data. Mischo, Schlembach, and O’Donnell (2014) were able to identify data repositories that

researchers were using to meet NSF requirements, while Parham and Doty (2012) discovered the infrequent mention of Georgia Tech’s institutional data repository through their analysis, as well as a high incidence of duplicate text across multiple, unrelated data management plans.

Despite the similar goals of these research studies, the specific research questions addressed varied widely and the findings are difficult to generalize or apply to other researchers or institutions. There is a tacit understanding that, “local studies can inform libraries and librarians about the behaviors, needs, interests, and concerns of researchers at individual institutions” (McLure et al. 2014, 158). Thus, employing qualitative methods builds a foundation for understanding the particular institutional context and needs of researchers at UVM by using qualitative methods, while also situating the study within the larger narrative of data management practices and research data services to understand more broadly how institutional contexts are similar or different.

Purpose Statement

This article reports on the first phase of an exploratory sequential mixed methods research design aimed at understanding data management behaviors of faculty at the University of Vermont (UVM). Mixed methods research (MMR) draws on the strengths of both qualitative and quantitative research, resulting in “multiple ways of seeing and hearing” (Greene 2007, 20) data. In an exploratory design, qualitative data is first collected and analyzed, and themes are used to drive the development of a quantitative instrument to further explore the research problem. Results of these two phases are then integrated and linked to develop a more nuanced understanding of the research question (Creswell and Plano Clark 2011; Teddlie and Tashakkori 2008; Onwuegbuzie, Bustamante, and Nelson 2010). The exploratory sequential mixed methods design (Figure 1) was selected in order to develop a more generalized understanding of data management activities at UVM rooted in researchers’ own lived experiences (Creswell 2014).

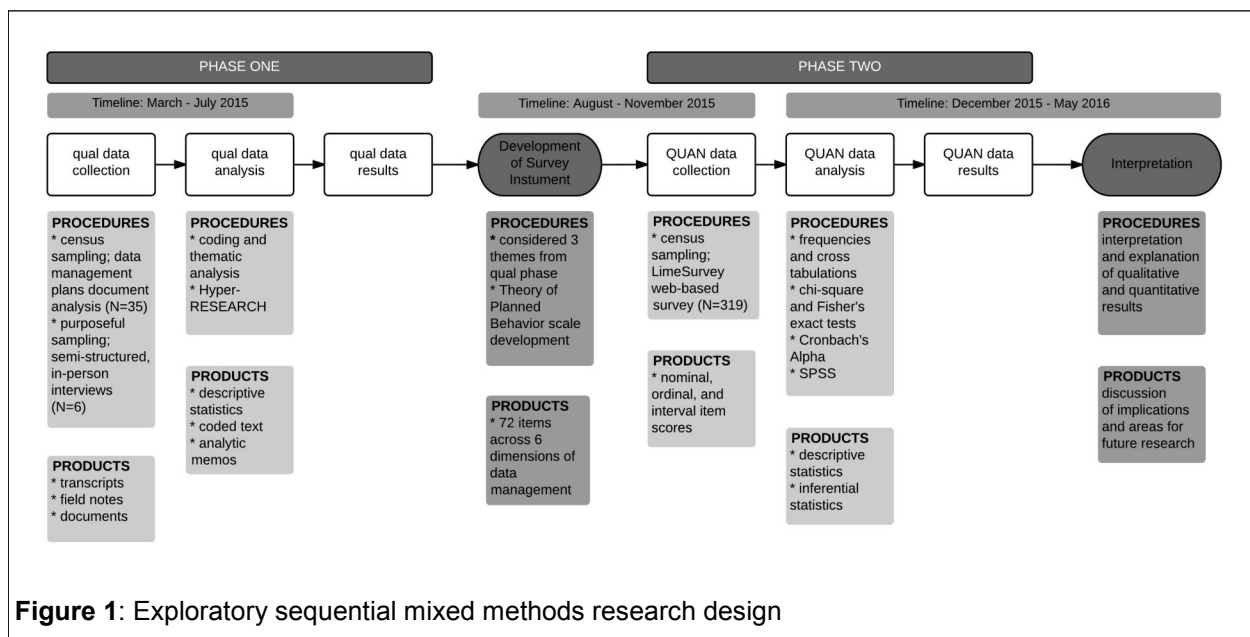


Figure 1: Exploratory sequential mixed methods research design

In the first phase of this MMR study, qualitative data was collected from UVM faculty who received NSF grants between 2011 and 2014 who had submitted a data management plan (DMP). The focus specifically on NSF grantees is a direct result of the NSF requirement, which has proven to be a key stimulus for academic institutions to begin addressing research data services (Fearon et al. 2013). A document analysis of successful DMPs was coded for emergent themes related to data management planning. In-person, semi-structured interviews were also conducted with a purposeful sample of this population to gather more in-depth information about data management planning, including data management activities (e.g. creation and use of metadata; short-term storage of data; long-term data storage and preservation; data sharing practices) and related challenges and issues of institutional support. The qualitative phase of this research was guided by the following research questions:

RQ1: How do faculty at UVM manage their research data, with a particular focus on data sharing and long-term data preservation?

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

The primary objective of the first phase of this pragmatic MMR research study is to gain an in-depth understanding of the current data management behaviors of a strategic sample of UVM researchers who have been actively required to address data management planning. The results of this phase will be used as a foundation for creating a survey for the second, quantitative phase of the study, which will be deployed to all research faculty at UVM in order to better generalize data management behaviors to the entire campus population. The qualitative and quantitative data will then be integrated in order to guide the development of research data services at UVM.

Methods

Qualitative Population

The University of Vermont (UVM) is a higher research activity Research University (The Carnegie Classification of Institutions of Higher Education 2017) with a student enrollment of 12,000 undergraduate and graduate students and a faculty of 1,200 (University of Vermont 2017b). Statistics provided by the Office for Vice President of Research demonstrate UVM's research impact: in FY 2014, UVM received 615 grants and contracts amounting to \$128 million; 70% of that funding was through federal grants (University of Vermont 2017a). During the qualitative research phase, the target population was all UVM faculty who received National Science Foundation (NSF) grants between 2011 and 2014, who were required to submit a data management plan (DMP).

