2016

A Study of Data Sharing Practices within Scholarly Research Communities

Mary Harward
Nova Southeastern University, mary@nova.edu

A Study of Data Sharing Practices within Scholarly Research Communities

by

Mary Harward

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Information Systems

College of Engineering and Computing
Nova Southeastern University

2016

We hereby certify that this dissertation, submitted by Mary Harward, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

_____ _____

Steven D. Zink, Ph.D — Date
Chairperson of Dissertation Committee

_____ _____

James Parrish, Ph.D. — Date
Dissertation Committee Member

_____ _____

Amon Seagull, Ph.D. — Date
Dissertation Committee Member

Approved:

_____ _____

Yong X. Tao, Ph.D., P.E., FASME — Date
Dean, College of Engineering and Computing

College of Engineering and Computing
Nova Southeastern University

2016

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

A Study of Data Sharing Practices within Scholarly Research

by
Mary Harward
August 2016

Recent literature acknowledges the importance of data and effective data management strategies to facilitate collaboration between disciplines of research. Likewise, understanding the policies and practices that support data sharing is a growing area of research in the fields of information and social studies of science. Shared data allow researchers to build on fellow researchers' work to enrich and facilitate advancements in science.

While much has been written to identify the elements that adversely affect data sharing in scholarly research, a definitive framework remains unclear. Several theories have been presented to explain this shortfall; however, the reasons are highly diverse. Some suggest the factors that impact data sharing practices include delays in the peer review process, ineffective data management practices, mistrust, financial considerations, and vague data sharing policies and procedures.

Those who support data sharing have acknowledged the important role of funding agencies to leverage the sharing of data in scholarly research in return for researcher support. Likewise, advocates suggest that scientific societies should establish data sharing as standard procedure. Respected organizations such as the National Academy of Sciences (NAS), National Science Foundation (NSF), and National Institutes of Health (NIH) are tasked with developing modern strategies to ensure that policies and procedures regarding data management and dissemination meet the evolving needs and computational capabilities of the 21st century.

While the NSF has proposed recent, updated regulations to guide the scientific community to adopt a culture that promotes the sharing of research data, literature suggests that regulations have been ineffective in advancing data sharing practices. The purpose of this research study was to review the NSF federal grant application process and its influence on timely data sharing practices. Most importantly, the goal of this study was to identify definitive ways in which the NSF grant application process may be improved to expedite the sharing of research data in the future.

ACKNOWLEDGEMENTS

I would like to express my most sincere appreciation and thanks to my trusted advisor, Dr. Steven Zink, who provided support and encouragement throughout this long journey. Thanks so much for guiding me along the way. Without your words of encouragement, seeing this study to completion would have been an impossible task. Looking back, I will always remember your thought-provoking courses and how much I enjoyed attending each one. Thank you for the opportunity to work with you and for challenging me to strive for excellence.

Special words of thanks and gratitude to Dr. James Parrish and Dr. Amon Seagull for your time and willingness to participate as members of my committee. Your suggestions were welcome enhancements to my dissertation and allowed me the opportunity to improve and perfect my work. Many thanks also to Dr. Eric Ackerman and Dr. Marilyn Littman, as I greatly appreciate your past support and guidance. Another important note of thanks is extended to Dr. Yair Levy, my first professor in the doctoral program. Thank you for being such a memorable and inspiring teacher.

To all the respondents of the survey and especially each interview participant, you have my warmest appreciation and thanks. This was a rewarding experience for me and I was humbled to meet so many gifted researchers who were willing to share their valuable time, wisdom, and insight, without which this research study would not be possible.

Finally, thank you to my family, friends, and coworkers for your faithful understanding and support, who, through the years, never seemed to grow tired of hearing me say that this journey would only take a few more months to complete.

This work is dedicated with love to my greatest treasure – my family:
Alec, Tamara, Hailey, Alessandra, and Isabelle.
Thank you for believing in me.

Table of Contents

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Leadership in science depends on reliable and effective access to scientific data (Borgman, 2012; Fecher, Frieske, & Hebing, 2015; Tenopir et al., 2011; Wilbanks & Friend, 2016). A review of literature supports the importance of data and its successful management to facilitate collaboration between disciplines of research. According to Tenopir et al., "data are the infrastructure of science" (p. 1). In keeping with recent discussions on the use of open access environments that foster knowledge and data sharing, well-respected organizations such as the National Science Foundation (NSF), National Academy of Sciences (NAS), and National Institutes of Health (NIH) are developing long-term strategies to ensure that policies and procedures regarding data management and dissemination meet the evolving needs and computational capabilities of the 21st century.

As advancements in digital technologies increase, new opportunities exist to ensure data integrity through increased openness and transparency (Holden, 2013; NSF, 2016a). The emergence and growth of openly accessible databases such as Genbank and Sloan Digital Sky Survey, for example, illustrate exciting data sharing opportunities among diverse disciplines of science (MacMillan, 2014). Recent technological changes, especially those related to networked computing, have made global access to research data sets a reality (Pampel et al., 2013; Tenopir et al., 2011).

A notable advancement in March, 2016 by Sage Bionetworks, for example, facilitates timely sharing of clinical data captured via smartphone interface within the scope of the mPower study (Wilbanks & Friend, 2016). mPower is a health research

study for Parkinson's disease conducted with 9000+ study participants. The mPower research study data, including quantitative sensor data and self-reported outcomes, is readily available for immediate sharing in an effort to improve patient health for those afflicted with the disease.

To promote transparency and spur economic growth, President Obama in May 2013 signed an Executive Order and introduced an Open Data Policy to allow easier access to information generated and stored by the U.S. Federal Government. The Executive Order requires that government generated data be made available in accessible, readable formats, while safeguarding citizen privacy, security, and confidentiality (Executive Order No. 13,642, 2013).

In keeping with the Open Data policy, likewise NSF published on May 9, 2013 an Open Data Memorandum that acknowledges three goals to guide the use, reuse, and accessibility of agency data. NSF's open data goals were designed to expand and publish data assets; enrich by improving the management and usability of agency data; and foster openness by ensuring accessibility of data in a readable format. NSF defines open data as "publicly available data structured in a way to be fully accessible and useable" (NSF, 2014, p.1).

The increasing importance of data sharing in scientific research is clear. Tenopir et al. (2011) explained that data sharing "includes the deposition and preservation of data; however, it is primarily associated with providing access for use and reuse of data" (p. 1). Data sharing practices should be present in each phase of the data and research lifecycles, which includes collecting and generating data, managing and analyzing data, and sharing the data (Tenopir et al.)

Even so, facilitation of effective data sharing practices is not understood fully (Borgman, 2012; Fecher et al., 2015; MacMillan, 2014 Tenopir et al., 2011). Although few empirical studies exist, several theories have been introduced to explain this shortfall, and the reasons are as diverse as the disciplines of science themselves. Some suggest the progress of scientific research has been adversely affected due to delays in scholarly publishing's peer review process (Harris, 2009). The peer review process is a cornerstone of academic writing to ensure that the information in academic publications is verifiable and of good quality (Harris). The basis of the peer review process is that research papers are subject to review by experts in the field to ensure accuracy and quality.

Other related factors that may adversely affect the progress of scientific research include deficiencies in documentation and the absence of metadata that support data sharing (Tenopir et al., 2011) or an unwillingness to share data based on mistrust (MacMillan, 2014); competitiveness (Committee on Science, Engineering, and Public Policy [COSEPUP], 2009); or relationship-related factors (MacMillan). Nelson (2009) suggested that data management practices affect the data sharing process. Further, Pittman (2010) argued that current data sharing policies themselves have contributed to the problem, citing vagueness in existing guidelines and federal regulations. Moreover, Tenopir et al. suggested that time constraints and lack of funding are key factors that explain why data is not made available for sharing.

Longo and Drazen published an editorial in *The New England Journal of Medicine* in 2016 that discussed data sharing and the authors' belief that "data sharing should happen symbiotically, not parasitically" (p. 276). The authors presented two main

concerns when researchers opt to share data between independent clinical studies. Misunderstanding and inaccurate findings can occur when a researcher who did not participate in generating or collecting the data does not have a sound understanding of study eligibility criterion, study population, and data collection. Further, the authors suggested that some researchers fear that "research parasites" may steal from research productivity or use data to disprove the original researcher findings (Longo & Drazen).

Problem Statement

Understanding the practices and policies that support data sharing is a growing area of research in the fields of information studies and social studies of science. Related research has been limited to case studies conducted within individual disciplines of science (Borgman, 2012). Research increasingly argues for additional empirical studies to explore data sharing efforts from diverse aspects, including studies to identify specific elements that affect data sharing practices, comparative studies to understand sharing between diverse disciplines, and exploratory studies to gain insight from those who successfully share and reuse data (Haeussler, 2011; Savage & Vickers, 2009; Tenopir et al., 2011).

Shared data allow researchers to build on fellow researchers' work to achieve timely results (Pampel et al., 2013, Tenopir et al., 2011). As research in science, engineering, and education becomes increasingly data intensive, digital data that scientists and engineers produce and store are increasing in volume, driven as an outcome of study simulation, observation, and experimentation. Likewise, the development of new scientific methods that adapt to evolving data sharing needs is essential (Kitchin, 2015).

In keeping with the need for advancements, and given its growing importance, recent literature calls for empirical research to determine methodologies to improve data sharing strategies within the realm of scholarly research (Haeussler, 2011; Pampel et al., 2013). Haeussler argued the practice of sharing data in scholarly research continues to be affected by diverse factors.

Research Goals

The main goal of this case study was to explore the NSF grant application process in federally funded research to determine the degree and nature of its influence on the timely sharing of scholarly data. In a sense, this study highlighted the various efforts of numerous groups to move toward the open data movement. While federal funding agencies, such as the NSF, envision future knowledge communities within open networked environments where data sharing is considered the norm, few studies exist to identify factors that influence data sharing (Haeussler, 2011; Pittman, 2010).

The growing importance of data policies that guide researchers and the methodologies set forth by federal funding agencies play a significant role in modern research (Pampel et al., 2013). On January 18, 2011, NSF revised its grant application policy to mandate that supplementary documentation be included with grant applications in an attachment known as the Data Management Plan (NSF, 2011). Recent modifications to NSF policy require grant applicants to share data and supporting materials gathered within the scope of NSF funded research. This study explored the impact of NSF policy changes within the grant application process.

Research Questions

The need for advanced research to identify new strategies for data sharing policies and practices has been documented in the recent work of Fecher et al. (2015), Pampel et al. (2013), and Tenopir et al. (2011). Likewise, there is growing momentum to expand traditional knowledge boundaries by fostering expanded research partnerships and funding collaboration between universities, government agencies, and other industry (Tenopir et al.). The objective of this study was to address specific research questions (RQs):

RQ1: Within the past two years, what specific relationships, if any, can be drawn between the Principal Investigator (PI) and the sharing of research data and materials, in keeping with current NSF federal grant application guidelines?

RQ2: In what ways can the NSF federal grant application process be improved to facilitate timely sharing of research data?

Relevance and Significance

Literature acknowledges that scientific inquiry needs to be an open process (Holden, 2013; MacMillan, 2014). Therefore, the ability to access research data is critical to the progress of analysis and inquiry, particularly in light of today's ubiquitous networking to facilitate open access. The scientific community must adopt effective strategies to facilitate data sharing and efficiency, enabling greater capabilities, and empowering enhanced collaboration over distance and across varied disciplines of science.

Advocates of data sharing have acknowledged the important role of funding agencies to promote data sharing in scholarly research in return for researcher support

(Fecher et al., 2015, Kitchin, 2015; Tenopir et al., 2011). Recent literature calls for new data practice policy and practices to ensure the effective use of public resources and funds (Tenopir et al.). Likewise, advocates suggest that scientific societies should establish data sharing as an accepted norm, while scholarly journals should make data sharing assurance a prerequisite of publication.

Federal funding agencies, such as the NSF and NIH, have proposed updated regulations in an attempt to guide the scientific community towards adopting a culture that promotes the sharing of research data (Holden, 2013). To date, the literature suggests that regulations have been ineffective in advancing data sharing practices (Tenopir et al.). The NSF (2007), itself, has documented its concern that data set collections have been crafted in a piecemeal fashion and have not been considered collectively for the sake of future accessibility.

While much has been written in related literature to help understand the elements that adversely affect data sharing in scholarly research, a definitive framework remains unclear, substantiating the need for future empirical studies (Pampel et al., 2013; Tenopir et al., 2011). According to the National Mental Health Council:

> Incentives for data sharing need to be offered that offset the investigators' loss of control over their databases… Ultimately there has to be a procedural framework that makes sharing sensible, efficient, and value-added. If all those pieces are in place, fewer external or coercive forces are needed to convince researchers to share (Arzberger et al., 2004, p. 2).

Much in recent literature argues for a full understanding of the factors that affect data sharing, which is critical to ensure the progress of future scientific research.

Barriers and Issues

One significant barrier was the the definition of data that varies greatly between collaborators and likewise between disciplines. Borgman (2010) cited a definition from the National Research Council, "*Data* are facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors" (p. 15). Paradoxes may be found in trying to define data's value, as some types of data may have both immediate and enduring value. Further, data may gain value over time or possess changing values, while other types of data may be easier to recreate than create (Borgman). Even so, the NSF (2007) suggested that despite data's importance, "there exists no standard or widely accepted definition of exactly what research data are" (p. 22).

In order to mitigate this barrier, it was important to understand that clear distinctions exist between data and depends upon the data themselves, whether categorized as observational, computational, or experimental (COSEPUP, 2009). Observational data include data retrieved from instruments, such as data found in surveys or weather mapping. Computational data result from the execution of a simulation or computer model. Experimental data are products of laboratory studies, such as controlled behavioral studies (Borgman, 2010).

Another constraint was the ability of this researcher to identify people who were knowledgeable of the grant application process within several academic institutions. More specifically, this researcher identified and contacted PIs, co-investigators (Co-PIs), and grant administrators to request their participation in the study. To engage potential candidates to participate in this study, the researcher reiterated the value of this research to improve future data sharing strategies. Study participants were limited to a select

population of NSF grant applicants from universities in the state of Florida. The survey instrument and interviews were designed to reveal participant perceptions and actions in their institution and that the population selected was generally representative of other research applicants located elsewhere in the United States.

Definition of Terms

Data sharing is the practice of making data available to other investigators within scholarly research. Data stewardship may be defined as the preservation of data to ensure enduring value (COSEPUP, 2009). Stewardship takes data accessibility to a new level, intending to preserve data and metadata for future use by researchers from disparate and similar disciplines alike. Stewardship implies the active preservation of data over long durations of time. According to COSEPUP, stewardship involves a conception of research where data are "both an end product of research and a vital component of the research infrastructure" (p. 27).

Towards a consistency of understanding, metadata may be defined as a subset of data, or rather, information about data. Metadata are used often to summarize data content, structure, context, and inter-relationships. Metadata add purpose to data and enables the identification of similar data in unlike data sets.

Approach

This study examined the federal grant application and related grant application policies and procedures of research grantees and the NSF. Using a multi-method approach, the study used two qualitative methods: a qualitative survey instrument and semi-structured interviews. Morse (2003) supported multi-method research to achieve a more robust understanding of a study subject and to facilitate research goals. Likewise,

Tashakkori and Teddlie (2003) argued that qualitative research has emerged in popularity over the past two decades, which they suggested was due to researchers' dissatisfaction with traditional methodology practices of quantitative research.

The researcher performed an in-depth review of the federal grant application and detailed instructions the NSF provided to grantees. The researcher also reviewed the roles of people responsible for completing and submitting the application to NSF for approval (e.g. PIs and Co-PIs). The researcher needed to identify, select, and contact specific schools, PIs, Co-PIs, and NSF employees to request their participation in the study survey and interviews.

As an initial step, a survey was refined and administered, via Survey Monkey.com, to select grantees and NSF administrators. The researcher developed an interview guide, which was used as a framework for administering in-depth interviews to gain an understanding of the steps required to complete a grant proposal. The interview data was gathered and analyzed to address the research questions. Findings from the qualitative survey and interviews were summarized and triangulated to identify potential patterns towards theory development.

Resources

Resources for this study included universities and researchers who submitted proposals to the NSF for grants. Another required resource was Survey Monkey.com, a cloud-based survey service, which provided a secure platform by which to administer the survey.

Chapter 2

Review of the Literature

Introduction

An extensive review of the literature revealed a history of data sharing practices within the natural sciences. As the National Research Council reported in 1992, the general norms of science emphasize openness as a standard principle, with the expectation that scientists should exchange research data and materials to achieve replication of study findings (Fienberg, 1994). While this norm has been acknowledged for centuries, literature suggests that the actual practice of sharing research data has varied greatly, even within individual disciplines (Fienberg).

Fienberg published an article in 1994 that provided in-depth insight into the ethical, legal, and professional dimensions of statistical data sharing in the 1980s. Describing ethical issues in the health sciences, Fienberg explained how data sharing was brought to the forefront as a result of high profile cases of fraud and scientific misconduct during that era. These indiscretions captured the attention of the national media, university officials, and ultimately the U.S. Congress.

Such cases revealed the inadequacies of policies and regulations related to scientific misconduct. As a result, federal funding agencies, such as the NSF, began to establish mechanisms to review inappropriate conduct within the sciences. Issues related to data sharing and the falsification and misrepresentation of data became the topic of discussions at conferences and debates during that era (Fienberg, Martin, & Straf, 1985; 1994).

Fienberg et al. published one of the first comprehensive studies devoted to the subject of sharing research data in 1985. Describing data as "the building blocks of empirical research" (p. 3), the authors provided insight into the controversies, benefits, and consequences of sharing data. Using the density of the earth as a classic example of open research (as cited in Cavendish, 1978), Fienberg et al. argued that withholding data from fellow scientists adversely affects scientific progress and understanding.

Authored as a collaborative effort by a group of social scientists, Fienberg et al.'s (1985) publication was a comprehensive multidisciplinary work that provided insight into the challenges of data sharing from diverse perspectives: cultural, legal, technical, financial, and social. Fienberg et al. explained that from an historical perspective, data sharing has been more prevalent within the realm of natural, rather than the social sciences, which introduced significant complexities.

Acknowledging the need for careful documentation as a key facilitator of effective data sharing, Fienberg et al. (1985), in an era prior to ubiquitous networks, suggested that data might be shared in various forms and venues, whether informally as appendices to books and papers or via formal archives and libraries. Feinberg observed that data sharing affects the interests of at least five parties or stakeholders: the possessor of the data set (primary researcher), the data requestor, the research participants, the scientific community, and society. Unfortunately, these interests are typically in conflict with one another. Fienberg et al. pointed out that data requestors and society typically favor data sharing practices, while primary researchers and research participants generally oppose it. Those who participate as research subjects tend to be concerned

about the security of their personal data, while primary researchers have issues related to trust and personal recognition that may be threatened by sharing their data.

As a continuation of his earlier work, Fienberg (1994) characterized the early 1990s as a time of increasing awareness of data sharing practices amongst researchers and health professionals. Fienberg attributed this shift in attitude to the allegations of scientific misconduct of the 1980s, as well as increasing public pressure for greater accountability by research institutions and universities. NSF Director Erich Bloch made clear the influential agency's expectations of conduct within scientific research in 1989,

> The NSF advocates and encourages open scientific communication. The NSF expects ... investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections, and other supporting materials created or gathered in the course of the research. (Fienberg, 1994, p. 6)

A landmark data sharing initiative in the 1990s that gained international attention was the Human Genome Project (Contreras, 2010). In 1996, policymakers met with a team of scientists to find a means to aid timely collaboration between researchers on an international scale. The result was the Bermuda Principles, which set the foundation for future use of data repositories on a global scale (Contreras).

The Researchers International Network (RIN) published a comprehensive study in 2008 that highlighted the value of data sharing within scholarly research. Their findings were summarized in a report depicting data sharing practices amongst the diverse disciplines of science. Further, the RIN report acknowledged that research data

should be made publicly available for two essential reasons: to be tested and validated further and to be reused in future research.