Working with the UVM Office of Sponsored Programs Administration, a list was generated including 47 researchers who met the criteria. To achieve the recommended minimum sample size of 6-12 interview participants (Onwuegbuzie and Collins 2007), a critical case sampling schema (Collins 2010) was employed to identify 12 potential participants. This purposeful sampling schema was used to reflect diversity in academic rank, gender, discipline, and NSF granting directorate in order to yield unique perspectives on data management planning (Maxwell 1997; Miles and Huberman 1994; Sandelowski 1995). Using census sampling, all researchers in this population were contacted for permission to access their DMPs.

Qualitative Research Design

Qualitative data was collected from multiple sources to provide richness and depth to the experience of data management planning. One of the most common approaches to collecting qualitative data is the interview because it allows a researcher to understand the meanings of activities for individuals, explore or construct complex phenomena, and yield a quantity of data (Marshall and Rossman 2011; Tracy 2013). Primary qualitative data included: (1) semi-structured, in-person interviews, including a pre-interview web-based questionnaire; and (2) document analysis of DMPs from funded grant applications.

To triangulate the information from the interviews and document analyses, two other types of qualitative data were collected and analyzed: (3) unstructured interviews with data management stakeholders (systems administrator, statistical analyst, archivist, and sponsored programs officer); and (4) NSF procedural documents, including Frequently Asked Questions and DMP guidance documents from each of the NSF Directorates. Qualitative data collection took place between March and June 2015; an overview of the data collected can be found in Table 2.

Table 2: Summary of qualitative data gathered.

Type of Data	Hours Spent Collecting Data		Single-Spaced Typed Pages
	<i>Collect</i>	<i>Transcribe</i>	
Semi-structured interviews with faculty (N=6)	6.0 hours	28 hours	198
Unstructured interviews with stakeholders (N=4)	3.5 hours	19 hours	78
Data management plans (N=35)	---		61
NSF procedural documents (N=7)	---		39
Subtotal	9.5 hours	47 hours	
TOTAL	56.5 hours		376 pages

Interview Protocol

Interview questions were designed using the Data Lifecycle Model as conceptual model (DDI Alliance Structural Reform Group 2004), focused broadly on data management activities (e.g. creation and use of metadata; short-term storage of data; long-term data storage and preservation; data sharing practices), and were informed by prior research (in particular Peters and Dryden 2011; Jones, Ross, and Ruusalepp 2009; Witt and Carlson 2007). An interview protocol was submitted to the UVM Research Protections Office and received a Protocol Exemption Certification.

Interview Procedures

Interviews occurred in two stages in the spring of 2015. Once participants agreed to be interviewed, they were sent a short pre-interview questionnaire, administered in UVM's LimeSurvey survey software. These questionnaires collected basic data about participants' NSF-funded research: data types collected, data formats collected, metadata standards used,

size of data, data storage, lifespan of data, and data sharing practices; they also primed the participant to begin to explicitly reflect on their data management practices (see Appendix A). The questionnaire took approximately 10 minutes and was completed prior to the interview.

All in-person interviews occurred in the participant’s office or research space. Participation was voluntary, and participants understood that they could refuse to answer any question or terminate the interview at any time. Prior to commencing the interviews, participants were given statements of consent for the study, and permission was obtained for audio recording the interviews using QuickTime Player.

Participants were prompted to talk specifically about their NSF-funded research that required a DMP: an overview of the research; different roles of collaborators in the research process; sharing data; long-term preservation of data; challenges in managing data; and UVM support for data management. Scripted questions were often followed by additional unscripted questions to elicit more detailed information. Interview protocol can be found in Appendix B. The interviews ranged in length from 42 to 71 minutes, and transcripts were produced using HyperTRANSCRIBE 1.6.

Documents

Data management plans were obtained from the UVM Office of Sponsored Programs Administration with the permission of the grantee. The UVM Research Protection Office provided a certification of Not Human Subjects Determination for this research.

Findings

Qualitative Participants

From a population of 47 researchers who had received NSF grants between 2011 and 2014 and who submitted a DMP, 12 researchers were identified through purposeful sampling for interview. Due to timing and scheduling constraints, six researchers were interviewed, resulting in a 50% response rate. From this same population, 35 faculty members (74.5%) granted

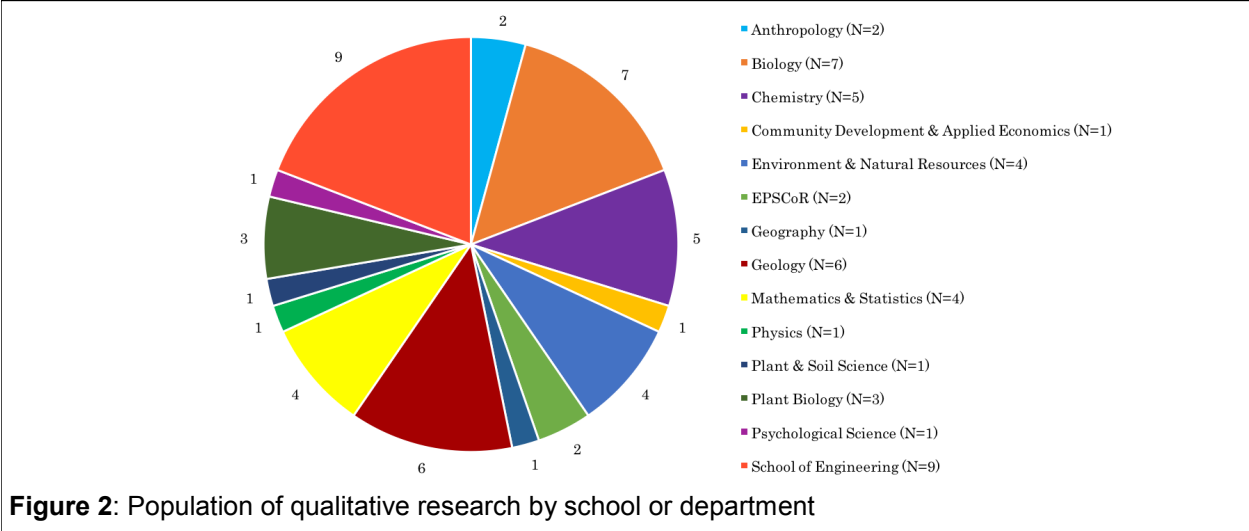


Figure 2: Population of qualitative research by school or department

Table 3: Descriptive statistics of qualitative population, document analysis sample and interview sample.

	Population (N=47)	DMP Sample (N=35)	Interview Sample (N=6)
Department			
Anthropology	2 4.3%	2 5.7%	0 0.0%
Biology	7 14.9%	3 8.6%	1 16.7%
Chemistry	5 10.6%	4 11.4%	1 16.7%
Community Development & Applied Economics (CDAE)	1 2.1%	0 0.0%	1 16.7%
Environment and Natural Resources	4 8.5%	2 5.7%	1 16.7%
EPSCoR	2 4.3%	2 5.7%	0 0.0%
Geography	1 2.1%	1 2.9%	0 0.0%
Geology	6 12.8%	3 8.6%	0 0.0%
Mathematics & Statistics	4 8.5%	3 8.6%	1 16.7%
Physics	1 2.1%	1 2.9%	0 0.0%
Plant & Soil Science	1 2.1%	1 2.9%	0 0.0%
Plant Biology	3 6.4%	3 8.6%	0 0.0%
Psychological Science	1 2.1%	1 2.9%	0 0.0%
School of Engineering	9 19.1%	9 25.7%	1 16.7%
Gender			
Female	6 12.8%	6 17.1%	2 33.3%
Male	41 87.2%	29 82.9%	4 66.7%
NSF Directorate			
Biological Sciences (BIO)	13 27.7%	9 25.7%	1 16.7%
Education & Human Resources (EHR)	1 2.1%	1 2.9%	0 0.0%
Engineering (ENG)	6 12.8%	5 14.3%	1 16.7%
Geosciences (GEO)	7 14.9%	3 8.6%	0 0.0%
Mathematical & Physical Science (MPS)	12 25.5%	9 25.7%	2 33.3%
Office of Integrative Activities (OIA)	2 4.3%	2 5.7%	0 0.0%
Social, Behavioral & Economic Sciences (SBE)	6 12.8%	6 17.1%	2 33.3%

permission to access their DMPs. Overview of the population by disciplines can be found in Figure 2; descriptive statistics of the phase one participants can be found in Table 3.

Qualitative Data Analysis

The qualitative data was analyzed through a multi-step process: (1) organization of data; (2) data immersion; (3) construction of categories and themes; (4) data coding; (5) creation of analytic memos; and (6) interpretation of the findings (Creswell 2014; Marshall and Rossman 2011; Tracy 2013). After the data was organized, read, and re-read, the interview transcripts and DMPs were uploaded into HyperRESEARCH 3.5 qualitative data analysis software for coding. Descriptive primary cycle codes were assigned to the raw data (Tracy 2013). Using a constant comparative method (Charmaz 2006; Glaser and Strauss 1967), new data assigned to each code was constantly compared to previous data; as necessary, definitions for codes were updated or data was broken into new codes. In the secondary-cycle coding, the codebook was examined in order to organize, synthesize, and categorize the codes into interpretive concepts or themes (Tracy 2013). Thematic memos were then written that focused on the meaning of codes and on the relationships among and between the codes within a theme. Finally, the data from the interviews and the document analysis were triangulated with data from the NSF procedural documents and the stakeholder interviews to corroborate, elaborate, and illuminate the research (Marshall and Rossman 2011).

Qualitative Results

The initial coding phase was completed through the process of structural coding, which is designed to start organizing the raw data around the research questions (Saldaña 2009). While DMPs may differ in form and content based on the NSF funding agency, they typically address the following elements: data and metadata description; security, ethics, and intellectual property issues; data access, sharing, and re-use provisions; short-term (five years or less) storage and management; long-term (more than five years) storage, management, and preservation; and personnel and infrastructure resources (Krier and Strasser 2014). Initial data labels were derived from these elements and the constructs of the Data Lifecycle Model (DDI Alliance Structural Reform Group 2004); in vivo codes were also identified and added to the codebook. These primary-cycle codes were descriptive in nature, and included 12 categories: data types; data size; data description; data storage; data back-up; data retention; data sharing; data preservation; funding; infrastructure; support; and guidance (for a sample coding structure, see Figure 3).

During the secondary-cycle coding process, pattern coding was utilized to examine initial codes, identify trends, patterns, and relationships, and assign codes into categories or themes (Hatch 2002). A concept map was created to visually cluster codes in order to understand how the data fit together (see Figure 4 for sample concept map). Triangulation of qualitative data sources was used as a means of confirmation of measures and validation of findings (Casey and Murphy 2009; Creswell 2014; Leedy and Ormrod 2004). As a result, three major themes were identified: data management activities, institutional research support, and perceptions of DMPs.