In 2008, Griffiths published an article that discussed the RIN (2008) report and provided an assessment of the various disciplines of science outlined in the report and their tendency to share and publish data (see Table 1). Based on study findings revealed in the RIN report, Griffiths summarized the detailed findings and assigned each discipline the categories of low, medium, or high to reflect differing levels of attitude towards sharing data, infrastructure-related barriers to publishing data, the effect of data publishing policy initiatives, and willingness to publish datasets, including metadata and documentation.

Table 1

RIN Study Summary of Data Sharing Attitudes by Discipline (Griffiths, 2008)

	Culture of sharing data	Infrastructure related barriers to publishing data	Effect of policy initiatives encouraging data publishing	Likelihood to publish datasets (metadata & documentation)
Astronomy	High	Low	Medium	High
Crystallography	Medium	Low	Low	High
Genomics	High	Low	High	High
Systems biology	Medium	Medium	High	Medium
Classics	High	High	Medium	Medium
Social and Public Health Sciences	Low	Low	Low	Low
RELU	Medium	Low	Medium	
Climate Sciences	Low	Low	Medium	Low to Medium

The complexities associated with the social sciences, for example, are described in the RIN study report and helps to explain why the social science culture does not typically champion the sharing of research data. Conversely, data sharing has proven to be the norm within the disciplines of astronomy, meteorology, and the physical sciences (RIN).

In 2015, the NSF espoused a comprehensive vision for the 21st century that highlighted its support for effective data sharing practices within scientific communities. Large volumes of digital data produced by scientists and engineers have become increasingly difficult to manage and access in a timely manner. Moreover, the NSF's appointed NSB Task Force on Data Policy acknowledged the need for action to understand the elements that contributed to effective data sharing practices and to establish comprehensive data sharing policies and procedures across all scientific disciplines (NSF, 2015). As a result, funding agencies that support human subject research has increasingly encouraged data sharing, as evidenced by the NSF's new default requirement in its grant applications that mandates a description of how data used in such funded research would be shared in a timely manner with fellow researchers (NSF, 2015).

Another factor for consideration when discussing data sharing strategies is the increasing volume of data being captured today. Large data sets, also known as "Big Data", is a common component in successful grant applications (Manyika et al., 2011). The proliferation of data coupled with the growing popularity of the Internet and social

media, is contributing to growth in size as well as the number of data sets. Big Data is adding significant complexity to the effectiveness of data sharing strategies (Manyika et al.).

Benefits of Data Sharing

A review of the literature illustrates not only the increasing importance of data sharing practices within the scientific community, but also the key benefits that data sharing provides. As Feinberg et al. (1985) argued, if all scientific research were conducted in an ideal fashion, the entire scientific community would have access to all scientific findings. As such, scientific societies would promote data sharing and journal editors would require it as a precursor to publication (De Wolf, Sieber, Steel, & Zarate, 2005).

Data sharing enables fellow researchers to re-analyze original research (De Wolf et al., 2005). This is important for several reasons. Re-analysis promotes new ideas, methodologies, and hypotheses, while facilitating creativity. The sharing of research data promotes suggestions for improvements. Further, re-analysis provides an opportunity to assure quality research, allowing refinements to proposed theories, or to provide a means to refute original study findings (Feinberg et al., 1985).

Advantages associated with the sharing of scholarly data are numerous and well documented. Arzberger et al. (2004) acknowledged access to publicly funded data empowers decision-makers with factual information to address complex issues. Greater data transparency allows citizens to participate in the decision making process and guarantees greater legitimacy and accountability of administration within a democratic system. In healthcare, for example, the advantages have been lifesaving. Data sharing

facilitates timely research to cure disease and to reduce suffering, providing quality healthcare via improved collaboration. Within the natural sciences, data sharing strategies have been considered the norm for generations (NSF, 2007).

De Wolf et al. (2005) suggested most researchers welcome the opportunity to share information with fellow researchers. Data sharing provides an efficient way to make comparisons across populations and to build upon one's own data with additional related data. Further, researchers who work across diverse disciplines may even *require* shared data, leading to reciprocal sharing between researchers. Moreover, shared data provides common ground for resolving controversy over disputed analyses of research findings (De Wolf et al.).

Various types of data sharing take place between researchers; for example, informal sharing with close colleagues or students is common. De Wolf et al. (2005) described how informal sharing involves elements of trust and integrity, integrity being the basis of confidentiality between parties. Collaborative re-analysis, reciprocal exchange of data, and unilateral sharing where confidentiality is implicit and a contract is unnecessary are characteristics of informal sharing. More formal data sharing arrangements include specific projects organized for sharing, public data archives, and research data centers or restricted access archives.

In addition to the timely dissemination of research findings, an important aspect of quality research is that experimental observations must be reproducible in order to be accepted as credible (MacMillan, 2014, Roche et al., 2014). Replication and validation offer opportunities for errors in data collection and interpretation to be identified and

corrected. Fienberg et al. (1985) asserted the replication and verification of research findings are the most frequently discussed advantages of data access in related literature.

To substantiate further the need to replicate research, Feldstein, as cited in Fienberg et al. (1985), argued that large data sets and complex analyses in social science research increases the potential for programming and statistical errors in study findings and "if anyone relies on one study, he runs the risk of being misled by an error or fluke" (p. 125). Another key point is that replication is valuable to facilitate refinements in original research (Fienberg et al.). Even so, Fienberg et al. cautioned a potential consequence is that replication may uncover discrepancies with original study findings.

Accordingly, Collins (2001), in discussing research replication and validity, argued that tacit knowledge and trust are essential to successful to replication of research. Collins' argument was that three elements must first exist to achieve reproduction of a research measurement: the experimenter must be able to master the previous researcher's tacit knowledge, the subsequent researcher must be sure the previous experimenter achieved the result (trust), and the experimenter must understand fully the difficulty of the procedure for the sake of perseverance. Replication requires that the methods and tools used to generate and manipulate study data be available to other researchers. As an outcome of his comprehensive study, Collins discovered a paradox of replication; successful repetition leads to trust, but more importantly, trust leads to successful repetition.

The Inter-University Consortium for Political and Social Research (ICPSR) is a long-standing membership-based organization that was established in 1962 at the University of Michigan. Today, this consortium of more than 700 academic institutions

provides access to the world's largest archive of training and research data for the social sciences (ICPSR, 2016). ICPSR supports social scientists globally by offering training and quantitative social analysis.

To ensure that data resources are available to future generations, ICPSR preserves data and migrates datasets to new storage media as dictated by modern technological changes. For long-term archiving purposes, ICPSR prefers that data be presented in a non-proprietary, text-based format, such as XML. The consortium assists researchers in conducting projects and identifying data for analysis. ICPSR uses encryption and various security measures for archiving restricted-use data files to protect confidential personal data (ICPSR, 2016).

Pampel et al. (2013) described a holistic approach to ensure that research data repositories (RDR) provide sustained accessibility, reliability, and stability of research data. The authors discussed the value of a RDR registry within the scope of project re3data.org - Registry of Research Data Repositories. As of February 2016, this registry had indexed and listed over 1,400 data repositories and had facilitated the identification of data storage and provided for comprehensive data search capabilities via information icons (re3data.org, 2016).

Regulations that Support Data Sharing

Changes in U.S. funding agency guidelines and federal regulations have been evolving over time to encourage a culture that acknowledges and supports effective data sharing. Even in light of increasing regulations, Schofield, Bubelaz, and Weavers (2009) cautioned that past policies and procedures to address data sharing have not been successful in practice. Schofield et al. explained that NSF data sharing policies, for

example, are scrutinized increasingly in literature as ad-hoc, piecemeal solutions. Moreover, Schofield et al. argued that enforcement of existing policies regarding data and resource archiving has been inconsistent, despite the increasing attention data sharing has received in recent journals and funding organizations.

Geneticists and genomic researchers increasingly withhold data and research materials, even in light of stringent regulations (Schofield et al., 2009). While guidelines and policies may encourage data storage and sharing, enforcement is difficult in practice. Likewise, COSEPUP (2009) explained that many researchers are not versed in data annotation or database management. Consequently, lack of experience in these areas has affected researchers' ability to create research data, as agreed upon within the scope of NSF and NIH sponsored projects (Schofield et al.).

The NIH introduced new requirements in October 2003, requiring proposals for grants with costs greater than $500,000 within a single year to document data sharing plans or to provide an explanation of why such plans were not possible. At that time, some criticized the 2003 NIH mandate, which literature suggested was developed to change the culture of data sharing (Nelson, 2009). Critics noted key shortcomings within the NIH requirements, including the vagueness of guidelines to make research data available. As a result, some researchers have opted to disregard NIH requirements. Even so, NIH representatives defended allegations of policy vagueness; they argued that flexibility was built into regulations to avoid grouping research into a single category (Nelson).

In 2007, in an effort to promote improved data sharing practices, the NAS challenged researchers to make all research data, methods, and supporting documents

publicly accessible in a timely manner. NAS identified the stewardship of research data as a critical long-term initiative for stakeholders and research institutions. Likewise, Arzberger et al. (2004) explained that open access to publicly funded data provided greater returns from public investment in research, enabling decision makers with factual information to address complicated and often, international issues.

To increase investments in research and development and to facilitate entrepreneurship, the American Competitiveness Initiative (ACI) was signed into law in 2007. The Act introduced federal assistance to promote competitiveness in scientific research and development and supported partnerships between the private sector, government, and educational facilities.

Landmark legislation subsequently replaced ACI in 2007 when President George W. Bush signed the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act (America COMPETES Act). This law requires federal agencies to provide policy and procedures to facilitate the exchange of data and research between agencies, the public, and policy makers. Since then, further legislation to refine the COMPETES Act was signed into law on January 4, 2011, when President Obama signed the American COMPETES Reauthorization Act of 2010.

In 2011, the NSF published papers that described its vision to facilitate the storage of research data in a documented form that is secure, accessible, and well-managed (NSF, 2011). Even so, critics at that time suggested the NSF revise its data-sharing policy to address inadequacies in its guidelines. The NSF mandates that investigators share with fellow researchers the samples, primary data, collections, and

other supporting materials gathered under NSF study grants. Grantees are expected to encourage and facilitate such sharing. Privileged or confidential information should be released only in a format that protects the privacy of individuals and subjects involved. General adjustments and, where essential, exceptions to this sharing expectation may be specified by the funding NSF Program or Division/Office for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interest of investigators.

According to the National Science Board (2011), "the progress of science and engineering has always been dependent on the collection of data" (p. 1). In 2010, the National Science Board established the Task Force on Data Policies under the Committee on Strategy and Budget whose goal was to review and refine NSF data policies to address modern challenges and to identify workable solutions that more effectively used digital research data to meet the objectives of the NSF. The NSF encourages openness and transparency, which are key components to the progress of science and engineering initiatives. Further, data management strategies are essential to the development of data sharing policies and guidelines. Key stakeholders are required to define modern data sharing polices, most notably the nation's science and engineering research enterprise (National Science Board).

On January 18, 2011, the NSF implemented a change in its data sharing policy by requiring that all research proposals include a data management plan (DMP) in the form of a two-page supplementary document (See Appendix A). The changes were tailored to address modern trends and needs of the research community (NSF, 2011). There is a growing consensus that researchers from diverse disciplines must rely on

improved collaboration and communication to address complex problems. To accommodate such disciplinary diversity, the NSF acknowledged that a modern research digital data sharing policy must accommodate diverse business models (National Science Board, 2011).

More recent legislation was proposed with the Federal Research Public Access Act (FRPAA), intended to stimulate data sharing and open access. FRPAA was introduced to Congress three times, in 2006, 2010, and 2012. Many believed FRPAA would "liberate more research literature than any other policy ever proposed in any country" (Hane, 2010, p. 1). While the FRPAA bill was not passed, a more robust version of the bill, the Fair Access to Science and Research Technology Act (FASTR), was introduced to Congress in 2013 and 2015 (Wright, 2015). FASTR would require U.S. departments and agencies with annual research expenditures in excess of $100 million to make final manuscripts from funded research publicly available online (Wright). The Senate Committee for Homeland Security and Governmental Affairs approved the bill unanimously on July 29, 2015. FASTR will be presented to Senate for final review and approval (Wright).

Data management plans are becoming an increasingly important requirement of modern research (Kitchin, 2015). In addition to the formal NSF statement on data sharing and the introduction of DMP requirements, on October 1, 2015 the Department of Energy (DOE) released a formal policy for digital research data management which also requires a data management plan with all new proposals (DOE, 2015). The policies do not define how the data management plan should be implemented, however, but rather mandates that a plan be developed and submitted with federal grant proposals.

NSF's Public Access Plan was also released in 2015 and highlighted several short and long term initiatives planned to enhance current operations related to public access of federally funded research, including a plan to offer guidance with DMPs and a repository to store research articles (NSF, 2015):

- The implementation of a system to enable NSF-funded researchers' articles (either manuscript or version) to be made available to the public via the NSF Public Access Repository (PAR); This system will be voluntary for researchers in 2015 and mandatory for NSF proposals submitted from the January 2016 effective date;
- Enhancements that may offer guidance with DMPs;
- Communications with the research community and stakeholders to identify gaps where guidance is needed;
- System flexibility for expansion and growth to allow for changing technologies.

Barriers to Data Sharing

Within the social sciences, literature suggests there are numerous barriers to data sharing. Known barriers to sharing practices in related literature may broadly be attributed to the following criterion (MacMillan, 2014):

- cultural or disciplinary constraints;
- inadequate data preservation infrastructure;
- lack of researcher rewards and acknowledgement; and
- perceived expense to support data sharing practices.

Another barrier that may impact data sharing is the lack of adequate training and guidance on data management plans required by funding agencies (Akers & Doty, 2013;

Tenopir et al., 2011). Tenopir et al. suggested organizations often do not provide support for long or short term data management. Likewise, data management needs vary by discipline, contributing to the complexity (Akers & Doty; Borgman, 2012).

To further illustrate the diversity of data sharing barriers, in 2012 the Emory University Libraries in Atlanta, Georgia conducted a study with 330 of the university's faculty researchers (Akers & Doty, 2013). Study outcomes revealed the top three reasons their researchers elected not to share data:

- Nature of the data - sensitive or personal;
- Concerns with researcher recognition or acknowledgement; and
- Data misrepresentation.

Likewise, Savage and Vickers conducted a study in 2009 that revealed some researchers opted to not share data because it was too much effort to provide raw data. Savage and Vickers concluded that researchers often fail to annotate their data and therefore lose the understanding of their datasets over periods of time. The authors noted the following concerns impacted data sharing practices of researchers in their study:

- concerns about publishing opportunities in the future;
- patient privacy with clinical data; and
- retaining data rights.

Literature suggests that policies and procedures themselves can sometimes create barriers to sharing data. Interestingly, Campbell et al. (2002) acknowledged federal agencies often impose strict policies to maintain secrecy in conducting federally funded research.

Another controversial barrier to data sharing discussed in literature is the usability of metadata as a tool to describe accurately the characteristics of research data. Critics have argued that the use of metadata as a tool to describe other data remains a highly controversial practice, asserting metadata fails to capture data context and tacit knowledge of the original study. Birnholtz and Bietz (2003), for example, explained data standards have been difficult to achieve because metadata models are not as simple as they appear. Further, they argued metadata models are not an effective long-term solution to describing research data. Likewise, the RIN report documented that the effectiveness of metadata schemes in data sharing practices have received varied reviews from researchers (Griffiths, 2008).

Ensuring that data is usable to researchers in different subject domains is a difficult task and can also create barriers. Advocates of metadata suggest data sets must be supplemented by people other than the primary investigator (NSF, 2007). Likewise, the COSEPUP (2009) described the important value of annotating preserved data so that data is understandable and retains value on a long-term basis. Annotation may include algorithms or other processing techniques used in research. Fienberg et al. (1985) argued researchers from the same discipline require less metadata than researchers from disparate areas.

Effective data sharing practices often require written instructions to supplement data sets. Knowledge transfer is not achieved easily by sharing written sets of instructions, but rather is a highly social process of learning practices that are not easily documented. Birnholtz and Bietz (2003) emphasized the social aspect of data sharing,

asserting knowledge of data sets is tacit in nature, and documentation frequently does not provide the understanding that others require.

Using the TEA laser as evidence (as cited by Collins, 2001), Birnholtz and Bietz (2003) described how the laser could not be replicated in different environments by following explicit instructions alone. Rather, successful replication required personal interaction with someone who previously assembled the laser. Similarly, Alavi and Leidner (2001) explained the value of specific knowledge as a prerequisite to processing data. In their paper, Alavi and Leidner described the intricate relationship between data, information, and knowledge.

One of the most controversial and highly discussed issues in data sharing is the protection of confidentiality and privacy of research study participants and patients (De Wolf et al., 2005). The NIH regulations are important to Institutional Review Boards (IRBs) that provide oversight for informed consent and confidentiality of data sharing, and likewise for researchers who have a responsibility to protect the identities of research participants. In complying with provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the Privacy Rule permits the use of health information without an individual's authorization when an IRB approves a waiver of authorization (De Wolf et al.).

Scientific disciplines differ in their needs and require more specialized attention to achieve successful data sharing strategies. The increasing complexity of research questions warrants data from diverse disciplines (Arzberger et al., 2004). Much is written in the literature on the difficulties in standardizing documentation to support the

diverse disciplines of science. Further, Arzberger argued, "diversity in science suggests that a variety of institutional models and data management approaches will be needed" (p. 1).

While some communities are open to sharing and use established repositories, Nelson (2009) argued there are no guarantees on the success of data repositories. For example, Nelson described how physicists, mathematicians, and computer scientists, in particular, currently access a centrally located repository at Cornell University, while molecular biologists successfully use the Protein Data Bank and GenBank archives, amongst others. Nelson's argument was that institutional repositories are not always valuable resources that promote data sharing practices, further substantiating the phrase "if you build it, they may not come" (p. 2).

To buttress Nelson's (2009) position, Birnholtz and Bietz (2003) argued that data standards have been difficult to achieve because they do not consider the individual role of data within individual research communities. Nelson cautioned that there are many reasons why scientists do not choose to archive their data, even when an available archive exists, as evidenced by a $200,000 data sharing initiative located at the University of Rochester (U of R) in upstate New York (Parry, 2010). The U of R's digital archive remained virtually untouched for years. Parry explained that professors failed to use the repository because it that did not assist article-writing management nor recognize researchers individually. After identifying the source of the problem, open-source institutional repository software was created to spur interest by both graduate students and professors (Parry).

Social scientists are challenged to exploit advancements in technologies to define new access models that preserve data and create scientific communities (Lane, Heus, & Mulcahy, 2008). For example, the Interagency Working Group on Digital Data under the National Science and Technology Council was established in 2009 as a federally regulated working group to analyze data collection, sharing, and management strategies. As a result of its research, the working group concluded that communities of practice are essential to today's evolving digital landscape (COSEPUP, 2009).

Scientific communities from diverse disciplines tend to use data in different ways (Lane et al., 2008) and the nature of usage makes the standardization of data storage requirements for storage in repositories a significant challenge. For example, Lane et al. discussed the differences between theoretical and experimentalist modelers, explaining that theoretical modelers require empirical data to validate their models, where experimentalists are able to benefit from theories enabled by modeling.

Another key barrier to data sharing is data accessibility. Simply because data is available does not mean it is readily accessible. Griffiths (2008) discussed some of the obstacles to data accessibility in the social sciences, such as licensing requirements, data sets that are too large to download, fees and charges, and confidentiality issues. Even so, legitimate reasons for keeping data private or delaying its release may be warranted. Granting access to research data prior to reporting results based on those data can undermine the incentives for generating the data (COSEPUP, 2009).

In scientific communities, having data often signifies status. Researchers who have access to their own data are viewed as *better* than using data sets borrowed from other resources. Thus, some researchers are reluctant to use others' data, as well as

reluctant to share their own data (Birnholtz & Bietz, 2003). Borgman (2010) explained that researchers use various methods to collect data depending on the experiment's purpose. Borgman argued that those purposes and methods influence what researchers consider their personal datasets and likewise the conditions under which they are willing to share data sets with others.