Data Management Activities

PRIMARY CODES		SECONDARY CODES		THEMES
XML	Data Types			Data Management Activities (RQ1)
Excel				
Transcripts				
SPSS				
Specimen				
EML	Metadata		Data Description	
Darwin Core				
FGDC				
ReadMe				
Hard Drive	Data Storage Short-Term	Data Storage Long-Term	Infrastructure	
Campus Server				
DropBox				
LTER	Data Sharing			
GenBank				
DataOne				
Publications				
Presentations				
Website				
Data Back-Up				
Data Security				

Figure 3: Sample qualitative coding structure for data management activities (RQ1)

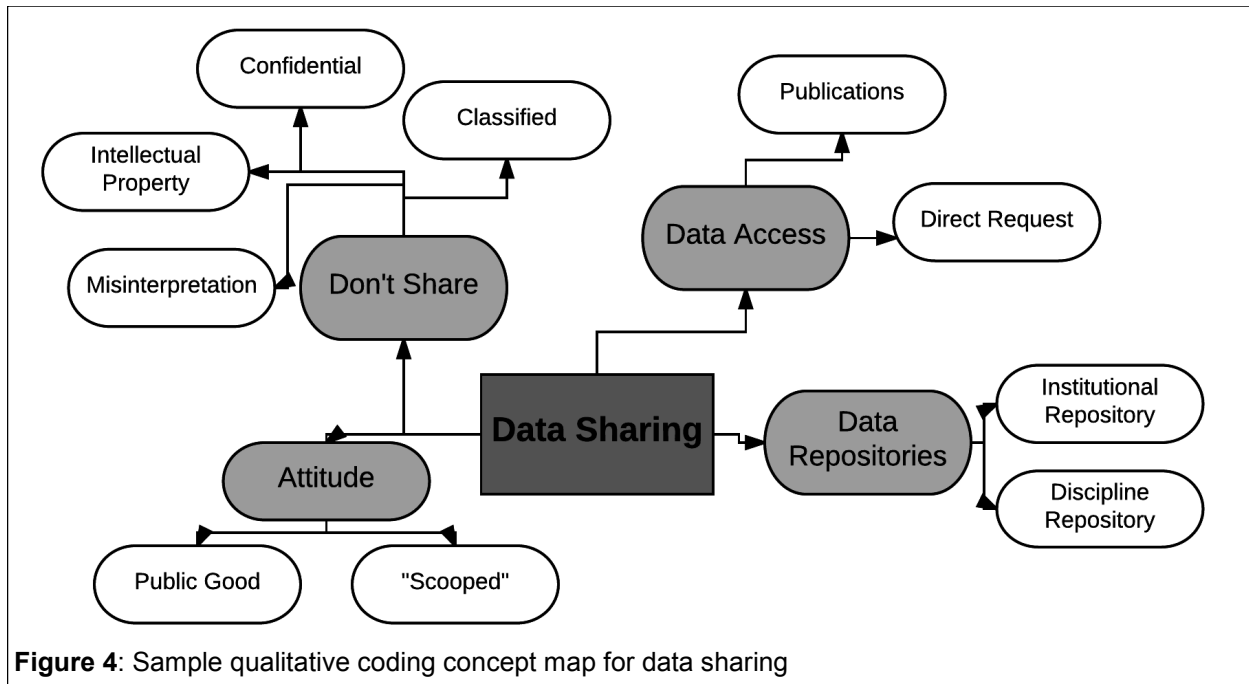


Figure 4: Sample qualitative coding concept map for data sharing

Despite specific requirements of funding agencies, the information regarding data management in the DMPs (N=35) was variable in nature. Eighteen plans (51.4%) addressed “data description” generally, while nine (25.7%) mentioned specific metadata schema. Twenty-nine (82.9%) of the DMPs addressed storage, while 14 (40.0%) addressed back-up protocols. Information about data management activities was also collected during the interview process. Three of the six participants discussed metadata: one used a specific metadata scheme, while the other two created ReadMe text files to document data within their research group. All six participants used campus servers to store their data; five (83.3%) also mentioned backing up data on external hard drives. During three of the interviews, external hard drives were physically pointed out in the offices and research spaces; the researcher from Mathematics & Statistics noted, “I’m pretty comfortable with a big box of hard drives.”

As to data sharing and preservation specifically, 28 of the 35 (80.0%) DMPs indicated a general willingness to share data. Of these 28, seven (20.0%) addressed specific data sharing restrictions, four (11.4%) addressed sharing data through open source or creative commons licenses, and eight (22.9%) addressed sharing via a direct request. All 35 DMPs addressed data access through a variety of means, including publications and presentations (94.3%), data repositories (48.6%), and websites (42.9%). These preservation pathways were mirrored in the interviews, where all six interviewees indicated they would be sharing research data via publications, four mentioned sharing via websites and collaborative online spaces, and two mentioned specific disciplinary data repositories.

In the pre-interview questionnaires, participants highlighted a variety of reasons why they restricted sharing their data, including: intellectual property concerns; confidential, proprietary, or classified information; and license and usage restrictions. Attitudes elicited during interviews demonstrated varied opinions on data restriction. Interviewees from Chemistry and Mathematics & Statistics were worried about being “scooped,” while the interview participant

from Environment & Natural Resources said, “We have this general feeling that it helps the public good, and that someone has paid for it. As long as whoever has funded our work is okay with us [sharing the data], we’ll put [our data out there]... If it’s going to be for the greater good, you want it to be as simple as possible [to access].”

Another theme related to sharing research data surfaced during the interviews: potential misinterpretation of data. The interview participant from Biology was anxious that their research would be under unfair scrutiny by individuals who don’t understand the methods they used to analyze the data. The researcher in Chemistry offered a divergent attitude: “I think people are nervous about having data reinterpreted and finding out they’re wrong. But that’s science! I’m okay with that, as long as you did a reasonable, honest effort.”

Institutional Research Support

In the pre-interview questionnaire, interview participants were asked a multiple-choice question about challenges they faced in preparing their DMP. Three of the participants indicated they faced no challenges, while the other three indicated that a lack of appropriate infrastructure to store their data in the short-term was a significant challenge. This latter group also highlighted other issues, including a lack of guidance from UVM, a lack of guidance from the funding agency, and a lack of appropriate infrastructure to archive and make data accessible in the long-term.

During the interviews, participants were prompted to discuss challenges they’ve faced managing data generally, whether they indicated they faced challenges in preparing their DMPs or not. Five of the researchers expressed frustration in terms of base-level research support they were receiving from the University, in particular, highlighting how “woefully understaffed” (Engineering researcher) campus IT is. The interviewee from Community Development and Applied Economics (CDAE) opined, “What do you really get in terms of research support? One of the things I always wonder when I get these big grants and I see the overhead taken off is, ‘What does my overhead fee go towards, exactly?’”

In talking about whether UVM should create its own institutional data repository to preserve and provide access to data, the participant from Biology said: “I would rather UVM do that, to be honest. I would rather not maintain [data] myself because I am nervous that there will be some problem or error and lots of data would go away and that would be bad.” When asked whether the researcher from Mathematics & Statistics had ever considered putting data into a repository, they said:

That’s something that I think is a little more interesting than depositing publications [into an institutional repository]. But is it expensive to do this? Because I’m riding high on these grants now, but ten years from now? Is there a permanent fee? If it’s free, of course that would be great. But I think that the issue is that we don’t have these long-term storage facilities now and so people have come to not expect them. And so why would they spend a lot and have a commitment to keep paying? They’d rather just let the data expire somehow.

The participant from Environment & Natural Resources was more reserved about the need for a data repository, emphasizing that they use different systems to store and share their research data based on the data type collected and the audience.

I think the challenge with all these repositories now is that all of them have strengths and weaknesses. FigShare works really well when I have nice graphics I want to put up and get a DOI, or if I have tabular data. But what happens when they're sold to Elsevier or another publisher? It's also not the best for geospatial data and discovery, and that's why we have projects using Google Maps... I think the challenge of all these data repositories is your chances of making everyone happy are slim. Slim to none.

Perceptions of Data Management Plans

During analysis and coding, a theme emerged related to writing a DMP and the perception of them, especially during the grant review process. When talking about their process in creating a DMP, the researcher in Chemistry said: "I took the NSF guidelines and read what they want in the plan. And I think part of it was the way they just want you to structure it the way they want. They talked about three or four big questions that the DMP should address. So I wanted to make sure to address the issues because that's making the program officer happy." To the same question, the researcher in Mathematics & Statistics said, "NSF does not provide a lot of guidance. And I've been on a review panel, and everybody is like, 'Well, we have to put in a data management plan.' It's just one of those situations where we don't have a lot of guidance as to what to put in, so we're just kind of throwing things against the wall. And everybody was complaining about it."

Similarly, during the interview with the Engineering researcher, they said: "I've been on review panels, and nobody says anything about the data management plan. Everybody reads it to check that they're there, but nobody makes any comment." They continue, saying the data management plan is merely a checkbox because,

I don't think we know what to say. We don't know what the standard is! NSF says the standard will come through the community of practice, so it means there is going to be an established norm. And then after a while, they are going to look at it and say, 'This is the norm.' And in three years, there will be review panels that go, 'Why is this person doing this and not this that I've seen in all the other data management plans in the last three years?' And then the investigator gets harassed, and then they'll conform. We'll all conform.

This lack of guidance and confusion is seen in the variability of the analyzed DMPs: while DMPs typically require a two-page submission, or the rough equivalent of 500 typed words double-spaced, the 35 plans analyzed ranged in length from 66 words to 1,317 words (mean=560, standard deviation=344).

Discussion

Data Dissemination and Data Sharing

As part of the grant submission process, NSF Directorates require specific elements to be addressed in DMPs to promote data sharing and long-term preservation of data. As is demonstrated by the document analysis, researchers do not account for all required elements, with nine of the 35 DMPs (36.0%) failing to address at least one of the major components of metadata, long-term data preservation, or data sharing.

The NSF requirement focuses on dissemination and sharing of research data: “Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data... created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing” (National Science Foundation 2014, pt. VI.D.4). Analyzed and summarized data – such that is included as tables and figures in presentations and publications – are a representation of the data, and not the primary data itself. The document analysis showed that 33 out of 35 researchers (94.3%) planned to share their research data through publications and/or presentations, but if we remove those responses, 13 (37.1%) of the DMPs do not address sharing primary data at all. It becomes apparent that there is a lack of understanding about the intention behind data sharing.

The interviews provided more context to this discussion, in particular regarding restrictions placed on disseminating primary data. The interviewee from Chemistry highlighted intellectual property concerns and competition, saying, “Chemists are the worst for sharing data, as a profession. I don’t know people that share their data!,” while the participant from Biology told an apocryphal story about a colleague who released a partial dataset, which was then harvested, analyzed, and published by a “predatory” researcher. This concern is reflected in Diekema, Wesolek, and Walters (2014), where the most common reason for researchers not sharing data was because it hadn’t been published yet.

The other issue that surfaced was a desire to restrict data to individuals who either understood how to analyze it or who were willing to be trained. The fear of invalidating results due to misunderstanding or misinterpretation is not uncommon: in an international survey on data sharing practices and perceptions, 75% of the survey respondents replied that “data may be misinterpreted due to complexity of the data across their research field,” 71% “agree that data may be misinterpreted due to poor quality of data across their research field,” and 74% “believe that data may be used in other ways than intended across their research field” (Tenopir et al. 2011, 7-8).

In addition to explicit restrictions, external and internal factors also impacted the researchers’ ability to share data. In the DMPs, the most common external factor was that data contained personal or sensitive information, while the most common internal factor related to broad issues of data organization (e.g. time spent to classify and organize data for public dissemination). In talking about data sharing requests, the interviewee from CDAE joked that they dreaded receiving an email requesting data because it was cumbersome to organize, difficult to send securely, and overall a time-consuming process. The participant from the Engineering department lamented how long it took for them to find the most current version of their data for their own use, and couldn’t imagine how they’d sort through all the files to share the data in a meaningful way with anyone else.