In their 2003 paper, Birnholtz and Bietz discussed task uncertainty, or the degree to which researchers agree on problems to be solved and the appropriate methodologies to solve them. In areas with low task uncertainty, for example, a stigma may be attached to researchers who use shared data. Where there is high task uncertainty, there is more variation between researchers. Likewise, there is greater innovation in the experiment's design and unique ways of collecting data. The process of analyzing another researcher's data bypasses this step, which is perceived as *less* prestigious in the competition for scientific reputation.

Competitiveness and personal recognition for a researcher's work are other key barriers that affect data sharing efforts (Borgman, 2012; Nelson, 2009). Nelson suggested another issue facing journals and data banks is how to ensure proper citations for data sets. According to COSEPUP (2009), "most researchers prefer to pursue new goals rather than devote effort to making their existing and past data useful for others" (p. 99). Nature (2014) suggested co-authorship should be the norm when sharing datasets.

Other factors that limit the sharing of scholarly data are financial costs associated with data sharing, storage, and reuse. Fienberg et al. (1985) argued data sharing should also require the distribution of costs beyond the primary investigator. Fienberg et al.

argued that data sharing should require distribution of documentation and data transfer costs to future research analysts. Often, a researcher's decision to retain study data and documentation varies by the cost or the ability to reproduce the experiment (Borgman, 2010). Even so, 2007 NSF guidelines mandate that researchers share primary data with others "at only incremental cost and within a reasonable time" (Pittman, 2010, p. 46).

Data sharing adversely affects economic interests of researchers in lost revenues (Birnholtz & Bietz, 2003). Rent, or revenue, can be earned from data via publications. Pre-publication sharing practices often risk the loss of such financial benefits. Grudin's 1989 research (as cited in Birnholtz and Bietz), made an interesting point in this regard arguing it is unlikely for people to use a system if it requires additional work from which they will receive no benefit. Likewise, Cecil and Griffin (as cited in Fienberg et al., 1985), asserted "few of the benefits and most of the burdens fall to the possessor of the data set" (p. 148).

Blumenthal, Campbell, Anderson, Causino, and Louis (1997) conducted a comprehensive empirical study of 2,100 life science faculty to gain an understanding of data sharing practices in academia. Blumenthal et al. explored data withholding behaviors among life-science faculty. Outcomes from their survey suggested an association between higher publication rates and data withholding behaviors, especially affecting researchers' intentions to share requested research results. Blumenthal et al. argued that industrial funding of university research has most frequently been cited as a cause of data withholding among academic scientists.

Literature suggests that withholding data sets is often required to commercialize university research through activities such as patent applications. NIH estimated a 60-

day delay was reasonable. However, more recent studies have shown that many sponsors of academic research from life science companies require researchers to refrain from publishing research results, often for more than six months to protect the commercial value of the results (Arzberger et al., 2004).

Blumenthal et al. (1997) argued that strong personal and external pressures often affect a researcher's ability to share data. Personal pressures include competition between researchers for priority and recognition. External pressures include the requirements of the promotion and tenure process, competition for funding, and processes and procedures related to the commercialization of university research.

In their discussion on the value of retaining data in scholarly research, Fienberg et al. (1985) suggested that the loss of valuable data takes place far more often in small research projects as compared to large projects for a few key reasons. Fienberg explained that researchers of small studies often have difficulty deciding which data to retain for future use, how to document data thoroughly, and how to find available funds within limited budgets for archiving data properly.

In 2011, Tenopir et al. published a comprehensive paper that discussed the findings of an NSF-funded survey conducted by the research team of DataONE. The purpose of the 2010 survey was to gain an understanding of modern scientists' data sharing practices, barriers, and enablers. This international survey polled 1329 scientists, 75% of whom were from North America. Tenopir et al. explained that PARSE Insight had conducted a similar study in 2009 with 50% of respondents from the European Union (EU). Interestingly, findings from the two studies revealed distinct similarities and differences regarding the data sharing practices of scientists.

The PARSE Insight study identified legal issues and data misuse as key barriers to data sharing. Conversely, the DataONE study identified time constraints of researchers and lack of funding as key barriers. Both studies suggested that researchers' ability to address scientific questions was limited because their access to data generated by others was restricted. Based on the DataONE study, a key recommendation to improve data sharing strategies is a strong focus on organizational policies and procedures. Study outcomes suggested organizations tend to promote individualized approaches to science, overlooking important tools, such as metadata (Tenopir et al., 2011).

Significant cultural differences in data sharing practices exist between the U.S. and EU. For example, effective January 18, 2011, NSF began to mandate the submission of a two-page maximum Data Management Plan (DMP) with all proposals involving data collection. The NSF allows grantees principal legal rights to intellectual property developed under their grants, although the NSF acknowledges that researchers have a responsibility to share data collections and results with others in the scientific community.

Rather than imposing a similar mandate, the European Commission strongly urges member states to develop policies that recognize the importance of data access and dissemination (Tenopir et al., 2011). In July, 2013, the European Commission held a public consultation in Brussels to discuss the direction of open access to research data (Dimitrova, 2013). The recommendations of key stakeholders at the consultation influenced subsequent negotiations to revise EU's data sharing policies in their major research program, Horizon 2020 (Dimitrova). EU has supported open access in research,

however, Dimitrova explained that many exceptions to data sharing exists in several areas (e.g. privacy, national security, and intellectual property).

Table 2 presents a summary of common factors that have adversely affected data sharing efforts as outlined in this literature review. Technology-related obstacles and inadequate documentation were identified as the most prominent factors affecting data sharing in the 1980s. Over time, as advancements in technology have increased, literature acknowledged other contributory factors, such as time, funding, trust, privacy, metadata, confidentiality, and behaviors of researchers.

Frameworks in the Literature

A variety of frameworks to promote data sharing strategies has been suggested in literature. Fienberg et al. (1985), for example, proposed one of the most creative sharing strategies, arguing that data sharing practices may be enhanced by assessing financial penalties for *not* sharing. In keeping with Feinberg's discussion, COSEPUP published a comprehensive examination in 2009 of the consequences of how change influences research data in terms of stewardship, integrity, and accessibility. In its study, COSEPUP acknowledged the increasing need for a fresh approach in the design and the management of research projects, recommending that researchers receive adequate training in managing research data.

Table 2

Literature Review Summary by Common Factors That Affect Data Sharing Practices

Factor	Author(s)
Careful documentation	Fienberg et al. (1985)
Competitiveness	Blumenthal et al. (1997) Harttner, Ryan, Mackenzie, Parker, & Strasser (2013)
Cultural norms and expectations	Griffiths (2008) Harttner et al. (2013) NSF (2007) Tenopir et al. (2011)
Data integrity, accessibility, stewardship	COSEPUP (2009)
Funding/cost	Roche et al. (2014) Tenopir et al. (2011)
Innovation in digital technologies	COSEPUP (2009)
Lack of technological infrastructure	NSF (2007)
Metadata models	Birnholtz & Bietz (2003)
Policy concerns/lack of standards	Arzberger et al. (2004) NSF (2007)
Privacy/confidentiality/ethical protection	Birnholtz & Bietz (2003) COSEPUP (2009) Lane et al. (2008)
Training concerns	COSEPUP (2009)
Trust	Birnholtz & Bietz (2003)
Recognition of ownership/reward systems	Akers & Doty (2013) Borgman (2012) Nature (2014) Tenopir et al. (2011)

Birnholtz and Bietz (2003) proposed that effective data sharing systems rely on an understanding of the roles that data play in scientific communities. The authors explained that data play two general roles: to serve as evidence to support scientific inquiry and to make a social contribution to the establishment and maintenance of communities of practice. The authors suggested that use of these roles may greatly enhance the future development of effective data sharing systems. Based on their research, Tenopir et al. (2011) explained that modern scientists have a special interest in protecting their data, which may be due, in part, to professional development and concerns related to tenure and promotion.

Chapter 3

Methodology

Introduction

This study utilized a qualitative and exploratory approach. A literature review of current data sharing practices in academic research and NSF federal grant application and procedures guided the research. Recent literature confirms the lack of empirical studies to explore factors that affect modern data sharing practices, even in light of its increasing importance. Maxwell (2013) supported the use of qualitative methods when little information is known about a subject or phenomenon. Likewise, Fink (2003) recommended qualitative survey analysis for the exploration of meanings and experiences in the social sciences.

This multi-method study used two qualitative approaches: semi-structured interviews and a qualitative survey instrument. Morse (2003) explained that a broader dimension can be achieved in research by combining and increasing research strategies. A multi-method design promised a better understanding of the research topic and a more comprehensive achievement of the study's research goals. Findings from both the qualitative interviews and qualitative survey were triangulated to discern common patterns as the basis for theory development.

Literature Review

The value of conducting interviews in qualitative research is well-documented (Maxwell, 2013; Tashakkori & Teddlie, 2003; Warren, 2004). Warren explained that qualitative interviewing in research is an invaluable tool to understand how people live

and operate in their worlds. Without this understanding, Warren argued that useful research is impossible.

As an advocate of qualitative research, Warren (2004) discussed the importance of interviews as the most frequently used research method in the social sciences. Since the 19th century, interviews have been used in the social sciences as scholars sought answers to questions about the human condition. Even so, while widely used, the qualitative interview has been subject to considerable debate in the research community. Advocates of quantitative methods and qualitative methods alike have been critical of rigor in research procedures, research methods, and the validity of research findings.

In their 2003 handbook, Tashakkori and Teddlie provided insights into the history of quantitative, qualitative, and mixed methods in modern research. The authors explained that qualitative research has emerged in popularity over the past two decades, most likely in reaction to researchers' dissatisfaction with the traditional methodology practices of quantitative research. Qualitative research is an inductive approach based on a worldview of constructivism, unlike qualitative research, which is based on the worldview of post-positivism and deductive approach.

Maxwell (2013) provided further insight on best-use cases for each research method. The author explained that within a research design, its method should reflect the nature of the study as well as the research questions and goals. Qualitative research is indicated, in part, when there is a need to develop causal explanations, understand meanings and context, and identify unanticipated phenomena or influences.

Research Methods and Procedures

A series of logical decision-making choices characterizes a research design framework and highlights the steps that link philosophical assumptions to specific methods (Creswell & Plano Clark, 2011). As outlined in current literature, this study followed key elements that comprise research design, including study purpose, type of investigation conducted, role of the researcher and extent of interference, study setting, measurement and measures, sampling design, and data collection and analysis. Figure 1 illustrates the research design for the study.

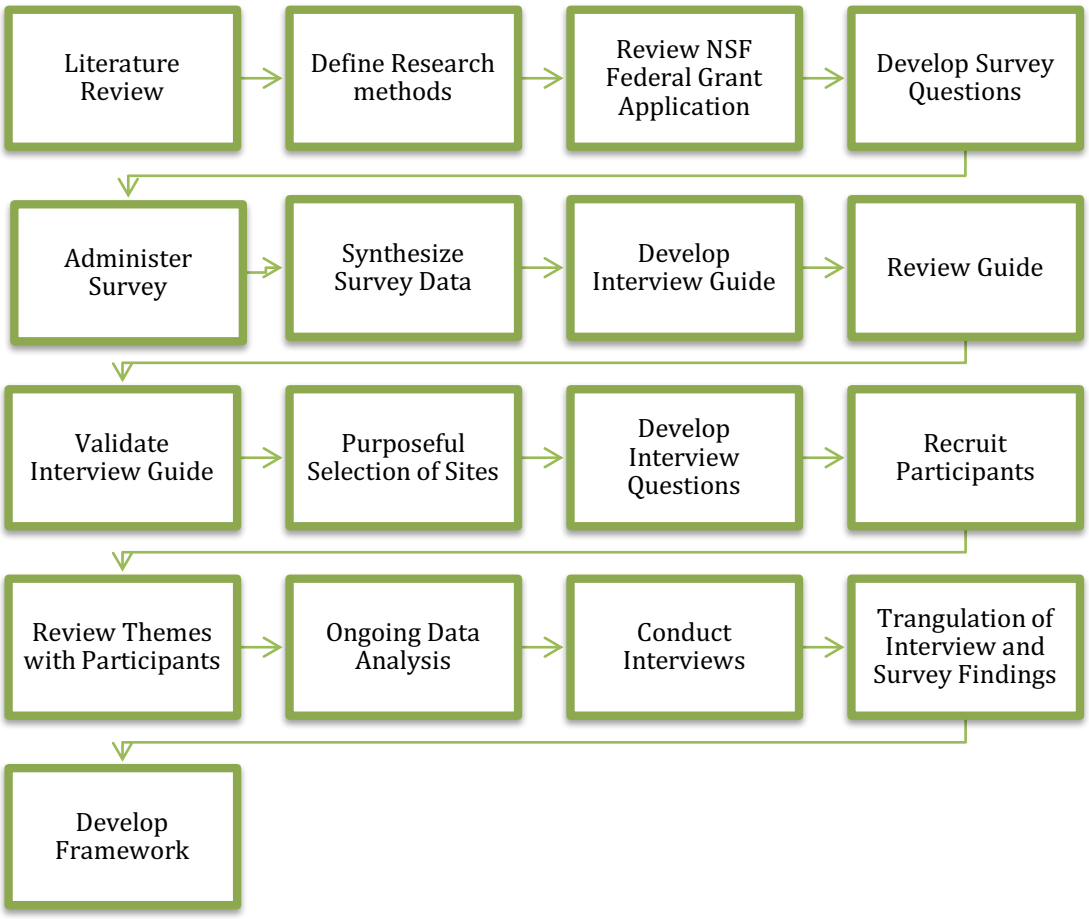

Figure 1. Research design framework.

The Qualitative Paradigm

The intent of this study and the nature of research questions guided the selection of qualitative research design over other research methods, such as quantitative or mixed methods research. Qualitative research empowers investigators to discover rich, in-depth knowledge from subject experts based on real-life settings and experiences. Maxwell (2013) explained that qualitative design is often the research method of choice when little is known about the research subject. A multi-method qualitative research design was selected for this study due to its exploratory nature to investigate data sharing practices in academic research.

Qualitative studies are typically considered exploratory and data is gathered through interview and observation. Unlike quantitative research where hypotheses are derived from theory, qualitative research is characterized by the understanding that is derived from patterns identified in aggregated data (Creswell, 1994). Qualitative research is an inquiry process of understanding, where researchers develop a holistic perspective of a process or phenomenon (Creswell & Plano Clark, 2011).

Literature suggests that within qualitative research, grounded theory methodology is a valuable approach for use in the interpretation of interview data (Maxwell, 2013). With grounded theory, an in-depth understanding of a situation may be gained through observation, conversation, and interview. The resulting theory is an explanation of categories and the relationships that exist between them. The emergent theory is grounded in the data.

Research Questions (RQs)

The goal of this case study was to develop a framework to facilitate data sharing within the scope of the NSF grant application process by addressing specific research questions:

RQ1: Within the past two years, what specific relationships, if any, can be drawn between the Principal Investigator (PI) and the sharing of research data and materials, in keeping with current NSF federal grant application guidelines?

RQ2: In what ways can the NSF federal grant application process be improved to facilitate timely sharing of research.

Role of the Researcher

In this study, the researcher learned from participants using observation of documents, open-ended interviews, and qualitative survey data. Documentation, such as audio-visual materials and meeting minutes were gathered in conjunction with qualitative interview data. The researcher asked probing questions to guide participants to share information from their personal perspectives and life experience. While semi-structured interviews followed an interview guide, the researcher welcomed open discussion with participants.

Maxwell (2013) explained that the researcher plays an influential role in the interview process and their behavior affects the validity of interview data and surveys. To minimize researcher bias or reactivity, it was important for study questions to be thoughtfully presented and delivered in such a way to elicit genuine responses from participants. It was also important for the researcher to practice administering the interviews and write notes from observations.

The relationship established between researcher and participant is the catalyst that facilitates the research. This relationship will change throughout the course of the study (Maxwell, 2013). The researcher acknowledges the importance of this relationship and must maintain a positive, communicative working rapport with participants, which is essential to the interview process. The researcher must respect the views and anonymity of study participants at all times.

NSF Federal Grant Application

The researcher reviewed the 2016 NSF federal grant application and the grant application's instruction guide (NSF, 2016b). This review provided a basis of understanding and ensured that interview and survey questions were relevant. A review of current NSF policies and procedures was beneficial to ensure a holistic understanding of the NSF federal grant application process.

Data Collection

A key component of research design is data collection methods. Unlike quantitative research, where literature review and planning of the research process takes place in the beginning of the process, data gathering begins as a first step in qualitative research to formulate plans and discover the nature of the research questions (Heath & Cowley, 2004). Data collection was achieved in this study by two approaches: semi-structured interviews and a qualitative survey instrument. The use of multiple research methods in data collection is a preferred approach to minimize the risk of bias in study findings (Maxwell, 2013).

The multi-method model this study followed was the QUAL + qual model, as defined by Tashakkori and Teddlie (2003). When using two distinct data collection

methods in multi-method research, the authors explained that the research project conducted first (QUAL), which is the dominant research and is considered the foundation of the study. The second project (qual) supports the primary research. In this study, the survey phase was the dominant research and the interview phase was the supporting research.

Survey Phase

Selecting Study Participants for the Survey

Survey participants included PIs and Co-PIs from universities located in Florida, Georgia, Alabama, North Carolina, and South Carolina. Purposeful sampling of survey participants was achieved by targeting universities from this region that were awarded NSF federal grants within the past two years to fund their academic research. The total population of survey participants selected for this study was 350 and the anticipated response rate was 30-35%. Creswell and Plano Clark (2011) suggested a smaller number of study participants allows a researcher the ability to gain more in-depth insight into perspectives that can become less manageable with larger numbers of study participants.

This researcher located PI and other stakeholder contact information at each proposed site. Once PI and stakeholder contact information was gathered, a written document was drafted that outlined this study's objectives and requested PI participation. This communication was distributed to selected participants via email communication.

While surveys are traditionally viewed as a quantitative source of data, they are increasingly used as a valuable component of qualitative research (Creswell & Plano Clark, 2011). In contrast to quantitative surveys that are characterized by closed

questions, qualitative surveys feature open-ended questions. A qualitative survey is the study of diversity, rather than distribution in a population.

The survey was developed and distributed via SurveyMonkey.com. Survey design followed traditional design principles, such as attention to the wording of questions with a focus on the appropriateness of question content, the type and sequencing of questions, and the use of professional language. The researcher organized the survey so that questions were delivered in order from general questions to more specific, with increasing difficulty as the survey progressed. The qualitative survey consisted of open-ended and multiple-choice questions.

Survey questions were developed to reveal data practices, perceptions, and attitudes of participants towards data sharing, the NSF grant application, and the NSF federal grant application process. Survey questions explored relevant topics, such as the collection and use of research data, views on data sharing, the type of data being used, collection and use of research data, the relationship between their organization and their data, the use of data across their area of research, responsibility to complete and submit the grant application, and views on the NSF application and process (See Appendix B).

Interview Phase

Selecting Study Participants for the Interviews

Qualitative research is characterized by purposeful selection of individuals and sites that provide essential information relevant to the study (Sekaran, 2003). In survey research, respondent selection is addressed in terms of representativeness, or how well study participants represent the phenomenon to be studied. Sample selection is often a

complicated process that involves several considerations, such as how to identify and locate participants and assessing cost.

In this study, interview participants included PIs and Co-PIs from five universities located in the state of Florida with the largest number of NSF grantees for the year 2011: University of Florida, University of South Florida, Florida International University, Florida State University, and the University of Central Florida. Creswell and Plano Clark (2011) explained that a smaller number of study participants allow the researcher a greater opportunity to gather individuals' perspectives along with their context, which becomes increasingly difficult with larger numbers of study participants.

Purposeful sampling of research sites was achieved by targeting universities awarded NSF federal grants within the past two years to fund their academic research. Further investigation was required to gain an understanding of NSF-funded projects at study research sites. The researcher located PIs and other stakeholder contact information at each proposed site. Once PI and stakeholder contact information was gathered, a document was drafted outlining this study's objectives as well as a request for PI participation. The researcher made initial contact with potential study participants for interview sessions via email.