In addition to organization, factors of time, funding, and available personnel were challenges for researchers. Tenopir et al (2011) found that lack of time is one of the top reasons why researchers don't make their data available, and Bardyn, Resnick, and Camina (2012) similarly discovered that researchers found data management costly and labor-intensive: "Focus group participants agreed that term funding from research grants is not likely to be sufficient to provide for long-term stewardship of research data. One researcher stated that financial support requested for data management and curation in grant proposal budgets was always the first item cut by the funding agencies" (280-281).

Despite these internal and external barriers, the general attitude expressed in the interviews, and corroborated by the DMPs, favored disseminating research data; general support for data sharing is confirmed by the literature (Akers and Doty 2013; Diekema, Wesolek, and Walters 2014; Diekmann 2012; Scaramozzino, Ramirez, and McGaughey 2012; Tenopir et al. 2011; Walters 2009). Regardless of this support, data sharing remains problematic within the research community. The Mathematics & Statistics researcher said that reproducibility was a significant problem in science, and there was an urgent need for having open, shareable data. Tenopir et al (2011) reported that two-thirds of their survey respondents reported that lack of access to data generated by other researchers was a major impediment to scientific progress, and half of the respondents reported that a lack of access to data restricted their ability to answer scientific questions. Yet only 32.3% strongly agreed that they share their data with others and 11.6% strongly agreed that others can access their data. Diekema, Wesolek, and Walters (2014) similarly discovered that faculty use data stored in repositories more frequently than they deposit data in repositories. Overall, the theoretical need for making data accessible is understood, but what becomes clear is that there are structural and/or psychological barriers that need to be addressed in order to help researchers move from intention to practice.

Institutional Support & Infrastructure Barriers

One significant barrier emerged in the qualitative analysis related to research support – or lack thereof – from the University. Despite high levels of research activity on campus, there are no centralized research support services at UVM. Instead, research faculty need to navigate a myriad of disconnected services, including the Office of the Vice President of Research, the Office of Sponsored Project Administration, Enterprise Technology Services (ETS), the Vermont Advanced Computing Core, Statistical Consulting, and UVM Libraries. Because of the relative isolation of campus units who support different aspects of data during its lifecycle, misconceptions, misunderstandings, and information asymmetry about available infrastructure or services permeated the interviews. For example, interview participants expressed frustration about base-levels of support available through ETS (campus IT), particularly in terms of data storage. In both the interviews and the DMP document analysis, the most common method for storing data was through campus infrastructure in terms of local, departmental, collegial, or institutional servers. But what was available to researchers was unclear: During the interviews, one participant thought they had unlimited data storage, one participant thought they had up to 1 terabyte without cost, and a third said they had to start buying additional storage space once they reached 400 megabytes. In conversation with a systems administrator in ETS, it was confirmed that there is a base-level of support all researchers receive, but that it is not advertised. This is compounded by the fact that ETS wants to remain flexible in trying to meet researcher needs, and so the base-level of support can be variable. This confusion and lack of

awareness was also reflected in the absence of available UVM services in the analyzed DMPs.

When shifting towards issues of long-term data storage and preservation, the interviews provided a mix of views on the need for an institutional data repository, whose explicit mission would be to archive and provide access to curated primary research data. Several researchers were unsure whether the repository would meet their overall needs because they collect different types of data that have different requirements, such as geospatial data or video ethnographies. Three researchers were intrigued by the idea of not having to be responsible for the long-term management of their data, while one researcher did not distinguish the repository as anything more than a place for long-term storage. This attitude concerns the university archivist, who notes that the system may be viewed as “free parking” for data, a solution simply to get out of paying additional storage fees to ETS. These mixed opinions did not provide clarity on whether UVM should be investigating an institutional data repository; layered within that question is whether there is the necessary funding and personnel available at UVM to create such a system.

Beyond available UVM resources, there is a larger concern about the infrastructure available to preserve and share research data through existing data repositories. A number of discipline-specific data repositories, such as GenBank or Inter-University Consortium for Political and Social Research (ICPSR), fill this open-data need, but they have their share of weaknesses. There are more disciplines without repositories than there are disciplines with established repositories, a shortcoming that is also documented by Westra (2010). Those that do exist do not cross traditional disciplinary boundaries, which is limiting for interdisciplinary or transdisciplinary researchers. Interview participants also indicated that some of the repositories, like DataONE, are so complex that they make it difficult for even the most dedicated researcher to upload and share their data. Other repositories, especially those that have been created with grant funding, may not be sustainable in the long-term. What becomes clear is that infrastructure barriers, at both a local and disciplinary level, significantly limit the ability of researchers to preserve and share their data.

Perceptions of Data Management Plans & Attitudes Towards Data Management Planning

Several published studies addressed researcher attitudes towards data creation, data sharing, and data preservation (D'Ignazio and Qin 2008; Parham, Bodnar, and Fuchs 2012; Scaramozzino, Ramírez, and McGaughey 2012; Tenopir et al. 2011). Scaramozzino, Ramírez, and McGaughey (2012) used paired questions in their survey to evaluate inconsistencies between what researchers believe is important and what they are actually doing with their data. They found that while 84% of respondents believed it was important to have a data preservation plan in place, fewer than 15% actually had such a plan. In a different study, researchers found that data management practices had not been affected by the federal mandates (Diekema, Wesolek, and Walters 2014).

The perception of DMPs, and their role in the grant application process, was a topic that came up unprompted in three interviews. The DMPs were seen as something that needed to be submitted, but served merely as an item to check off a list and not something that was critically evaluated or had any bearing on whether or not the grant application was successful. In part, this checklist mentality can be seen as a result of limited support and guidance from the

granting agencies, as well as a lack of available funding and infrastructure to realize the DMPs. Diekmann (2012) quoted a researcher who said:

'If I were a reviewer of a proposal, for example, and somebody had a section in there that talks about how the data would be handled afterwards – how it would be stored and processed – I think that would be very impressive, because we never see that. It is just never included in any of the proposals. All we see is, data will be collected on this, and will be analyzed by this method, and that's it. Not any systematic process by which it will be collected and stored, housed, or managed after the facts' (p. 27).