To identify subject matter experts from NSF for interview sessions, this researcher reviewed the NSF organization chart for the Division of Grants and Agreements Office (DGA) from the NSF website. Similar to the process for contacting PIs, a written document was drafted that outlined this study's objectives and requested NSF participation. Although email requests were sent to the appropriate DGA Branch

Manager at NSF to request participation, no response was received and therefore, NSF's viewpoint is not included in this study as planned.

Once the survey phase was complete, semi-structured interviews were scheduled and conducted with PIs at the previously mentioned universities to gather detailed, open-ended information. Twenty-five interviews were planned with PIs and Co-PIs from the selected universities. The use of semi-structured interviews as a data collection method allowed participants the opportunity to express themselves and facilitate reliable qualitative data (Maxwell, 2013).

The researcher developed an interview guide that detailed the interview themes to be covered and to frame the direction of the interview (See Appendix G). Even so, discussions were not dictated by the guide, but rather were used in conjunction with open discussions during the interview. Mason (2004) discussed the value of an interview guide to conduct semi-structured interviews. This approach ensures that important questions are not overlooked during the course of the interview. Even with the use of an interview guide, flexibility can be retained and the researcher can change the direction of the interview as deemed appropriate (Maxwell, 2013).

When possible, the interviews were audio recorded for subsequent transcription and analysis by the researcher. In the event a recording was not possible, the researcher took field notes of interview conversations to provide contextual details of the interview. Each interview was conducted in 30 to 45 minutes and the researcher began interviews with introductions and requested permission from the participant that the session be audio recorded. Upon approval, the session recording began and

the researcher reiterated the objective of the study and proceeded with the interview questions following the format of the interview guide.

As the interview progressed, general questions became more specific. Interviews began by reviewing the nature and scope of the study. Pre-determined, open-ended interview questions were asked during the interview to explore data-related and NSF grant application procedures, such as the collection and use of research data, views on data sharing, the type of data being used, collection and use of research data, the relationship between their organization and their data, the use of data across their area of research, responsibility to complete and submit the grant application, and views on the NSF application and its process.

Data Analysis

Qualitative data analysis is an inductive approach that should be conducted simultaneously with data collection (Coffey & Atkinson, 1996). Ideas discussed during interviews were recorded as memos as data was analyzed and coded. Maxwell (2013) cautioned that a common problem in qualitative studies is allowing a backlog of unanalyzed transcripts and field notes to accumulate. Therefore, the researcher began data analysis immediately after completing each interview and throughout the course of the study to avoid a backlog.

By utilizing a grounded theory methodology in the data analysis, study findings were well-grounded in data. Four fundamental characteristics are evident in this approach. First, the resulting theory should fit the phenomenon being studied, as theory should be crafted from diverse data collected during the study. Secondly, the theory should provide a better understanding of the phenomenon being studied. Because the

data is comprehensive, it should also provide generality; theory should include variation and be sufficiently abstract to apply to various contexts. Lastly, the theory should provide control, stating the conditions to which the theory applies, while describing a reasonable basis for action (Strauss & Corbin, 1990).

In the final steps of the data analysis process, the research findings from survey data and coded interview data were triangulated and synthesized to identify common themes towards theory development. Common patterns, themes, and topics were analyzed and summarized. Detailed study findings are discussed in Chapter 4.

Data Analysis - Survey Phase

This researcher reviewed each participant's survey information and transposed the survey question data to the corresponding question on the Microsoft Excel survey spreadsheet matrix. Since this was a qualitative survey, all comments to the open-ended questions were labeled on the spreadsheet and categorized. Once all surveys were received and transposed to the matrix, the spreadsheet was reviewed to identify common themes. Data aggregation of survey data took place manually by the researcher; no software was used for this purpose. In addition, since the survey was housed on SurveyMonkey.com, several report options were available in order to summarize survey data and generate findings in the form of percentages and illustrative bar graphs.

Data Analysis - Interview Phase

Warren (2004) explained that transcripts are often used in the interview process as the basis for researcher analysis. The analysis of data collected during the interview consists of reading and rereading transcripts in the form of a Microsoft Excel spreadsheet to identify conceptual patterns. As an initial step in the analysis of interview

data, this researcher listened to each interview recording before transcription and made notes and memos for initial thoughts on relationships or categories. After each interview session, coding was achieved by examining notes and transcripts a sentence at a time, making comparisons between statements[1].

The processes of data collection, memos, and notes was achieved simultaneously. Memos are notes about a hypothesis formulated about a category or potential relationships between categories. The researcher used index cards during memo taking for sorting purposes. Brief biographical data about participants were also documented in the interview notes.

Using the first interview as the basis for subsequent interviews, common categories were identified. The second interview was coded based on elements from the first interview, and subsequent interviews helped to begin formulation of emerging theory. This approach was in keeping with grounded theory methodology that is characterized by constant comparison, initially from data set to data set, and from data set to theory.

As the interviews proceeded, the researcher identified core categories that began to emerge. Once core categories were identified, connections between categories were identified using the researcher's memo notes. Saturation occurred as interviews failed to add new information about a category and at that point, coding was complete.

---

[1] There are differing views in literature whether multiple data analysts are needed to perform the coding and analysis of qualitative data. For example, Janesick (2003) argued one person is sufficient and preferred to conduct all coding, where Patton (2002) suggested two or more data analysts are beneficial to ensure research validity.

Validity of Research Findings

A critical step in the research process was to ensure validity of research findings, surveys, and reports. Validity in qualitative research is also referred to as trustworthiness or credibility. Maxwell (2005) cautioned that "validity is a goal rather than a product; it is never something that can be proven or taken for granted" (p. 105). While researcher bias and reactivity are two key threats to validity, this researcher lacked any preconceived knowledge about the NSF federal grant application process and therefore was able to maintain a neutral perspective in conducting the study.

To help ensure the validity of research study findings, data from Phases One and Two were triangulated, or converged, as this research was a multi-method study consisting of survey and interview data. According to Schwandt (2015), triangulation "is a procedure used to establish the fact that the criterion of validity has been met" (p. 307). Likewise, Turner, Cardinal, and Burton (2015) discussed that while single research methods are inherently flawed, the flaws may be mitigated through a mixed methods research approach. Similarly, Mathison (1988) suggested "good research practice obligates the researcher to triangulate, that is, to use multiple methods, data sources, and researchers to enhance the validity of research findings" (p. 13).

Summary

This chapter discussed the methodology that was followed in this study. A qualitative and exploratory approach was used to review the NSF federal grant application process and data sharing practices in academic research. Specific research questions (RQs) were addressed:

RQ1: Within the past two years, what specific relationships, if any, can be drawn between the Principal Investigator (PI) or Co-Investigator (Co-PI) and the sharing of research data and materials, in keeping with current NSF federal grant application guidelines?

RQ2: In what ways can the NSF federal grant application process be improved to facilitate timely sharing of research data?

To broaden the dimension of the study and facilitate research goals, the researcher used a multi-method approach in its design: a qualitative survey instrument and semi-structured interviews. Findings from each method were analyzed and synthesized to identify common themes towards theory development.

Chapter 4

Results

Introduction

The following chapter presents research study findings from Phase One and Phase Two. Also included in this chapter is the triangulation of findings from both Phases that converges survey and interview session data. The purpose of this research was twofold: to identify potential relationships between PIs and timely data sharing and to identify recommendations for potential enhancements to the NSF proposal process to encourage the sharing of data in research.

Ethical Considerations

Three fundamental principles guided this study to ensure that ethical considerations were addressed, including respect for persons, justice, and beneficence. The researcher was committed to protect the anonymity of research participants. All participants underwent an informed consent process to confirm that they understood their involvement in the study so they could determine if they wished to participate. Lastly, the researcher sought to minimize social and psychological risks to study participants at all times. Study participants were ethically treated and the researcher respected their rights and perspectives.

Phase One Survey Methodology

Following the methodology described in Chapter 3, the survey was developed and distributed to prospective study participants. Using the NSF awards database accessible via the NSF website, this researcher identified contact information for

researchers who met the study's selection criterion: researchers who submitted an NSF grant proposal that was funded within the past two years, and at the time of proposal submission, was employed by an institution located in one of the following states: North Carolina, South Carolina, Georgia, Alabama, or Florida.

SurveyMonkey.com was the website the researcher selected on which to develop and then distribute the online survey (See Appendix C). Once researcher contact information was identified, an introductory email was drafted and distributed to 350 PIs and Co-PIs randomly selected requesting their participation in this study. A survey Participation Letter was drafted and included in an introductory email as an attachment, along with a brief introduction, the study objective, and a link to access the survey at SurveyMonkey.com (See Appendices D and E).

Because the participant response rate from the initial online survey distribution was relatively low (31 respondents or 8.9% of the population selection), a second request was distributed via Survey Monkey.com four weeks later. This distribution yielded an additional 18 survey responses. In an effort to improve the survey response rate, it was decided to print and distribute paper surveys via U.S. postal mail to members of the original population selection requesting their participation.

Nulty (2008) conducted a literature review of studies that compared response rates of online and paper surveys. His findings suggested that while there are advantages and disadvantages associated with each method, typically paper surveys achieve much higher response rates than online surveys. Therefore, 100 paper surveys were distributed to a random sampling from the original survey population via U.S. postal mail. The

paper survey yielded an additional 21 responses, bringing the total overall number of survey respondents to 70 or 18.4% of the total survey population.

Survey Data Analysis

The data analysis of survey data revealed that 98% of survey respondents were PIs, 1% were Co-PIs and 1% of respondents held another role in their institution (e.g. office clerks). When asked the total number of years of experience in their role as a researcher at their current institution, 35% responded 5-10 years, 29% 10+ years, 21% 1-5 years, and 15% of respondents had less than 1 year of experience, as detailed in Figure 2.

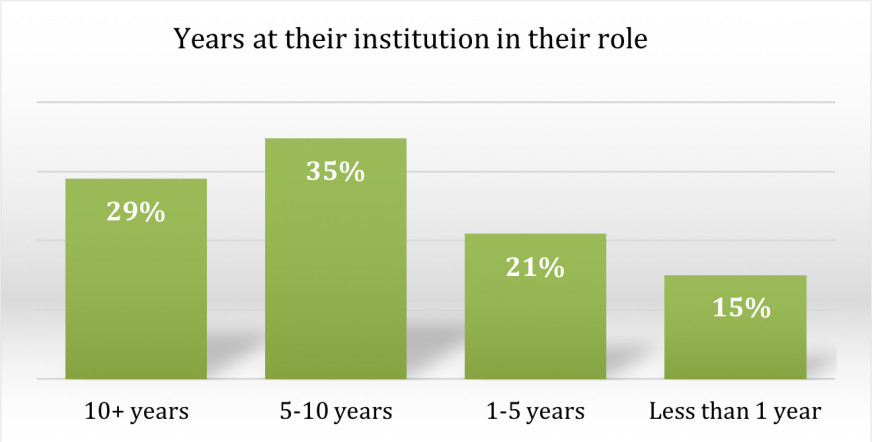

Figure 2. Researcher years of experience at their institution.

When asked how many NSF proposals survey participants had submitted to NSF in the previous two years, Figure 3 illustrates that out of 70 survey respondents, 8 participants (11%) had no response, 39 participants (56%) submitted 1-3 applications, and 23 participants (33%) submitted 3 or more proposals to NSF during that timeframe.

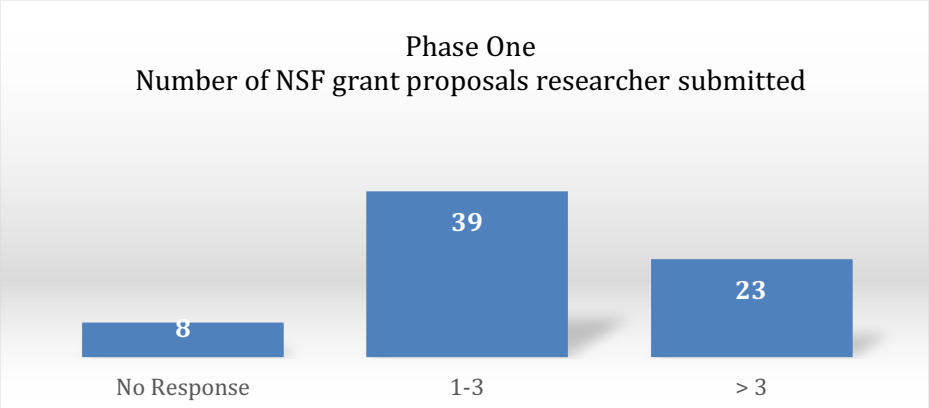

Figure 3. Number of NSF proposals the researcher submitted in previous two years.

Researchers were then asked the importance of data sharing in the scope of their current role as PI or Co-PI. In response, 73% identified data sharing as either very important or an important aspect of their current research, as illustrated in Figure 4.

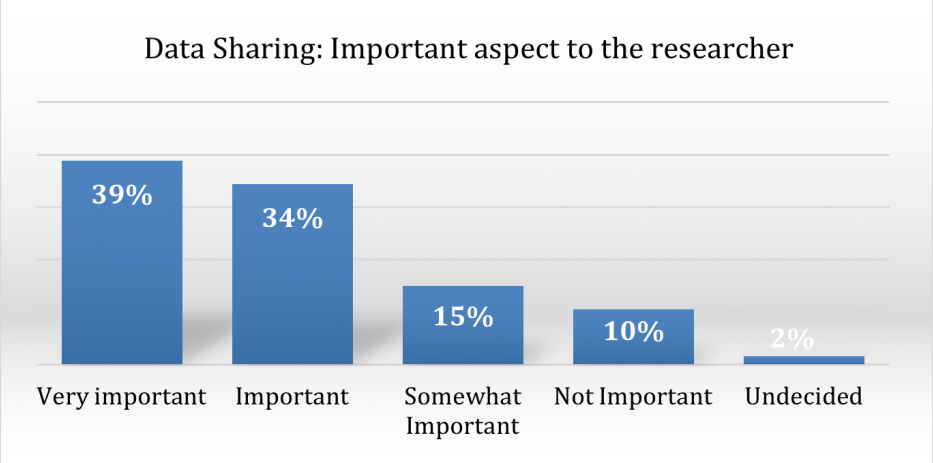

Figure 4. Importance of data sharing to the researcher.

Similarly, participants were asked about the overall importance of data sharing at their institution. Figure 5 illustrates that 19% of survey participants responded that data sharing was a very important aspect of research at their institution. In a broader context, 40% indicated data sharing was important at their workplace, 24% indicated it was

somewhat important, 6% responded that data sharing as not important at their institution and 11% of those surveyed were undecided.

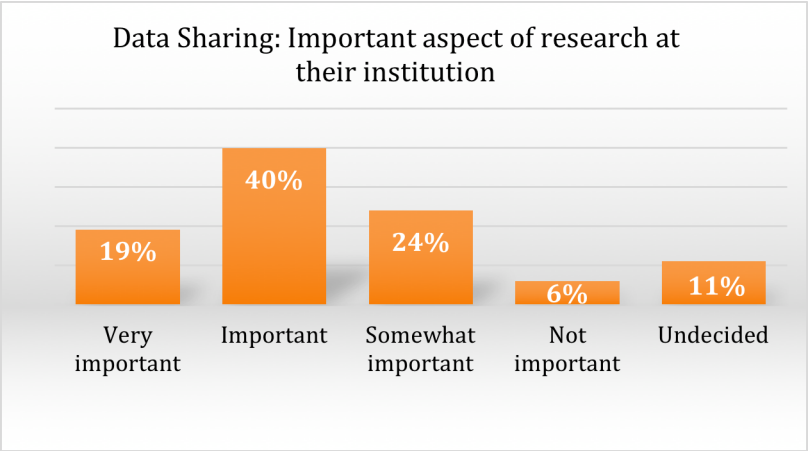

Figure 5. Importance of data sharing at the researcher's institution.

To gain an understanding of data sharing practices with researchers outside their institution, Figure 6 illustrates 69% of participants responded that they always or often shared data outside their institution and 29% seldom shared. Interestingly, no respondent indicated that they never shared data.

Figure 6. Data sharing practices outside the researcher's institution.

Another aspect of the NSF grant proposal is the supplemental DMP introduced as a requirement by NSF on January 18, 2011 to encourage data sharing practices in research. According to one researcher, "the data management plan requirement is a good one. Tools and training to create data management plans have raised awareness and standards of what should be done. The institution [NSF] now recognizes the need and is more willing to create infrastructure to facilitate data sharing." The researcher did not provide examples of action by NSF, but rather was stating an opinion that NSF may address infrastructure in the future. Conversely, other researchers said their institution "regards it [DMP] as a hoop to jump through. Once you've successfully jumped, you can ignore it in the future" and likewise another stated "I don't think it really serves much purpose."

When asked about the effectiveness of the DMP to encourage their data sharing, some researchers acknowledged the plan as an improvement in the NSF proposal process, however there were differing opinions about the effectiveness of the DMP to facilitate data sharing. Some survey respondents acknowledged benefits of the plan, especially in terms of its value to help organize researchers' thoughts and formalize the NSF proposal process. Survey results highlighting the overall effectiveness of the DMP to facilitate data sharing is illustrated in Figure 7.

Further analysis of researcher comments revealed a wide range of acceptance of the DMP to encourage effective data sharing. Some researchers acknowledged they support the DMP and accept the plan as a positive enhancement to the original NSF grant proposal. One researcher commented, "it's valuable as it forces all PIs to think about this [data sharing] explicitly. I am not sure if it has facilitated my data sharing

practices, but it has made me happy to see best practices recognized and I hope rewarded with more grants." Another suggested they would perform the practices outlined in the plan even if they didn't have to submit a DMP. Another PI suggested that the true purpose of the DMP was to provide, "accountability and visibility into taxpayer-funded research." Similarly, some acknowledged the DMP, "encourages the sharing, which is good. Data sharing should be done in any case as it is a part of doing good research."

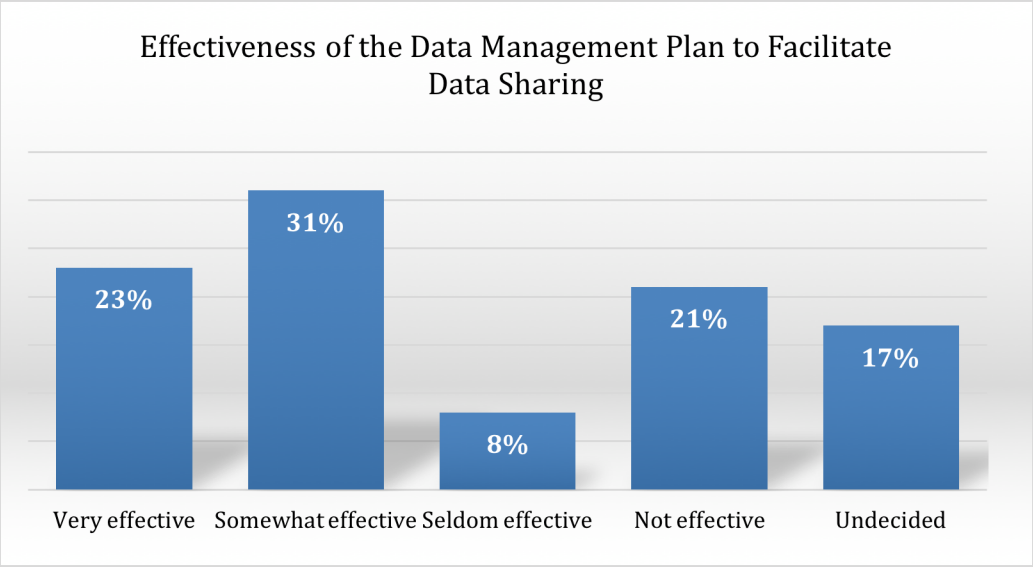

Figure 7. Effectiveness of the DMP to facilitate data sharing.