DMPs are a federal-policy intervention predicated on several assumptions: that open data is valued as a public good and a scholarly necessity, that federally funded researchers aren't providing access to their research data, and that digital data are more prone to loss or corruption and there is an explicit need to preserve digital research data. If researchers have low esteem for DMPs, it is reasonable to question whether these plans are fulfilling their intended purpose of providing access to and preserving digital data in the long-term. The question that remains is whether there is a psychological barrier, in terms of attitudes and beliefs about DMPs, that negatively impact researchers preserving and providing access to their data.

Conclusion

This article focuses on the first qualitative phase of a two-phase exploratory sequential mixed methods research (MMR) design (Creswell 2014): an initial phase of qualitative data collection and analysis, followed by a phase of quantitative data collection and analysis (Figure 1). During phase one, qualitative research methods were used to understand UVM researchers' experiences with data management and to explore potential institutional barriers or challenges to managing data. This was accomplished through semi-structured interviews with research faculty and document analysis of NSF data management plans. While the qualitative results presented here are specific to a narrow population of researchers at UVM, when combined with other institutional research, studies can provide a more complex picture of research data management practices.

The data from the qualitative phase of research express multiple, and occasionally divergent, viewpoints. This qualitative study demonstrates that there is no singular reality for researchers or data management practices, and that there are multiple physical and psychological challenges they face in managing data. This suggests that there can be no singular and straightforward approach to supporting data management at the University of Vermont.

The sample for this qualitative study was necessarily small – six interview participants and 35 documents analyzed, all National Science Foundation grantees. This allowed in-depth exploration of the issues that are salient to UVM researchers who have submitted DMPs, but the findings are not generalizable to the campus and alone do not provide the necessary guidance for the development of data management support or services. While attempts were made at securing a diverse sample through critical-case sampling, not all disciplines and NSF Directorates were proportionally represented. Additionally, the document analysis only focused on data management plans from successful grants; there may be additional, or differing,

information available in DMPs from unsuccessful grants, and this warrants a follow-up study. Similarly, because the population criteria did not include National Institute of Health (NIH) grants, biomedical science researchers who have submitted data sharing plans were not included in this phase of the study, and therefore may have influenced the data analysis and findings.

Despite these limitations, the themes from this qualitative study provide a foundation on which to build a survey instrument in order to more broadly understand the research data management activities and challenges of faculty at UVM across all disciplines. A survey will also allow us to gain a broader understanding of the researchers' attitudes towards data management planning, namely mandates surrounding data sharing and preservation, and how these attitudes influence data management behaviors. Applying rigorous mixed methods research design, the results of this qualitative analysis will be linked in the interpretation-phase of the study with the results of the quantitative analysis in order to understand the broad spectrum of data management behaviors and challenge in order to develop appropriate research data services at the University of Vermont.

Supplemental Content

Appendix A and B

An online supplement to this article can be found at <http://dx.doi.org/10.7191/jeslib.2017.1097> 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 Cynthia Gerstl-Pepin, professor, Department of Education, for their advice and feedback at multiple stages of this study.

Disclosure

The author reports no conflict of interest.

References

- 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>
- Bardyn, Tania P., Taryn Resnick, and Susan K. Camina. 2012. "Translational Researchers' Perceptions of Data Management Practices and Data Curation Needs: Findings from a Focus Group in an Academic Health Sciences Library." *Journal of Web Librarianship* 6(4): 274-287. <http://www.dx.doi.org/10.1080/19322909.2012.730375>
- Berg, Bruce L., and Howard Lune. 2012. *Qualitative Research Methods for the Social Sciences*. 8th ed. Upper Saddle River, NJ: Pearson Education, Inc.
- Bracke, Marianne Stowell. 2011. "Emerging Data Curation Roles for Librarians: A Case Study of Agricultural Data." *Journal of Agricultural & Food Information* 12(February): 65-74. <http://www.dx.doi.org/10.1080/10496505.2011.539158>

- Casey, Dympna, and Kathy Murphy. 2009. "Issues in Using Methodological Triangulation in Research." *Nurse Researcher* 16(4): 40-55. <http://www.dx.doi.org/10.7748/nr2009.07.16.4.40.c7160>
- Charmaz, Kathy. 2006. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Thousand Oaks, CA: SAGE Publications.
- Coates, Heather, and Stacy Konkiel. 2013. "University-Wide Data Management Services: Cross-Campus Collaborations at Indiana University." *Indiana University Libraries*. <http://hdl.handle.net/2022/16595>
- Collins, Kathleen M. T. 2010. "Advanced Sampling Designs in Mixed Research: Current Practices and Emerging Trends in the Social and Behavioral Sciences." In *Sage Handbook of Mixed Methods in Social & Behavioral Research*, edited by A. Tashakkori and C. Teddlie, 2nd ed., 353-377. Thousand Oaks, CA: SAGE Publications.
- 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." <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://dx.doi.org/10.1016/j.acalib.2014.04.010>
- Diekmann, Florian. 2012. "Data Practices of Agricultural Scientists: Results from an Exploratory Study." *Journal of Agricultural & Food Information* 13: 14-34. <http://dx.doi.org/10.1080/10496505.2012.636005>
- DuraSpace Organization. 2016. "E-Science Institute Course Description." <http://www.duraspace.org/esi-course-description>
- Fearon, David Jr., Betsy Gunia, Sherry Lake, Barbara E. Pralle, and Andrew L. Sallans. 2013. *SPEC Kit 334: Research Data Management Services (July 2013)*. Washington, DC: Association of Research Libraries. <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334>
- Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory*. New York: Aldine de Gruyter.
- Greene, Jennifer C. 2007. *Mixed Methods in Social Inquiry*. San Francisco, CA: Jossey-Bass.
- Hatch, J. Amos. 2002. *Doing Qualitative Research in Education Settings*. New York: State University of New York Press.
- Jones, Sarah, Seamus Ross, and Raivo Ruusalepp. 2009. "Data Audit Framework Methodology." *HATII, University of Glasgow* 1.8. http://www.data-audit.eu/DAF_Methodology.pdf
- Krier, Laura, and Carly A. Strasser. 2014. *Data Management for Libraries: A LITA Guide*. Chicago, IL: ALA TechSource.
- Lage, Kathryn, Barbara Losoff, and Jack Maness. 2011. "Receptivity to Library Involvement in Scientific Data Curation: A Case Study at the University of Colorado Boulder." *Portal: Libraries and the Academy* 11(4): 915-937. <http://www.dx.doi.org/10.1353/pla.2011.0049>
- Leedy, Paul D., and Jeanne Ellis Ormrod. 2004. *Practical Research: Planning and Design*. 8th ed. Upper Saddle River, NJ: Prentice Hall.

- Marcus, Cecily, Stephanie Ball, Leslie Delserone, Amy Hribar, and Wayne Loftus. 2007. "Understanding Research Behaviors, Information Resources, and Service Needs of Scientists and Graduate Students: A Study by the University of Minnesota Libraries." *Minnesota Libraries*.
- Marshall, Catherine, and Gretchen B. Rossman. 2011. *Designing Qualitative Research*. 5th ed. Thousand Oaks, CA: SAGE Publications.
- Mattern, Eleanor, Wei Jeng, Daqing He, Liz Lyon, and Aaron Brenner. 2015. "Using Participatory Design and Visual Narrative Inquiry to Investigate Researchers' Data Challenges and Recommendations for Library Research Data Services." *Program: Electronic Library and Information Systems* 49(4): 408-423.
<http://www.dx.doi.org/10.1108/PROG-01-2015-0012>
- Maxwell, Joseph A. 1997. "Designing a Qualitative Study." In *Handbook of Applied Social Science Methods*, edited by L. Bick and D. J. Rog, 69-100. Thousand Oaks, CA: SAGE Publications.
- 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>
- Miles, Matthew B., and A. Michael Huberman. 1994. *Qualitative Data Analysis: An Expanded SourceBook*. 2nd ed. Thousand Oaks, CA: SAGE Publications.
- Mischo, William H., Mary C. Schlembach, and Megan N. O'Donnell. 2014. "An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals." *Journal of eScience Librarianship* 3(31): e1060. <http://www.dx.doi.org/10.7191/jeslib.2014.1060>
- National Institutes of Health. 2003. "Final NIH Statement on Sharing Research Data."
<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
- National Science Foundation. 2011. "Data Management for NSF Engineering Directorate Proposals and Awards."
https://nsf.gov/eng/general/ENG_DMP_Policy.pdf
- . 2014. "Proposal and Award Policies and Procedures Guide."
http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/nsf15_1.pdf
- 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
- Onwuegbuzie, Anthony J., Rebecca M. Bustamante, and Judith A. Nelson. 2010. "Mixed Research as a Tool for Developing Quantitative Instruments." *Journal of Mixed Methods Research* 4(1): 56-78.
<http://www.dx.doi.org/10.1177/1558689809355805>
- Onwuegbuzie, Anthony J., and Kathleen M. T. Collins. 2007. "A Typology of Mixed Methods Sample Designs in Social Science Research." *The Qualitative Report* 12(2): 281-316. <http://nsuworks.nova.edu/tqr/vol12/iss2/9>
- 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>
- Parham, Susan Wells, Jake Carlson, Patricia Hswe, Brian Westra, and Amanda Whitmire. 2016. "Using Data Management Plans to Explore Variability in Research Data Management Practices across Domains." *International Journal of Digital Curation* 11(1): 53-67. <http://www.dx.doi.org/10.2218/ijdc.v11i1.423>
- Parham, Susan Wells, and Chris Doty. 2012. "NSF DMP Content Analysis: What Are Researchers Saying?" *Bulletin of the American Society for Information Science and Technology* 39(1): 37-38.
<http://www.dx.doi.org/10.1002/bult.2012.1720390113>

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

Peters, Christie, and Anita Riley Dryden. 2011. "Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston." *Science & Technology Libraries* 30(4): 387-403.

<http://www.dx.doi.org/10.1080/0194262X.2011.626340>

Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. 2007. "Sharing Detailed Research Data Is Associated with Increased Citation Rate." *PLoS ONE* 2(3): e308. <http://www.dx.doi.org/10.1371/journal.pone.0000308>

Saldaña, Johnny. 2009. *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: SAGE Publications.

Sandelowski, Margarete. 1995. "Sample Size in Qualitative Research." *Research in Nursing and Health*

18(2): 179-183. <http://www.dx.doi.org/10.1002/nur.4770180211>

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>

SPARC. 2016a. "Open Access." <http://www.sparc.arl.org/issues/open-access>

———. 2016b. "Open Data." <http://www.sparc.arl.org/issues/open-data>

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://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 22.

<http://carnegieclassifications.iu.edu/methodology/basic.php>

Tracy, Sarah J. 2013. *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact*. London: John Wiley & Sons Inc.

University of Vermont. 2017a. "Research at the University of Vermont." Accessed February 22.

<https://www.uvm.edu/research>

———. 2017b. "UVM Facts." Accessed February 22. https://www.uvm.edu/uvm_facts

Walters, Tyler O. 2009. "Data Curation Program Development in U.S. Universities: The Georgia Institute of Technology Example." *International Journal of Digital Curation* 4(3): 83-92. <http://dx.doi.org/10.2218/ijdc.v4i3.116>

Westra, Brian. 2010. "Data Services for the Sciences: A Needs Assessment." *Ariadne* 64.

<http://www.ariadne.ac.uk/issue64/westra>

Williams, Sarah C. 2013. "Gathering Feedback from Early-Career Faculty : Speaking with and Surveying Agricultural Faculty Members about Research Data." *Journal of eScience Librarianship* 2(2): e1048.

<http://www.dx.doi.org/10.7191/jeslib.2013.1048>

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>