Conversely, some argued that the DMP was a wasted effort and has little impact on their data sharing efforts, describing the plan as "worthless" and recommending that it be removed from the proposal. Others suggested that NSF clarify the plan's objective and provide a clear template for what information should be included. From yet another perspective, some argued there are no consequences for those who complete the DMP but fail to share their data, asserting, "I have always shared all of my data immediately… how is data sharing checked? Suppose that a PI with NSF funding refuses to share

strains or data. What then"? This researcher's perception of researcher data being shared "immediately" was that data was likely shared upon publication, rather than sharing primary data.

One researcher expressed concern that, in the Geosciences, the 2011 introduction of NSF's EarthCube initiative will introduce complexity for data sharing and data management by imposing strict policies. EarthCube is a long term joint venture between the NSF Directorate for GeoSciences and the Division of Advanced Cyberinfrastructure and is a virtual community of 2,500 contributors that supports open science in Geosciences research towards a holistic understanding of the Earth system (NSF, 2016d). From the researcher's perspective, in order to be effective, NSF should remain very flexible about data management.

Another researcher explained that Atmospheric science is very "data heavy" and unlike other disciplines, data archives go back a hundred years with "a billion bits of weather data from satellites and surface based platforms". While the researcher mentioned that he depends on current data sets from many field experiments, an important consideration for his field is that data has a short lifetime and requires long term data archive planning for high volume data sets to be available for decades and longer for trend and modeling analysis.

It is interesting to note that interview discussions by researchers from the Geosciences and Atmospheric Science disciplines seemed to align with findings from the 2008 RIN study, in terms of their data sharing culture and willingness to publish datasets, as previously discussed in Chapter 2.

In terms of aiding data management, participants mentioned that NSF sponsors key programs to assist the research community. For example, one valuable resource for large project research is the National Center for Atmospheric Research (NCAR) Federally Funded Research and Development Center, which is funded by NSF and plays a key role to assist researchers in coordinating data management in the field of atmospheric sciences (NSF, 2016c). Once researcher acknowledged that NCAR has provided a great service to aid scientists involved in large project research, although limited in scope to Atmospheric and Earth Science communities. In terms of human subjects research, some participants discussed how data sharing introduces significant security concerns and questioned how NSF plans to handle these growing concerns. Potential options were to continue sharing best practices with the research community and make special allowances, when possible, in light of confidentiality concerns.

In reviewing and summarizing survey comments, there were insights into several areas where researchers acknowledged ways the current NSF proposal process is effective. Some researchers suggested NSF protocols and requirements for certain types of data have significantly enhanced data sharing practices, for example shipboard data for oceanographic cruises and biological sequence data. Typically, this type of data would not readily be shared today without current NSF requirements and mandates in place.

Survey Findings

Based on the data analysis of Phase One, survey findings include suggestions about the process from the researchers' perspective. Responses also provided insight into potential relationships that may exist between PIs, data sharing practices, and

factors that may impact the timeliness in sharing data. Figure 8 illustrates an overview summary of Phase One findings.

❖ PHASE ONE - SURVEY FINDINGS ❖

Recommendations on the NSF proposal process

1. **Review and streamline the current NSF proposal application and process:**
    - Review the necessity and scope of the Broader Impact section of the proposal and streamline, if possible.
2. **Consider and implement ways to ensure confidentiality and the protection of human subjects:**
    - Consider innovative ways that sensitive stored data may be protected, especially as it relates to compromising human subjects (IRB);
    - NSF could share best practices among researchers with a focus on confidentiality concerns.

Recommendations towards improving data sharing practices in research

1. **Provide comprehensive training programs to educate academic researchers on data sharing techniques, best practices, and ways to address data confidentiality concerns:**
    - Provide professional development for researchers on data sharing best practices;
    - Develop and provide access to educational resources to aid researchers on data sharing techniques; and
    - Develop and publish a fundamental guide for data sharing protocols.

2. Review and implement modern, updated policies and practices to foster data sharing in academic research communities:

- Require submission of primary data directly to NSF from funded researchers and require public posting of that data after a specified embargo period on a website managed directly by NSF;
- NSF could consider remaining flexible about data management and avoid strictly mandated data sharing policies;
- Implement planning strategies for the long-term to accommodate the storage and retrieval of data sets. Data sets from research experiments tend to have a short lifetime that requires planning for archives to be available for decades;
- Consider that primary data sharing continues to be made through the peer review literature; and
- NSF could consider and define more explicit data sharing requirements (e.g. software must be released under an open source license at a permanent, accessible website).

3. Monitor funded projects to ensure DMP is being followed with a verification process:

- Increase accountability to ensure data sharing in keeping with the DMP. NSF could monitor funded projects with a verification process after year one and *before* year two funds are released; and
- Require demonstration of data sharing per DMP of awards for new proposals.

4. **Implement public web-based technologies to document researcher experiences in using data repositories and other data sharing sites:**
    - Consider developing and maintaining a portal for researchers to report attempts to obtain data from other researchers and record their data sharing experiences, whether success or failure; and
    - Develop comprehensive web pages to display data available. For example, the National Center for Atmospheric Research (funded at least, in part, by NSF) coordinates data management as a service to researchers involved in large projects.
5. **Implement innovative funding incentives to facilitate data sharing:**
    - Grant funding for data sharing programs and professional development for PIs on the topic of data sharing;
    - Establish clear, mandated prerequisites on data sharing before additional funding is granted;
    - Greater recognition of the funding needs is required to fulfill appropriate data sharing needs. Data sharing costs for time and funding are high and the proportion of funding to support them are currently insufficient; and
    - Grant financial support beyond overhead funds for storage and backup to cover the cost related to large data sets.

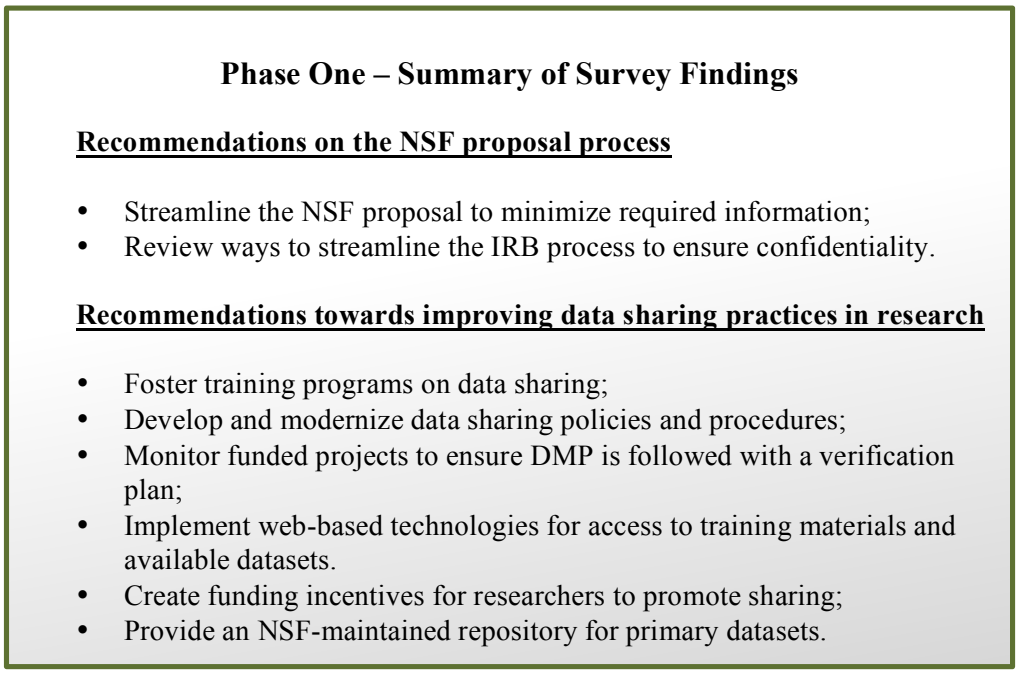

Figure 8. Summary of Phase One survey findings.

Phase Two Interview Methodology

In keeping with the study methodology outlined in Chapter 3, upon completion of the Phase One surveys, the aim of Phase Two was to prepare for, conduct, and document the findings of 25 one-on-one interviews with PIs and Co-PIs from five select Florida universities. The in-depth interviews allowed this researcher to gain an understanding of PIs personal experiences with the NSF grant proposal process and also to clarify the extent of the researcher's role in processing proposals for federal grants.

Each interview participant was recruited by reviewing PI and Co-PI contact information from the NSF awards archive database that is accessible via the NSF website. The selection criteria for interview candidates were PIs and Co-PIs who had submitted an NSF proposal within the previous two years that were funded by NSF and

and at that time were employed by one of the five institutions selected for this study. As previously mentioned, the five universities targeted for Phase Two were academic institutions with the largest number of NSF grantees for the year 2011, including the University of Florida, University of South Florida, Florida International University, Florida State University, and the University of Central Florida.

Before interview sessions were scheduled, this researcher's academic institution IRB requested two additional criteria be met in advance of Phase Two:

1. The receipt of written approval from each of the five institutions for interviews to be conducted with their participating researchers;
2. IRB approval from each institution to physically conduct interviews on-site at their institution.

To satisfy the first IRB requirement, this researcher contacted each institution and received written approvals to conduct interviews for this study. However, to satisfy the second IRB requirement, it was decided that telephone and Skype sessions would suffice, rather than conducting on-site sessions in an effort to avoid delays in obtaining IRB approvals from each institution.

Phase Two consisted of semi-structured interviews to gather detailed information from researchers who complete and submit NSF federal grant proposals. The goal was to also acquire responses from five NSF representatives familiar with grant processing and approvals in order to gain an understanding of the process from the NSF perspective. Contact information for individuals from the DGA was retrieved from the NSF website. Although email requests for an interview were sent to the appropriate DGA Branch

Manager at NSF, there was no response to these emails and therefore NSF's viewpoint is not included in this study.

Once email address contact information for the selected PIs and Co-PIs was identified, participation letters were distributed to the researchers via email with a brief overview of the study and the Adult Letter of Consent included as an attachment (See Appendix F). The researcher randomly selected contact information of potential interview participants in batches of approximately 20 email addresses per batch. Acceptance responses from PIs to participate in this study was relatively consistent, with one-to-two acceptances per batch of delivered emails. Several researchers responded that they were interested in participating in this study, but lacked the time due to busy work schedules.

A total of 297 email requests were distributed over a three-month period until 25 researchers agreed to participate. 297 was not the entire population of PIs that met the selection criteria, but rather the number of email requests distributed in order to receive 25 acceptances. As researchers responded and acknowledged their willingness to participate via email, an appointment was arranged at a mutually convenient time. Each semi-structured interview was approximately 30 minutes in length and was conducted via Skype or telephone. In keeping with study's stated methodology, sessions were audio-recorded for subsequent transcription and data analysis.

In preparation for the interview sessions, a detailed interview guide was developed to guide the discussions (See Appendix G). The interview guide was a framework, however, and each interview offered opportunities beyond the guide for open discussion. The interview process included the following dialogue:

1. Introductions and a brief overview of the study;
2. Explanation of the importance of the research and the study objective;
3. Explanation of the following study procedures:
    - Consent process - Reviewed the risks and benefits associated with the study, where loss of privacy and loss of confidentiality are the two primary associated risks.
    - Explanation that participation in the study was completely voluntary and the participant had the right to withdraw or refuse to participate in the study at any time.
    - Explanation that the session would be audio taped for reference purposes during the data analysis phase of the study. Also, reassured the participant of their anonymity.
    - The participant was asked if recording the interview was a concern. If not, this researcher began to record the session, readied the interview guide, and took notes throughout the interview.

An important component of the interview process was to develop and maintain the Microsoft Excel spreadsheet matrix to document interview questions, participant responses, and notes. By using the interview recordings and interview notes as reference resources, the final step after each interview concluded was to play back the recording, then analyze, code, and document researcher responses onto the matrix spreadsheet.

Interview Data Analysis

By following grounded theory methodology, the first interview was used as a framework for subsequent interviews and allowed this researcher an opportunity to

identify common elements and viewpoints. The second interview was coded based on elements from the first interview. Likewise, subsequent interviews began to form and categorize the emerging recommendations and theory.

Data analysis and review of the interview discussions provided interesting viewpoints and perceptions about data sharing from a researcher's perspective, including thoughts on the importance of data sharing to researchers, the positive and sometimes limiting aspects of NSF efforts, thoughts on NSF's current Fast Lane proposal submittal system, perceptions of the NSF proposal process as a whole, potential relationships between the researcher and the process, and recommendations for future improvements to the proposal process aimed at enhancing data sharing.

All Phase Two researchers were active PIs and some also participated in scientific projects as Co-PIs. The researchers held a wide range of scientific research experience ranging from 5-30 years and represented diverse areas of science, including molecular genetics, statistics and research, tropical cyclones and hurricanes, astronomy, physics, neuroscience, theoretical physics (quantum field theory), material science and research, environmental engineering science, engineering, oceanography, natural user interfaces and computer science, and chemical biology.

In keeping with the format of the interview guide, the first step in the interview process was to gain an understanding of each researcher's background by asking several questions related to their experience. What was the researcher's primary area of interest? Was the researcher in their field for many years or were they new to their current field of research? Were they active participants in the NSF proposal preparation and submittal process? How many proposals had they submitted in the past two years? What was their

view of the NSF's DMP to encourage data sharing in academic research? Is data sharing an important aspect of their research or did they perceive it to be important at their institution? What recommendations could they offer to improve data sharing efforts, the NSF proposal, or the process?

Figure 9 illustrates that out of 25 interview participants, 23 PIs and Co-PIs received funding for 1-3 NSF proposals between 2012 and 2014. There were no Phase Two interview participants awarded more than 3 NSF proposals during that timeframe.

Figure 9. Number of NSF proposals funded per researcher from 2012 to 2014.

A few researchers acknowledged that data sharing is not an important aspect of their research at all, most notably from the disciplines of mathematics, physics, and the material sciences. These researchers confirmed they seldom or never shared data with other researchers. One of these researchers explained that typically his data is not shared prior to publication because true value is achieved by sharing research ***results*** via publications and conferences with other researchers, rather than sharing physical datasets. The researcher explained that open forums and conferences, such as Gordon Research Conferences, are valuable venues for sharing research results with fellow scientists.

Even so, 19 of the 25 researchers interviewed within the scope of this study actively share data with other researchers and acknowledged that data sharing is important today and will become increasingly important in the future. One researcher acknowledged that young researchers tend to be "pro data-sharing" and foster "pro-transparent science" and mentioned that student research groups currently share data as the norm.

As a side note, researchers were asked about their experience with the NSF grant proposal submittal process. The majority of researchers acknowledged they use FastLane, which they indicated has proven to be a highly efficient and effective platform for submitting proposals. There were several positive comments on FastLane and its efficiency, which included "very user friendly," "very easy to use," and "easy to collaborate with other researchers." Figure 10 illustrates a process flow diagram of the NSF grant award process and timeline. FastLane and Grants.Gov are the two primary platforms researchers use to submit proposals to NSF.

PIs and Data Sharing

As the interview sessions progressed, it became increasingly clear that one of the more controversial aspects of the NSF proposal was the perceived effectiveness of the DMP to encourage data sharing. For some researchers, the DMP was viewed as an effective way to "organize thoughts," "is a commitment within the proposal process," and was also viewed as a step they would have completed "even if it weren't mandatory." For others, the DMP has proven to be a frustrating and often "worthless" exercise that created time delays in the process and did little to promote data sharing.

Some researchers mentioned there is a great deal of confusion on how much time should be given to the plan, what should be included in the plan, and how the plan should be completed. Some of the comments gathered from interview sessions about the DMP indicated, “it significantly slows down the process” and “holds little value because the plan is open to interpretation.”

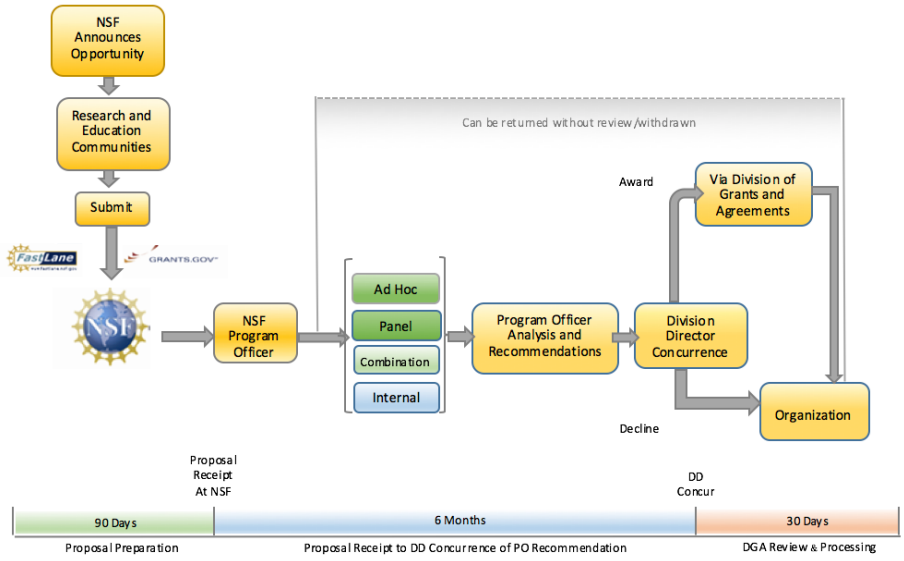

Figure 10. NSF proposal and award process timeline (NSF, 2016a, Exhibit III-1).

In an effort to offer clarity to fellow researchers in completing the DMP, one oceanographic scientist acknowledged the positive efforts of the Biological and Chemical Oceanography Data Management Office (BCO-DMO) to provide template guidelines via their official website. In addition, the BCO-DMO website highlights documentation on DMP best practices prepared by their staff members to assist fellow researchers. BCO-DMO also offers an annual seminar for researchers with a tutorial on how to complete and submit datasets to NSF.

In terms of factors that may restrict or delay the NSF proposal process, one researcher shared that the introduction of Broader Impact criterion has had a significantly negative impact on the grant proposal process from their perspective. The researcher explained that Broader Impacts was originally introduced by the Clinton administration in the 1990s in an effort to provide for increased federal regulation in research grant funding. In the researcher's opinion, Broader Impacts has had an adverse effect on the funding and proposal processes, whereby funding that used to be judged on science is oftentimes judged on projects that attract underrepresented groups for funding. From the researcher's perspective, Broader Impacts has introduced undue bureaucratic rules and regulations that create significant delays in the NSF proposal process.

The researcher continued that the Broader Impacts section of the proposal was further complicated by the 2008 Democratic administration, who believed there was not enough regulation in federal funding. Likewise, the researcher argued that grant paperwork has increased significantly since that time, especially noted in the last grant submission cycle in 2014. As an example, a relatively minor oversight on a proposal, such as an unchecked checkbox, was resolved in the past via email from a Program Manager to NSF. In light of increasing regulations, however, minor oversights must be corrected by the researcher and the proposal completely resubmitted to NSF, resulting in significant processing time delays.

Another source of proposal process delays identified by researchers were grants and budgetary staff resource limitations at academic institutions. Time delays can sometimes exist at the academic institution before the NSF proposal is submitted. For example, institutions may not have a grants specialist on staff to aid with either the

process or budget preparation. Some researchers acknowledge that “red tape” encountered at the institution can often be frustrating and time consuming. One researcher explained that paperwork must flow through their institution’s Division of Research that operates with long lead times. In addition, their Research Division staff may not have the expertise to understand fully the specific grant being proposed, creating undue complexity and delays.

Further, increasing IRB restrictions when dealing with human subjects is another aspect of the process that can sometimes create delays in grant processing. While IRB protection for human subjects is necessary, 3 of the 25 researchers voiced their opinion that regulations have become increasingly complicated and time-consuming. When asked for an opinion on how the NSF could address this concern, many were unsure what NSF can do to relieve time delays attributed to this sensitive issue. While acknowledging its importance, some researchers are increasingly frustrated by the lack of timeliness of the IRB process and question the need for increasing IRB requirements in research.

Another factor that may hinder sharing are security concerns about data repositories and their stability. For example, how long will the repository exist? Who is ensuring that repository security standards are adequate and maintained? What happens to the data if the repository is discontinued? In keeping with these concerns, a common theme amongst researchers interviewed was the need for a secure, accessible data archive housed by the NSF, where researchers could submit small and large data sets for easy access by fellow PIs and other researchers. Such a central data repository would be beneficial to reduce delays, improve security, and ensure stability when accessing data

housed in a remote repository. In addition, a portal was suggested for ready access to data.

Fear was another aspect that may hinder researchers from timely sharing of data. One researcher suggested the use of datasets by other researchers could be used against the original researcher's findings to refute initial study conclusions. Another fear expressed was that preliminary shared data could be misinterpreted by others, leading to confusion and unreliable research findings. Figure 11 summarizes key factors that researchers acknowledged may introduce delays in the sharing of data.

Interview Findings

Interview sessions provided valuable insight into how data sharing may be improved from a researcher's perspective. Establishing comprehensive data requirements may be an important first-step to establish new data sharing strategies. According to one researcher, the field of Neuroimaging should play a leading role in the development of modern data sharing strategies.

From the researcher's perspective, Neuroimaging should be considered by the NSF and other federal funding agencies as the foremost field because, "there is so much you can do with the data" and "the data is so expensive as compared to datasets from other disciplines." Therefore, in terms of return on investment, their data would provide the highest return due to its high cost.

Accordingly, Eickoff, Nichols, Van Horn and Turner published a comprehensive paper in 2016 that discussed the leading role Neuroimaging should play in modern data

sharing strategies. The authors suggested that neuroimaging "resonates well with the increasingly recognized need for transparent and reproducible science" (p. 1065). The authors suggested that monetary expenses associated with MRI scanning, along with increasing volume of study sample sizes that requires little effort to recruit new subjects,

IRB	Increasing IRB restrictions for human subject research creates significant delays in social science research.
Policies: Data, Metadata, and Data Standards	Comprehensive metadata standards are essential.
	Data file formats should be standardized.
	More formal guidelines on ways researchers should share data is needed.
	Shared data needs to be de-identified. Every discipline has a different view of data, therefore how should it be standardized? There is significant risk if data is uploaded and not de-identified.
	Misinterpretation of shared data is a concern.
	Working groups are essential to develop data sharing policies for a given discipline.
Security	Security is a concern if data is stored on a centralized computer. Strong need for a centralized permanent database for large and small datasets, preferably housed and maintained by NSF.
Proposal process	At times, multiple proposals are required to gain approval from different NSF reviewers.
Funding	Some suggest, for maximum return on investment, NSF should consider the field of Neuroimaging to lead the way in data sharing initiatives because their data is expensive.
	Sometimes researchers reduce the scope of their proposals to meet the allowed funding.
	Consider expanding funding for opportunities outside a research's specific field of study.
	NSF requests that researchers wait 1 year to resubmit proposals, creating a funding gap and delays.
Impact of new rules and regulations e.g. Broader Impacts, DMP	Some suggest the DMP significantly slows down the process and has done little to promote data sharing. The DMP is a data management tool, rather than data sharing.
	Federal regulations introduced by the Clinton Administration in the form of Broader Impacts mandates significantly more paperwork - especially noted in the last proposal submission cycle in 2014.
	Some researchers invent make-work projects within proposals to attract under-represented groups for funding.
	Before new legislation was implemented in 2008, minor errors on the NSF proposal were resolved via email. These same errors now require resubmission of the proposal that creates delays.
	Some suggest there should be consequences when data is not shared as indicated in the DMP.
Institution staffing	At times, the institution may not have a grants specialist to process proposals, creating frustrations, red-tape, and delays.

Figure 11. Overview of factors that impact relationships between PIs and data sharing.

provides robust information and, in part, makes data sharing and re-use "an inevitable prerequisite for further advancement in the field of neuroimaging" (p. 1065). The use of neuroimaging data for data sharing is economically prudent at a time where universities and funding institutions have limited financial resources (Choudery, Fishman, McGowan, & Juengst, 2014).

During interview sessions, some researchers mentioned framework standards are required to help ensure data accuracy, maintain data context, and minimize risk as follows:

- Ensure for high quality metadata;
- Establish comprehensive metadata standards;
- Develop guidelines to ensure data de-identification; and
- Identify standardization methodologies for file formats.

Literature supports the needs for quality metadata and standards, arguing that appropriate metadata to support data sets "is fundamental for effective data discovery and retrieval" (MacMillan, 2014, p. 543). Even so, Tenopir et al., 2011 explained that in a survey of 1329 researchers, 78% of researchers used no metadata and relied on their research laboratory standards. The lack of standardized metadata is a significant challenge to successfully storing data. Consistent metadata standards are essential for fellow researchers to understand the context of data for its effective use by others. Data standards are essential to facilitate accuracy in data sharing efforts. Data file formats between diverse disciplines of science should be standardized and well-documented.

Another researcher explained that every discipline has a unique view of "data" and outstanding questions exist on how to control associated risks in sharing data

between different disciplines. A key question to consider in shaping future data sharing strategies is how will the data be standardized between disciplines? If data is uploaded for sharing and not de-identified, there is significant risk in sharing the data. Therefore, comprehensive metadata standards are essential and data file formats must be de-identified beforehand.

Likewise, MacMillan (2014) acknowledged the need for a comprehensive data standardization strategy between disciplines, suggesting "the multiplicity of disciplines and data types complicates dissemination and discovery" (p. 543). With each discipline, there are inconsistencies on what makes the data discoverable, with unique patterns of data discovery and its own unique language (MacMillan).

Another creative way to encourage sharing that was discussed during one interview session would be to expand the current outcomes report process that researchers prepare as an NSF requirement for grant submission. Currently, researchers of funded projects provide NSF a project outcomes report that is published online. To encourage data sharing, NSF could consider a web-based enhancement to the outcome report process, whereby the outcomes report would link directly to the researcher's corresponding dataset(s) for use by fellow researchers.

NSF could also consider promoting data sharing efforts by developing and maintaining a secure, accessible data archive whereby researchers could submit small and large data sets for easy access by fellow PIs and other researchers. A central data repository housed by NSF would be beneficial and reduce delays when accessing remote data. While limited to the Earth Sciences, NSF funded in 2009 the Data Observation Network for Earth (DataONE) initiative, which is a virtual data network for

collaboration and provides a secure repository environment. According to Allard (2012), DataONE preserves and protect research objects in an open format approach responsive to researcher and user needs.

Another area of opportunity may be for NSF to become more involved in the training and educating researchers and institution data administrators on data sharing practices and methodologies. NSF could establish, maintain, and promulgate formal guidelines on the ways researchers could and should share data. NSF could also consider developing a dedicated work group to review and refine data sharing policies for ongoing enhancements, updates, and communications.

Regarding NSF's efforts to monitor the progress of projects post-funding to ensure funded projects are on track, Langfeldt and Scordato (2015) suggested federal funding agencies tend to monitor research projects in various ways and there is little discussion in the literature on definitive methodologies for monitoring of funded research. Typically, however, funding agencies conduct periodic conferences and knowledge exchange with stakeholders. The authors explained that NSF's approach to monitoring includes annual reports from grantees with "no general demands – but encouragement – for interaction with users" (p. 20). Langfeldt and Scordato argued NSF does not conduct mid-term monitoring/assessments of individual projects. Even so, continuing or additional grants may be delayed or terminated if annual reporting is not submitted by researchers to NSF for review.

Phase Two Summary

Based on the data analysis and summarization of Phase Two interview sessions, several recommendations emerged to improve data sharing in academic research from the researchers' perspectives. The following recommendations are based on the findings and analysis of the 25 semi-structured interview sessions, further summarized in Figure 12.

❖ **PHASE TWO - INTERVIEW FINDINGS** ❖

Recommendations on the NSF proposal process

1. Review of the current NSF proposal format to identify complexities driven by changes that may hinder the timeliness of the proposal and funding process:

- Some researchers suggest the introduction of Broader Impacts criteria in the NSF proposal has had an adverse impact on the proposal by introducing unwarranted, cumbersome rules and regulations that creates undue delays in the federal grant proposal process; and
- Researchers' minor omissions on the NSF proposal (e.g. overlooking a checkbox) can introduce significant delays in the grant approval process and affects progress of research. In the past, program managers had the authority to request minor corrections to the proposal by emailing NSF directly. Recently revised federal rules and regulations mandates researchers must resubmit corrected proposals for re-review, which significantly delays or causes proposals to be denied.

2. Improved transparency in the NSF grant submission process:

- Transparency and accountability to educate less experienced researchers on completing the proposal may relieve processing delays.

3. The approvals process at the academic institution level can be time-consuming and frustrating to researchers:

- At times, delays exist in the proposal process at the academic institution due to frustrating "red-tape" requirements in approving proposals for submission to NSF; and
- The institution's research or grants division may not have adequate expertise or knowledge to understand thoroughly the specific grant, which may delay the submission process.

4. Consider proposals from researchers outside the boundaries of their particular branch of study:

- Some researchers suggest it is difficult to receive funding for grants considered outside their given area of study. Broadening the scope of research opportunities would allow for diversity in research opportunities.

5. Review the need for increasing restrictions posed by Institutional Review Boards in working with human subjects within social science research:

- Some researchers suggest that Institutional Review Boards create unwarranted, significant delays in social science research when working with human subjects by imposing increasing requirements and restrictions.

Recommendations towards improving data sharing practices in research

1. **Expand current NSF forward-facing web technologies to share datasets:**
    - Researchers of funded projects provide NSF a project outcomes report that is currently published online. NSF could consider a web-based enhancement to this process, whereby the mandatory Outcomes Report would link directly to researcher datasets maintained by NSF.
2. **Develop and maintain a permanent, secure data archive for data sets:**
    - Consider promoting data sharing efforts by developing and maintaining a secure, accessible data archive whereby researchers could submit small and large data sets for easy access by fellow PIs and other researchers.
3. **Ensure that data standards are comprehensive and well documented to encourage accurate and effective data sharing in modern research. NSF could consider the field of Neuroimaging to lead modern data sharing efforts.**
    - Consistent metadata standards are essential for fellow researchers to understand the context of data for subsequent sharing with others. Data standards are essential to ensure accuracy in data sharing efforts;
    - Data file formats between diverse disciplines of science should be standardized and the need is well documented in literature;
    - Some researchers suggest data needs to be de-identified to ensure effective data sharing; and
    - NSF could consider the field of Neuroimaging as a leading discipline to assist in developing updated data standards and practices since their data is so expensive.

4. NSF could consider an active role to train and educate researchers on data sharing practices and methodologies:

- Formal guidelines about ways researchers should share data would be a welcome addition to the NSF proposal process. Every research community has different entities they refer to as "data".

5. NSF could consider developing a dedicated, collaborative work group to develop and review data sharing policies for ongoing enhancements, updates, and communications.

Phase Two – Summary of Interview Findings

Recommendations on the NSF proposal process

- Monitor the impact of new policies and regulations;
- Accountability and transparency with the proposal process;
- Streamline the NSF proposal to include essential information;
- Accept proposals outside a researcher's primary discipline;
- Monitor project progress post-funding; and
- Review increasing IRB restrictions when working with human subjects.

Recommendations towards improving data sharing practices in research

- Expand the use of web-based technologies;
- Implement an NSF-maintained data repository;
- Ensure for comprehensive data and metadata standards; consider the field of Neuroimaging as a leading discipline in data standard/sharing strategies;
- Foster comprehensive training initiatives to educate researchers on DMPs and best data sharing practices; and
- Establish a dedicated workgroup to develop and review data sharing policies and procedures.

Figure 12. Summary of Phase Two interview findings.

Triangulation of Data

Once the individual analysis and summary of Phase One and Phase Two data was complete, the final task in the data analysis phase was to triangulate, or converge the findings between the two phases. According to Schwandt (2015), triangulation "is a procedure used to establish the fact that the criterion of validity has been met" (p. 307). Schwandt suggested the strategy of triangulation "is often wedded to the assumption that data from different sources or methods must necessarily converge on or be aggregated to reveal the truth" (p. 307).

Other research by Turner et al. (2015) suggested that triangulation is a fundamental social sciences strategy to produce a more robust understanding of a subject or theory. Early social science researchers identified triangulation as a method to help reduce imperfections inherent to all research methodologies, reduce bias, and validate research study findings, rather than simply seeking a common theme (Turner et al.).

In her widely-cited work on triangulation strategy, Sandra Mathison (1988) described the value of triangulation as providing evidence of convergence, inconsistency, or contradiction between research findings in order to make sense of diverse study findings. Mathison maintained that once the data analysis phase of a study is complete, it is the role of the researcher to provide a holistic assessment to make sense of diverse research findings using multiple methods, resources, and personal knowledge.

Triangulation Methodology

In keeping with Mathison's triangulation strategy framework, this researcher adopted a comprehensive triangulation strategy as the best approach to converge Phase One and Phase Two data (see Figure 13). Applying Mathison's framework to this study

provides a multi-faceted perspective that incorporates 4 key areas for a holistic research assessment:

- Method triangulation (qualitative survey and interview between methodologies);
- Researcher triangulation (multi-disciplinary and experience/knowledge base);
- Data triangulation (literature review analysis, survey and interview data); and
- Environment triangulation (location, diversity of institution cultures).

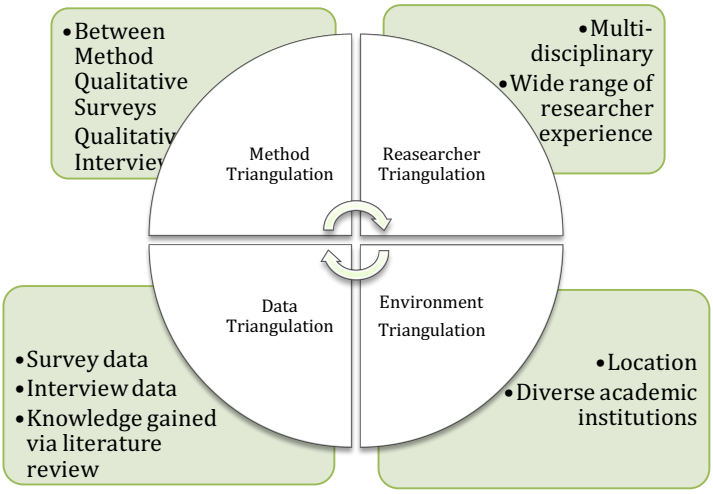

Figure 13. Triangulation strategy.

In 2002, Patton argued that analyst triangulation is a methodology used to analyze qualitative data and compare research findings by two or more data analysts. This researcher incorporated this methodology to help ensure the validity and trustworthiness of study findings. To ensure research validity, this researcher requested the assistance of two additional people (data analysts) to review coding and analyze the data to support or highlight differences in original theme assessment and no discrepancies were noted.

Triangulation Summary - Converged Findings

As an outcome of the data triangulation between Phase One and Phase Two, several common findings emerged to reduce delays and encourage data sharing within the provision of NSF's current grant proposal process. The following findings were common suggestions from Phase One and Phase Two:

❖ **DATA TRIANGULATION – CONVERGED FINDINGS** ❖

PHASES ONE AND TWO

Recommendations on the NSF proposal process

- Review implications of increasing IRB protocols that may delay the proposal process;
- Periodic review to determine the impact of new policies and regulations; and
- Monitoring of projects post-funding to ensure initiatives are on track and project team is intact.

Recommendations towards improving data sharing practices in research

- The need for a stable, accessible, NSF-maintained repository to house datasets;
- Funding incentives to promote sharing and offset additional expenses for sharing datasets;
- Consequences for those who submit DMPs but do not share datasets;
- The need for researcher training programs and workshops on DMPs to educate about data sharing strategies and best practices;
- Metrics to assess the effectiveness of NSF's DMP to facilitate data sharing; and

- Enhancements for modern data security, sharing, and access policies and procedures.

The next step in the process was for this researcher to identify contradictions or inconsistencies in the data responses gathered between the two spreadsheet matrices. After careful review, it was determined that 100% of the framework Phase One survey suggestions were also included in Phase Two findings. In terms of inconsistencies, three additional suggestions were noted on the interview matrix that were not included in the survey matrix, as follows:

- The ability of NSF to accept proposals outside a researcher's primary area of study;
- Adequate data and metadata standards and policies are essential to support reliable, accurate data sharing; and
- The field of Neuroimaging should be considered to lead the way in future data sharing strategies.

Chapter 5

Conclusions, Implications, Recommendations, and Summary

Conclusions

As advancements in digital technologies increase, opportunities exist to ensure data integrity through openness and transparency (NSF, 2016d). The growing importance of data policies that guide researchers and the methodologies set forth by federal funding agencies play a significant role in modern research (Pampel et al., 2013). In keeping with recent discussions on the use of open access environments that foster knowledge and data sharing, well-respected organizations are developing strategies to ensure that effective policies and procedures exist to guide data management and dissemination practices in research communities.

This study utilized a multi-method approach to reveal insight into potential relationships between PIs and timely data sharing practices in academic research within the guidelines of the NSF federal grant application process. Also presented in this study were suggestions for potential enhancements to the current NSF proposal process to encourage data sharing. The NSF proposal, its process, and the practices of researchers who participated as active NSF grantees guided this research.

In 2011, NSF revised its grant application policy to mandate that supplementary documentation be included with grant applications in an attachment known as the DMP (NSF, 2011). Recent modifications to NSF policy require grant applicants to share data and supporting materials gathered within the scope of NSF funded research.

While there were suggestions for improvements and enhancements in certain areas of NSF's proposal process, one of the most controversial aspects of this study was

the value of the DMP to promote the sharing of data. Borgman (2012) suggested that while DMP requirements encourage data sharing, instead the DMP is a plan designed for data management. Likewise, several researchers interviewed in this study acknowledged the DMP's value to help organize thoughts and formalize the grant proposal process, rather than promote the sharing of data.

As a potential solution, the NSF could consider additional metrics to gauge the effectiveness of the DMP to promote data sharing from the research community. Bishoff and Johnston (2015) discussed a study conducted by the University of Minnesota Libraries to analyze DMPs from their institution's funded NSF grant proposals submitted January 2011 to June 2014. The study included 7 colleges and academic units from the University of Minnesota. Study findings suggested that while useful to understand researcher philosophies on sharing data, more intervention is required to ensure that plans to share as defined in DMPs are carried out fully. Also noted in the study was the need for researchers to gain a more complete understanding about the types of data sharing that facilitates public access.

Likewise, the University of Illinois at Urbana-Champaign conducted a similar study of 1,260 DMPs submitted with their NSF proposals July 2011 through November 2013 (Mischo, Schlembach, & O'Donnell, 2014). A primary goal of the study was to analyze what researchers were proposing in the DMP and also to examine storage solutions researchers adopt to meet DMP requirements. Study findings revealed the data dissemination and sharing methodologies suggested most often by researchers were workshops, conferences, and publications. The authors suggested this may due in part to vague DMP guidelines, but also "NSF's focus on the sharing of processed data – as

opposed to raw data – and the PI's natural tendency is to associate processed data with publications" (p. 42).

As a side note, Antell, Foote, Turner, and Shults (2014) suggested that science librarians may play an increasingly important role to assist researchers with data management needs. While this would be a new and challenging role, the authors explained that science librarians are well equipped to assist researchers in the preparation of DMPs because of their traditional skills to apply metadata standards, organize information, and provide ready access to information.

In keeping with perceived limitations of the DMP, researchers in this study argued that NSF currently imposes no consequences to researchers if data is not shared. Therefore, while the DMP is a mandatory component of the proposal, there is no course of action for those who fail to comply to share datasets. Researchers suggested that NSF should consider implementing restrictions or consequences for those researchers who do not share data. Another suggestion was that NSF mandate primary data be submitted to NSF for all federally funded projects. This data could be publicly posted after an established embargo period on a site directly maintained by NSF.

Pryor authored a book in 2012 that discussed the complexities of managing research data and the lack of consequences for not sharing in the scope of DMPs. According to Pryor, "there exist neither carrots nor sticks to ensure that, once funds are released, there will be any rigorous adherence to the agreed [data management] plan" (p. 5). While Pryor acknowledged NSF's efforts to include the supplementary, two page DMP with proposals effective January 18, 2011, he argued that "it is a long way from ensuring that the data produced will be properly prepared, managed, and preserved for

long-term access and reuse. It certainly carries no sanctions to ensure compliance with a front-loaded statement of conformance" (p. 5).

Another recommendation in this study was for NSF to maintain a stable, accessible, secure environment for researchers to store their datasets. When discussing their views on data sharing, researchers expressed concerns about data security, repository stability, and concerns with additional expenses they must bear in order to share data. Some researchers suggested NSF may relieve some of these concerns with a stable data repository environment. In addition, an online portal maintained by NSF was also suggested as a welcome enhancement for the research community to provide a holistic platform for communications about available datasets, information on training opportunities, and overall collaboration.

Likewise, NSF's Public Access Plan released in 2015 highlighted several short and long term initiatives planned to enhance current operations related to public access of federally funded research, including a plan to offer guidance with DMPs and a repository to store research articles (NSF, 2015):

- The implementation of a system to enable NSF-funded researchers' articles (manuscript or version) to be made available to the public via the NSF Public Access Repository (PAR); This system will be voluntary for researchers in 2015 and mandatory for NSF proposals submitted from the January 2016 effective date;
- Enhancements that may offer guidance with DMPs;
- Communications with the research community and stakeholders to identify gaps where guidance is needed; and

- System flexibility for expansion and growth to allow for changing technologies.

Adequate training was also mentioned as an important aspect that may encourage sharing. Although the DMP became a mandatory requirement 5 years ago, some researchers in this study maintained they do not know how to fill out the DMP, nor information that should be included in the plan. Further, several researchers expressed interest in learning more about data sharing best practices and strategies. Based on interview sessions in this study, researchers would welcome additional NSF-sponsored training programs and forums. They expressed a need for enhanced, comprehensive training sessions on the DMP and data sharing, as well as programs or work groups that promote collaboration with fellow researchers.

In keeping with the belief that additional training programs are needed to educate researchers on the DMP, Akers and Doty (2013) discussed a research study conducted in 2012 by the Emory University Libraries. The study was conducted with 330 of their faculty members who were also active researchers. Analysis of survey results indicated that 82% of faculty researchers were somewhat or not at all familiar with DMP requirements of the federal funding agencies (NSF, NIH and National Endowment for the Humanities). Even worse, study findings revealed that Humanities and Arts researchers were completely unfamiliar with DMP requirements.

Some suggest IRB restrictions in social science research creates significant delays in the process and therefore, may be another opportunity to streamline the process. In 2015, Carl Schneider published a book that described the current IRB system as fundamentally misconceived. Schneider explained that the IRB process and growing number of regulations imposed by the system does little to protect human subjects.

When describing IRB's system, Schneider asserted "its design invites IRB imperialism, and it has invaded new areas (like social science research)" (p. xxx).

The majority of researchers that participated in this study recognized the growing importance of data sharing in modern research. While several recent advancements have been made to increase transparency and accountability in the form of new federal rules and regulations, some researchers suggest these changes have been ineffective. Further, new regulations have placed undue burdens on researchers creating unwarranted delays in their research.

Likewise, the increasing number of federal regulations was mentioned as a key area of concern in the scope of this study. In particular, some researchers argued the introduction of Broader Impacts in the NSF proposal introduced undue rules and regulations. Increasing federal regulatory requirements tends to restrict researcher efforts to focus on their research. Several researchers suggested the proposal could be streamlined to request minimal information to expedite the process. Facilitating the proposal process would, in turn, allow researchers more time to focus on their research. According to Watts, George, and Levey (2015), the Broader Impacts Criterion (BIC) has generated mixed reviews and reception by researchers. According to Watts et al., many researchers judge BIC as "a burden and a distraction from the science they were trained to conduct" (p. 397).

Lastly, a key opportunity discussed in this study was the significant contribution the field of neuroimaging could provide as a leader in modern data sharing strategies. In their comprehensive 2012 article, Poline et al. provided an interesting discussion on the benefits and social/technical challenges of sharing neuroimaging data. Recognizing

advancements and the increasing interest in sharing neuroimaging data within the past decade, the authors explained that electronic data capture (EDC) methods along with other new tools have provided a means to simplify and standardize components of sharing data, however barriers to sharing still remain.

A key challenge has been found with inaccurate or missing metadata related to the data, which in turn has resulted in reduced trust and compliance concerns. Poline et al. (2012) suggested that technology is not the main concern in data sharing practices, but rather the lack of standards, policy, and specifications to capture data during the neuroimaging lifecycle (e.g. what, how, when). Even so, as compared to other fields of neuroscience, the neuroimaging community has been acknowledged for their advancements in data sharing. The INCF Task Force on Neuroimaging Datasharing was formed in collaboration with partner groups worldwide to develop tools to allow researchers to minimize barriers in sharing primary, derived, and processed neuroimaging data.

Poline et al. (2012) suggested that it may be possible in the future for research to rely on knowledge management and technical systems rather than data management to aggregate information to assist with predictive modeling. The authors concluded that coordinated efforts on standards to generalize data sharing and associated metadata within the neuroimaging community is essential moving forward.

Accordingly, Eickoff et al. (2016) suggested that monetary expenses associated with MRI scanning, along with increasing volume of study sample sizes that requires little effort to recruit new subjects, provides robust information and, in part, makes data sharing and re-use “an inevitable prerequisite for further advancement in the field of

neuroimaging" (p. 1065). The use of neuroimaging data for sharing is financially a viable solution as universities and funding institutions have limited financial resources (Choudery et al., 2014).

Implications

Understanding the practices and policies that support data sharing is a growing area of research in the fields of information and social studies of science. Research has argued for empirical studies to explore data sharing from diverse aspects, including studies to identify specific elements that affect data sharing practices, comparative studies to understand sharing between diverse disciplines, and exploratory studies to gain insight about researchers who share and reuse data (Haeussler, 2011; Savage & Vickers, 2009; Tenopir et al., 2011).

There are several implications for future research to review data sharing practices in academic research. In this study, the scope of Phase One was limited to participants from North Carolina, South Carolina, Georgia, Alabama, and Florida. Phase Two interviews were limited to five major Florida academic research institutions. As an extension of this study, future research on federal grant proposals as related to data sharing could include a broader selection criteria of researchers located beyond the scope of North Carolina, South Carolina, Georgia, Alabama, and Florida.

In addition, future comparative research could be conducted to focus on the federal grant proposal process and its impact on data sharing between the various federal funding agencies, including NSF, NAS, DOE, and NIH. Yet another interesting comparative study would be to analyze researchers' data sharing attitudes and practices

between diverse disciplines of science. Scientific disciplines differ in their needs and require more specialized attention to achieve successful data sharing strategies.

Another extension of this research could focus on the DMP as compared to mandatory data management or sharing plans from other federal funding agencies. In this study, there was a wide range of responses from researchers on the DMP. Additional research might also be expanded on this study's interview selection criterion to include academic institutions located internationally for an historical or comparative study on data sharing strategies to highlight current and/or projected data sharing progress on a global scale.

A final suggestion for future research would be to conduct a study with a variation sampling of researchers selected by different disciplines, and include researchers from the field of social science due to its inherent complexity. Such research could serve as an extension of the extensive RIN study conducted in 2008 that presented attitudes of various disciplines of science to share and publish datasets. The RIN report was previously discussed within the literature review of Chapter 2.

Recommendations

According to Tenopir et al., 2011 "increasing the efficiency of current data practices in a world of increased data challenges requires a new comprehensive approach to data policy and practice" (p. 3). Several suggestions were presented in this study from researchers' perspectives to document relationships between researchers and data sharing practices and suggestions to enhance the current NSF grant proposal process, summarized as follows:

- The need for a stable, accessible, NSF-maintained repository to house datasets;

- Implications of increasing IRB protocols that may delay the proposal process;
- The need for researcher training programs and workshops on DMPs to educate about data sharing strategies and best practices;
- Metrics to assess the effectiveness of NSF's DMP to support data sharing;
- Enhancements for modern data security, sharing, and access policies and procedures;
- Funding incentives to promote sharing and offset additional expenses for sharing datasets;
- Periodic review to determine the impact of new policies and regulations;
- Consequences for those who submit DMPs but do not share datasets;
- NSF oversight to ensure adequate grants administration staffing at institutions; and
- Monitoring of projects post-funding to ensure initiatives are on track and project team is intact.

Summary

Recent literature acknowledges the importance effective data management strategies to facilitate collaboration between disciplines of research. Shared data allow researchers to build on fellow researchers' work to enrich and facilitate advancements in science (Roche et al., 2014). Those who support data sharing have acknowledged the important role of funding agencies in leveraging data sharing in scholarly research.

Within the framework of this research, Chapter 1 established the study objective by presenting the research problem and fundamental research questions to be examined. Recent literature confirms the lack of empirical studies to explore factors that impact

modern data sharing practices, even in light of its increasing importance (Tenopir et al., 2011). Chapter 1 highlighted the important role of federal funding agencies to implement working strategies for modern data management and dissemination, along with the goal of this research and its research questions, as follows:

To explore the NSF grant application process to determine the nature of its influence on the timely sharing of scholarly data. Written from the perspective of active PIs and Co-PIs who complete and submit NSF grant proposals, this study provided insight into potential relationships between researchers, the NSF proposal process, and the timely sharing of research data. This study addressed two specific research questions:

RQ1: Within the past two years, what specific relationships, if any, can be drawn between the Principal Investigator (PI) and the sharing of research data and materials, in keeping with current NSF federal grant application guidelines?

RQ2: In what ways can the NSF federal grant application process be improved to facilitate timely sharing of research data?

The literature review in Chapter 2 illustrated the evolving historical perspectives of data sharing practices and barriers in select literary works of well-respected researchers. The literature review highlighted many of the challenges faced by funding agencies and researchers alike to encourage and support the sharing of data. The diversity of cultures and disciplines, personal attitudes and practices, technology and infrastructure limitations, training, competiveness, fear, trust, security, privacy and ethical concerns, secrecy, policies and procedures, monetary gain, recognition, and data management limitations are all controversial elements that have contributed to this

important discussion. A review of regulations that support sharing was also presented, as well as data sharing frameworks in literature.

Chapter 3 provided a detailed discussion on the methodology adopted for this research and the reasoning behind its multi-method qualitative approach. Following the literary work of Maxwell (2013), qualitative research is indicated where little is known about a subject. A multi-method design promised a better understanding of the research topic and a more comprehensive achievement of this study's research goals.

Research outcomes were presented in Chapter 4 in the form of opportunities derived from interview sessions and anonymous survey data. Research participants were a random sample selection from a targeted population of active PIs and Co-PIs. Study findings were converged using Mathison's (1988) triangulation model, which is a holistic strategy to conduct method, researcher, data, and location triangulation. Common themes in research data emerged using a grounded theory approach. As illustrated in Chapters 4 and 5, several researcher recommendations proposed in this study are well-supported in related literature.

In conclusion, while factors that impact data sharing practices are complex and diverse, likewise there is a growing need for working methodologies that support data access, reuse, and sharing efforts. Several recommendations presented in this study are supported in literature by researchers who actively manage research data. This research provides a framework of opportunities that may assist NSF and other federal funding agencies as they continue to review and refine current data sharing strategies to minimize delays and likewise, encourage timely data sharing practices within scholarly research communities.

Appendix A

NSF Data Sharing Policy

NSF Data Sharing Policy

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing (NSF, 2011).

Award & Administration Guide (AAG) Chapter VI.D.4

a. Investigators are expected to promptly prepare and submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work conducted under NSF grants. Grantees are expected to permit and encourage such publication by those actually performing that work, unless a grantee intends to publish or disseminate such findings itself.

b. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. Privileged or confidential information should be released only in a form that protects the privacy of individuals and subjects involved. General adjustments and, where essential, exceptions to this sharing expectation may be specified by the funding NSF Program or Division/Office for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interest

of investigators. A grantee or investigator also may request a particular adjustment or exception from the cognizant NSF Program Officer.

c. Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.

d. NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.

e. NSF program management will implement these policies for dissemination and sharing of research results, in ways appropriate to field and circumstances, through the proposal review process; through award negotiations and conditions; and through appropriate support and incentives for data cleanup, documentation, dissemination, storage and the like (NSF, 2011).

Appendix B

Institutional Review Board Approval Letter

NOVA SOUTHEASTERN UNIVERSITY

MEMORANDUM

To: Mary Harward, M.S.
Graduate School of Computer and Information Sciences

From: David Thomas, M.D., J.D.
Chair, Institutional Review Board

Date: July 2, 2014

Re: *A Study of Data Sharing Practices within Scholarly Research Communities* – NSU IRB No. 02271427Exp.

I have reviewed the revisions to the above-referenced research protocol by an expedited procedure. On behalf of the Institutional Review Board of Nova Southeastern University, *A Study of Data Sharing Practices within Scholarly Research Communities* is approved in keeping with expedited review category # 6 and # 7. Your study is approved on **July 1, 2014** and is approved until **June 30, 2015**. You are required to submit for continuing review by **May 30, 2015**. As principal investigator, you must adhere to the following requirements:

1) CONSENT: You must use the stamped (dated consent forms) attached when consenting subjects. The consent forms must indicate the approval and its date. The forms must be administered in such a manner that they are clearly understood by the subjects. The subjects must be given a copy of the signed consent document, and a copy must be placed with the subjects' confidential chart/file.

2) ADVERSE EVENTS/UNANTICIPATED PROBLEMS: The principal investigator is required to notify the IRB chair of any adverse reactions that may develop as a result of this study. Approval may be withdrawn if the problem is serious.

3) AMENDMENTS: Any changes in the study (e.g., procedures, consent forms, investigators, etc.) must be approved by the IRB prior to implementation.

4) CONTINUING REVIEWS: A continuing review (progress report) must be submitted by the continuing review date noted above. Please see the IRB web site for continuing review information.

5) FINAL REPORT: You are required to notify the IRB Office within 30 days of the conclusion of the research that the study has ended via the IRB Closing Report form.

The NSU IRB is in compliance with the requirements for the protection of human subjects prescribed in Part 46 of Title 45 of the Code of Federal Regulations (45 CFR 46) revised June 18, 1991.

Cc: Dr. Ling Wang
Dr. Steven Zink
Ms. Jennifer Dillon

Institutional Review Board
3301 College Avenue • Fort Lauderdale, Florida 33314-7796
(954) 262-5369 • Fax: (954) 262-3977 • Email: *irb@nsu.nova.edu* • Web site: *www.nova.edu/irb*

Appendix C

Survey Questions

1. What is your primary role at your institution?

- Principal Investigator
- Co-Investigator

Other (please specify)

2. How long have you worked at your institution in this role?

- Less than 1 year
- 1-5 years
- 5-10 years
- 10+ years

3. How would you rate data sharing as an important aspect of research to you?

- Very Important
- Important
- Somewhat important
- Not important
- Undecided

4. Overall, how would you rate data sharing as an important aspect of research at your institution?

- Very important
- Important
- Somewhat important
- Not important
- Undecided

5. Which answer best describes current data sharing practices with researchers outside your institution?

- Always share
- Often share
- Seldom share
- Never share
- I am unaware of data sharing practices

6. At your institution, which individual(s) are responsible to complete and submit the National Science Foundation federal grant application? (check all that apply)

- Principal Investigators
- Co-Investigators

Other (please specify)

7. Within the past two years, have you completed a National Science Foundation federal grant application?

- Yes
- No

8. Approximately how many NSF federal grant applications have you completed in the last two years?

- 1-3
- more than 3
- Not applicable

9. Does the NSF federal grant application facilitate your current data sharing practices?

- Yes
- No

10. What enhancements to the NSF federal grant application are recommended, if any?

11. In the past two years, have you completed a supplemental Data Management Plan in the NSF federal grant application?

- Yes
- No

12. Approximately how many supplemental Data Management Plans have you completed in the last two years?

- 1-3
- More than 3
- Not Applicable

13. Describe the impact of the Data Management Plan to facilitate your data sharing practices.

14. What enhancements to the Data Management Plan are recommended, if any?

15. Overall, how would you rate the effectiveness of the NSF federal grant application process to facilitate your data sharing practices?

- Very effective
- Somewhat effective
- Seldom effective
- Not effective
- I am undecided

16. Overall, how would you rate the effectiveness of the NSF federal grant application process to facilitate data sharing practices at your institution?

- Very effective
- Effective
- Somewhat effective
- Not effective
- Undecided

17. What enhancements to the overall NSF federal grant application process do you recommend to facilitate data sharing in scholarly research?

18. Additional comments:

Appendix D

Survey Participation Letter of Consent

NOVA SOUTHEASTERN UNIVERSITY
Graduate School of Computer and Information Sciences

Participation Letter

Survey Participation Letter for the Study Entitled:
A Study of Data Sharing Practices within Scholarly Research Communities

Funding source: None

IRB protocol #:

Principal investigator:
Mary Harward,
Doctoral candidate for Information Systems
1851 NW 99 Avenue
Plantation, Florida 33322
Phone: (954) 383-9443

Co-Investigator
Steven P. Zink Ph.D.
Vice Chancellor, Information Technology
Nevada System of Higher Education
5550 N. Flamingo Road Suite C-1
Las Vegas, NV 89103-0137
(775) 784-3494

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board)
Nova Southeastern University
(954)262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

Description of the Study

What is the study about?
The aim of this research study is to review the NSF federal grant application process and its influence on timely data sharing practices. Most importantly, the goal of this study will be to identify definitive ways in which the NSF grant application process may be improved to expedite the sharing of research data in the future.

Why are you asking me?
You were selected as a potential candidate for this study in your role as Principal Investigator, Co-Investigator, or NSF employee.

What will I be doing if I agree to be in the study?
You have been selected as a potential candidate for the survey phase of this research study. 350 surveys will be sent to Principal Investigators, Co-investigators, and/or National Science Foundation staff located in the Southeast Sunbelt region and the National Science Foundation.

The survey will consist of 18 questions on the subject of data sharing and the NSF federal grant application process. The nature of the questions will focus on your experiences in completing the federal grant application and your suggestions for improving the current grant application process. Your participation in the study is completely voluntary and you may terminate participation at any time by not completing the survey.

Risks

What are the dangers to me?
The procedures or activities in this study may have unknown or unforeseeable risks, however, known risk to study participants is minimal. If you have any questions about the research, your research rights, or have a research-related injury, please contact Mary Harward, Principal Investigator and Steven Zink, Co-Investigator. You may also contact the IRB at the numbers indicated above with questions as to your research rights.

Benefits

Are there any benefits to me for taking part in this research study?
There are no direct benefits for participating in this study.

Compensation and/or Cost

Will I get paid for being in the study? Will it cost me anything?
There are no costs to you or payments made for participating in this study.

Confidentiality

How will you keep my information private?
All information obtained in this study will be retained for 36 months from the conclusion of the study. Data will be stored on a password-protected laptop that will remain at the Principal Investigator's residence. All information obtained in this study is strictly confidential unless disclosure is required by law. The record retention period will be a minimum of three years after the study is over, as required by the NSU-IRB. The NSU-IRB and other regulatory agencies may review research records, as well as the Principal Investigator and dissertation chair/thesis advisor.

Voluntary Consent

What if I want to leave the study?
You have the right to refuse to participate in or withdraw from the study at any time without adverse affects or loss of benefits to which you are otherwise entitled.

You have the right to leave this study at any time or refuse to participate. If you do decide to leave or you decide not to participate, you will not experience any penalty or loss of services you have a right to receive. If you choose to withdraw, any information collected about you **before** the date you leave the study will be kept in the research records for 36 months from the conclusion of the study but you may request that it not be used.

Additional Information

Other Considerations
If significant new information relating to the study becomes available which may relate to your willingness to continue to participate, this information will be provided to you by the investigators.

Completion of this survey implies my consent to participate in this research study.

Appendix E

Survey - Email Participation Request Letter

NOVA SOUTHEASTERN UNIVERSITY
Graduate School of Computer and Information Sciences

Participation Letter

Hello,

As a brief introduction, my name is Mary Harward and I am a current Ph.D. candidate studying Information Systems at the Graduate School of Computer and Information Sciences at Nova Southeastern University in Fort Lauderdale, Florida. I am currently working on my dissertation entitled "A Study of Data Sharing Practices in Scholarly Research Communities". My study offers insight into the National Science Foundation (NSF) federal grant application process to understand its impact on data sharing in research. Outcomes from this study may be valuable to identify factors that impact data sharing within the scope of the NSF grant application process.

This email is a request to participate in the study by completing an anonymous online survey. Please be assured that participation is at your discretion and is completely voluntary.

The survey consists 18 questions and should take approximately 20 minutes to complete. Information gathered during this survey will be anonymous and used exclusively for my dissertation research.

If you are willing to participate, please access the survey by following the link below:

http://surveymonkey.com/

Please feel free to contact me with any questions at mary@nova.edu.

Thank you in advance for your participation.

Best regards,

Mary Harward

NOVA SOUTHEASTERN UNIVERSITY
Institutional Review Board
Approval Date: JUL 0 1 2014
Continuing Review Date: JUN 3 0 2015

3301 College Avenue • Fort Lauderdale, Florida 33314-7796 • (954) 262-2000 • 800-541-6682, ext. 2000
Fax: (954) 262-3915 • Web site: *www.scis.nova.edu*

Appendix F

Adult/General Informed Consent Letter

NOVA SOUTHEASTERN UNIVERSITY
Graduate School of Computer and Information Sciences

Adult/General Informed Consent

Consent Form for Participation in the Study Entitled:
A Study of Data Sharing Practices within Scholarly Research Communities

Funding source: None

IRB protocol #:

Principal investigator:
Mary Harward,
Doctoral candidate for Information Systems
1851 NW 99 Avenue
Plantation, Florida 33322
Phone: (954) 383-9443

Co-Investigator
Steven P. Zink Ph.D.
Vice Chancellor, Information Technology
Nevada System of Higher Education
5550 N. Flamingo Road Suite C-1
Las Vegas, NV 89103-0137
(775) 784-3494

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board)
Nova Southeastern University
(954)262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

Site Information:

University of Florida 1765 Stadium Road Gainesville, FL 32611
University of South Florida 4202 E. Fowler Avenue Tampa, FL 33620
Florida International University 11200 SW 8th Street Miami, FL 33174
Florida State University 600 W College Avenue Tallahassee, FL 32306
University of Central Florida 4000 Central Florida Boulevard Orlando, FL 32816
The National Science Foundation 4201 Wilson Boulevard Arlington, VA 22230

Initials: _____ **Date:** _____

Description of the Study

What is the study about?
The aim of this research study is to review the NSF federal grant application process and its influence on timely data sharing practices. Most importantly, the goal of this study will be to identify definitive ways in which the NSF grant application process may be improved to expedite the sharing of research data in the future.

Why are you asking me?
You were selected as a potential candidate for this study in your role as Principal Investigator, Co-Investigator, or NSF employee. There are 30 participants that will be selected as potential candidates in the study.

What will I be doing if I agree to be in the study?
You have been selected as a potential candidate for the interview phase of this research study. Thirty qualitative interviews will be conducted with Principal Investigators, Co-investigators, and/or National Science Foundation staff located at select Florida Universities and the National Science Foundation.

The interview will consist of one 30-minute interview, either face-to-face at your institution or remotely via Skype or telephone. During the interview, specific questions about data sharing and the NSF federal grant application process will be asked. The nature of the questions will focus on your experiences in completing the federal grant application and your suggestions for improving the current grant application process. The session will be audio taped for reference purposes during the data analysis phase of the study. Your participation in the study is completely voluntary and you may terminate participation at any time.

Audio or Video Recording

Is there any audio or video recording?
This research project will include audio recording of the interview to discuss data sharing practices at your institution and will be recorded by using the Garage band application on a MacBook Air laptop. This audio recording will be available to be heard by the researcher, the IRB, and the dissertation chair or committee. The recording will be not be transcribed.

The recording will be kept on a secure computer drive that is only accessible by the Principal Investigator. The recording will be kept for 36 months and destroyed after that time by the Principal Investigator, who will securely delete the information from the laptop hard drive. Because your voice will be potentially identifiable by anyone who hears the recording, your confidentiality for things you say on the recording cannot be guaranteed, although the researcher will try to limit access to the tape as described in this paragraph.

Risks

What are the dangers to me?
The procedures or activities in this study may have unknown or unforeseeable risks, however, known risk to study participants is minimal. Loss of confidentiality and loss of privacy are the primary risks associated with this research study. Privacy deals with control over how much data is shared and others knowing that participants are part of the group. Confidentiality deals with how data will be safeguarded and stored. Information gathered during the interview will be anonymous and used exclusively for my dissertation research. If you have any questions about the research, your research rights, or have a research-related injury, please contact Mary Harward, Principal Investigator and Steven Zink, Co-Investigator. You may also contact the IRB at the numbers indicated above with questions as to your research rights.

Benefits

Are there any benefits to me for taking part in this research study?
There are no direct benefits for participating in this study.

Compensation and/or Cost

Will I get paid for being in the study? Will it cost me anything?
There are no costs to you or payments made for participating in this study.

Confidentiality

How will you keep my information private?
All information obtained in this study will be retained for 36 months from the conclusion of the study. Data will be stored on a password-protected laptop that will remain at the Principal Investigator's residence. All information obtained in this study is strictly confidential unless disclosure is required by law. The record retention period will be a minimum of three years after the study is over, as required by the NSU-IRB. The NSU-IRB and other regulatory agencies may review research records, as well as the Principal Investigator and dissertation chair/thesis advisor.

Voluntary Consent

What if I want to leave the study?
You have the right to refuse to participate in or withdraw from the study at any time without adverse affects or loss of benefits to which you are otherwise entitled.

You have the right to leave this study at any time or refuse to participate. If you do decide to leave or you decide not to participate, you will not experience any penalty or loss of services you have a right to receive. If you choose to withdraw, any information collected about you **before** the date you leave the study will be kept in the research records for 36 months from the conclusion of the study but you may request that it not be used.

Additional Information

Other Considerations
If significant new information relating to the study becomes available which may relate to your willingness to continue to participate, this information will be provided to you by the investigators.

Voluntary Consent by Participant:

By signing below, you indicate that

- this study has been explained to you
- you have read this document or it has been read to you
- your questions about this research study have been answered
- you have been told that you may ask the researchers any study related questions in the future or contact them in the event of a research-related injury
- you have been told that you may ask Institutional Review Board (IRB) personnel questions about your study rights
- you are entitled to a copy of this form after you have read and signed it
- you voluntarily agree to participate in the study entitled "A Study of Data Sharing Practices within Scholarly Research Communities"

Participant's Signature: _____ Date: _____

Participant's Name: _____ Date: _____

Signature of Person Obtaining Consent: _____ □

Date: _____

NOVA SOUTHEASTERN UNIVERSITY
Institutional Review Board
Approval Date: JUL 0 1 2014
Continuing Review Date: JUN 3 0 2015

Appendix G

Interview Guide

Interview Guide

Q1. What is your role at your institution?

Q2. What is your primary discipline of science and how many years have you been conducting research?

Q3. Have you recently accessed data shared by other researchers? If so, has access to shared data been readily available to you?

Q4. How would you rate data sharing as an important aspect of research to you? Importance at your institution?

Q5. Do you share data with other researchers outside of your institution? Do you access data from a repository?

Q6. What key factors hinder the sharing of research data at your institution with other researchers? In your opinion, which is the most important factor?

Q7. What role does the Principal Investigator at your institution play to complete and/or file the Federal grant application? If this is a shared responsibility, what other individuals are responsible to file this application?

Q8. In your opinion, does the Federal grant application process adversely impact data sharing efforts? If so, how?

Q9. Are you responsible for filling out Federal grant application proposals for your institution?

Q10. In what ways would you improve the Federal grant application process to improve data sharing opportunities?

Q11. Do you have any additional comments on how the sharing of research data may be improved at your institution?

Reference List

Alavi, M., & Leidner, D. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly, 25*(1), 107-136.

Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation, 8*(2), 5-26.

Allard. S. (2013). DataONE: Facilitating eScience through collaboration. *Journal of eScience Librarianship, 1*(1), 4-17.

Antell, K., Foote, J. B., Turner, J., & Shults, B. (2014). Dealing with data: Science librarians' participation in data management at Association of Research Libraries institutions. *College & Research Libraries, 74*(4), 557-574.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., . . . Wooters, P. (2004, March 19). An international framework to promote data sharing. *Science, 303*, 1777-1778.

Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. Proceedings of the ACM Conference on Supporting Group Work, Sanibel Island, FL, 339-348.

Bishoff, C., & Johnston, L. (2015). Approaches to data sharing: An analysis of NSF data management plans from a large research university. *Journal of Librarianship and Scholarly Communication, 3*(2), eP1231. http://doi.org/10.7710/2162-3309.1231

Blumenthal, D., Campbell, E. G., Anderson, M. S., Causino, N., & Louis, K. S. (1997). Withholding research results in academic life science: Evidence from a national survey of faculty. *Journal of the American Medical Association, 277*(15), 1224-1228.

Borgman, C. L. (2010). Who will share what, with whom, when, and why? *China-North American Library Conference, Beijing, China*, 1-21.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the Association for Information Science and Technology*, *63*(6), 1059-1078.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., & Blumenthal, D. (2002). Data withholding in academic genetics: Evidence from a national survey. *The Journal of the American Medical Association, 287*(4), 473–480.

Choudery, S., Fishman, J. R., McGowan, M. L., Juengst, E. T. (2014). Big data, open science, and the brain; Lessons learned from Genomics. *Frontiers in Human Neuroscience, 239*(8), 1-10.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: Complementary research strategies.* Thousand Oaks, CA: Sage Publications, Inc.

Collins, H. M. (2001). Tacit knowledge, trust and the Q of sapphire. *Social Studies of Science, 31*(1), 71-85.

Contreras, J. L. (2010, July 23). Prepublication data release, latency, and genome commons. *Science, 329*, 393-394.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative methods.* Thousand Oaks, CA: Sage Publications, Inc.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research (2nd ed.).* Thousand Oaks, CA: Sage Publications, Inc.

De Wolf, V. A., Sieber, J. E., Steel, P. M., & Zarate, A. O. (2005). What is the requirement for data sharing*? IRB: Ethics and Human Research, 27*(6), 12-16.

Dimitrova, V. (2013). Open access to research data: The European Commission's consultation in progress. Retrieved July 13, 2013 from http://openeconomics.net/2013/07/09/open-access-to-research-data/

Eickoff, S., Nichols, T. E., Van Horn, J. D., & Turner, J. A. (2016, January 1). Sharing the wealth: Brain imaging repositories in 2015. *Neuroimage, 124*(Part B), 1065-1068.

Exec. Order No. 13,642, 3 C.F.R. 78 FR 28,111 (2013).

Fecher, B. Frieske, S., & Hebing, M. (2015). What drives academic data sharing? *PLoS ONE 10*(2): e0118053. doi:10.1371/journal.pone.0118053

Fienberg, S. E., Martin, M. E., & Straf, M. L. (1985). *Sharing research data*. Washington: National Academy Press.

Fienberg, S. E. (1994). Sharing statistical data in the biomedical and health sciences: Ethical, institutional, legal and professional dimensions. *Annual Rev. Public Health, 15*(1), 1-18.

Fink, A. (2003). *The survey handbook*. Thousand Oaks, CA: Sage Publications, Inc.

Griffiths, A. (2008). The publication of research data: Researcher attitudes and behaviour. *The International Journal of Digital Curation, 1*(4), 46-56.

Haeussler, C. (2011). Information-sharing in academia and the industry: A comparative study. *Research Policy, 40*(1), 105-122.

Hammersley, M. (1989). *The dilemma of qualitative method: Herbert Blumer and the Chicago tradition*. London, New York: Routledge.

Hane, P. J. (2010). Bill introduced for open access to federally funded research – FRPAA revisited. Retrieved September 30, 2010, from http://newsbreaks.infotoday.com/NewsBreaks/Bill-Introduced-for-Open-Access-t%20o-Federally-Funded-ResearchFRPAA-Revisited-67078.asp

Harris, W. (2009). How scientific peer review works. Retrieved June 22, 2011, from http://science.howstuffworks.com/innovation/scientific-experiments/scientific-peer-review1.htm

Harttner, J., Ryan, S. J., MacKenzie, C. A., Parker, J. N., & Strasser, C. A. (2013). Spatially explicit data: Stewardship and ethical challenges in science. *PLoS Biology, 11*(9), e10016634-e1001635.

Heath, H., & Cowley, S. (2004). Developing a grounded theory approach: A comparison of Glaser and Strauss. *International Journal of Nursing Studies, 41*(2), 141-150.

Holden, J. P. (2013). *Increasing access to the results of federally funded scientific research* [Memorandum]. Washington, D.C.: Office of Science and Technology.

ICPSR. (2016). Membership in ICPSR. Retrieved Aprol 10, 2016, from https://www.icpsr.umich.edu/icpsrweb/content/membership/about.html

Janesick, V. (2003). The choreography of qualitative research: Minuets, improvisations, and crystallization. *Strategies of qualitative inquiry.* Thousand Oaks, CA: Sage Publications, Inc.

Kitchin, J. R. (2015). Examples of effective data sharing in scientific publishing. Retrieved July 17, 2016, from http://pubs.acs.org/doi/ipdf/10.1021/acscatal.5b00538

Langfeldt, L., & Scordato, L. (2015). Assessing the broader impacts of research: A review of methods and practices. Retrieved April 13, 2016, from https://brage.bibsys.no/xmlui//bitstream/handle/11250/282742/NIFUworkingpaper2015-8.pdf?sequence=1

Lane, J., Heus, P., & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinferstructure. *Transactions on Data Privacy, 1*(1), 2-16.

Longo, D. L., & Drazen, J. M. (2016). Data sharing. *New England Journal of Medicine, 374*(1), 276–277.

MacMillan, D. (2014). Data sharing and discovery: What librarians need to know. *The Journal of Academic Librarianship, 40*(5), 541-549.

Mason, J. (2004). Interview guide. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), (p. 519). Sage Publications, Inc. doi: 10.4135/9781412950589.n450

Mathison, S. (1988). Why triangulate? *Educational Researcher, 17*(2), 13-17.

Maxwell, J. A. (2005). *Qualitative research design: An interactive approach*. Thousand Oaks: CA, Sage Publications, Inc.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.) Thousand Oaks: CA, Sage Publications, Inc.

Manyika, J., Chui, M., Brown, B., Burghin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved July 10, 2013 from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Mischo, W. H., Schlembach, M. C., & O'Donnell, M. N. (2014). An analysis of data management plans in University of Illinois National Science Foundation grant proposals. *Journal of eScience Librarianship, 3*(1): Article 3.

Morse, J. M. (2003). Principles of mixed methods and multimethod research design. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 189-208). Thousand Oaks, CA: Sage Publications, Inc.

Nature. (2014, March 12). Share alike. *Nature, 507*(7491), 140.

Nelson, B. (2009, September 10). Data sharing: Empty archives. *Nature, 461*, 160-163.

Nulty, D. 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.

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., . . . Dierolf, U. (2013). Making research data repositories visible: The re3data.org registry. *PLoS ONE 8*(11): e78080. doi:10.1371/journal.pone.0078080

Parry, M. (2010). As open-access chatter grows, U. of Rochester debuts new repository software. Retrieved March 20, 2016, from http://chronicle.com/blogs/wiredcampus/as-open-access-chatter-grows-u-of-rochester-debuts-new-repository-software/20807

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks: CA: Sage Publications, Inc.

Pittman, D. (2010, September 10). NSF revamps data-sharing policy. *Chemical and Engineering News, 88*(39), 46-47.

Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., … Kennedy, D. N. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics*, *6*, 9. http://doi.org/10.3389/fninf.2012.00009

Pryor, G. (2012). *Managing research data.* London: Facet Publishing.

re3data.org. (2016). Enhancements to creating and updating the registry. Retrieved March 19, 2016, from http://www.re3data.org

Research Information Network. (2008). To share or not to share: Publication and quality assurance of research data outputs. Retrieved March 16, 2016, from http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf

Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., . . . Kruuk, L. E. B. (2014). Troubleshooting public data archiving: Suggestions to increase participation. *PLoS Biology, 12*(1), e1001779.

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE*: 4: e7078.

Schofield, P. N., Bubelaz, T, & Weavers, L. P. (2009, September). Post-publication sharing of data and tools. *Nature, 461*, 168-170.

Schneider, C. E. (2015). *The censor's hand: The misregulation of human-subject research.* Cambridge, MA: The MIT Press.

Schwandt, T. A. (2015). *The SAGE dictionary of qualitative inquiry* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Sekaran, U. (2003). *Research methods for business: A skill building approach* (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedure and techniques*. Newbury Park, London: Sage Publications, Inc.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research.* Thousand Oaks, CA: Sage Publications, Inc.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE, 6*(6), 1-21.

Turner, S., Cardinal, L. B., & Burton, R. M. (2015). Research design for mixed methods: A triangulation-based framework and roadmap. *Organizational Research Methods*, 1094428115610808, first published on November 25, 2015 doi:10.1177/1094428115610808.

U.S. Committee on Science, Engineering, and Public Policy (2009). *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Washington, D.C.: National Academies Press.

U.S. Department of Energy (2015). *DOE policy for digital research data management.* Retrieved April 9, 2016, from http://www.energy.gov/datamanagement/doe-policy-digital-research-data-management

U.S. National Science Board (2011). *Digital research data sharing and management.* Retrieved April 6, 2016, from https://www.nsf.gov/nsb/publications/index.jsp?timeframe=date+range&queryText=digital+research+data+sharing+and+management&firstDate=&lastDate=&docType=NSB+Board+Reports&search2=Search

U.S. National Science Foundation (2007, March). *Cyberinfrastructure vision for 21st century discovery*. Retrieved September 10, 2010, from http://www.nsf.gov/pubs/2007/nsf0728/index.jsp

U.S. National Science Foundation (2011). *Chapter II: Proposal preparation instructions.* Retrieved March 25, 2012, from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp

U.S. National Science Foundation (2014). *Open data at NSF.* Retrieved May 10, 2016 from http://www.nsf.gov/data/

U.S. National Science Foundation (2015). *Today's data, tomorrow's discoveries.* Retrieved April 10, 2016 from https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf15052

U.S. National Science Foundation (2016a). *The National Science Foundation: Proposal and award policies and procedures guide*. Retrieved August 6, 2016 from http://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp

U.S. National Science Foundation (2016b). *Chapter II: Proposal preparation instructions.* Retrieved April 2, 2016 from http://www.nsf.gov/pubs/policydocs/pappguide/nsf16001/gpg_2.jsp

U.S. National Science Foundation (2016c). *National Center for atmospheric research.* Retrieved April 6, 2016 from https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=12809

U.S. National Science Foundation (2016d). *Earthcube.* Retrieved May 10, 2016 from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504780

Warren, C. (2004). Interviewing in qualitative research. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), (pp. 522-525). Sage Publications, Inc. doi: 10.4135/9781412950589.n455

Watts, S., George, M., & Levey, D. (2015). Achieving broader impacts in the National Science Foundation, Division of Environmental Biology. *Bioscience*, *65*(4), 397-407.

Wilbanks, J., & Friend, S. H. (2016). First, design for data sharing. Retrieved March 20, 2016, from http://www.nature.com/nbt/journal/vaop/ncurrent/full/nbt.3516.html

Wright, J. (2015). Libraries cheer passage of strong open access legislation in U.S. Senate. Retrieved April 10, 2016, from http://www.ala.org/news/press-releases/2015/07/libraries-cheer-passage-strong-open-access-legislation-us-senate