This study seeks to determine the needs of the social science research and information professional communities regarding data management training and education. The needs assessment surveys conducted in this study work towards answering three main questions: How do researchers and the information professionals who serve them understand the term “data management”? What are some of the gaps in data management knowledge of the social science and information professional communities that need to be addressed? What preferences do researchers and information professionals have in regards to data management training and education dissemination methods? An online needs assessment survey was e-mailed to 1,974 participants with a total response rate of 10.5%. The results of this study emphasize a broad understanding of the term “data management” with a focus on storage and sharing, a need for education on long-term data storage best practices and data management tools, as well as a preference for short, online video courses with an open discussion forum.

Headings:

- Needs assessment
- Data management
- Massive open online course (MOOC)
- Information professionals
- Social science researchers
- Curricula (Courses of study)
DATA STORAGE AND SHARING: A NEEDS ASSESSMENT SURVEY OF SOCIAL SCIENCE RESEARCHERS AND INFORMATION PROFESSIONALS FOR DEVELOPING A DATA MANAGEMENT CURRICULUM

by
Amanda J. Gooch

A Master’s paper submitted to the faculty of the School of Information and Library Science of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in Library Science.

Chapel Hill, North Carolina
April 2014

Approved by

_______________________________________

Helen Tibbo
TABLE OF CONTENTS

LIST OF IN-TEXT TABLES AND FIGURES ................................................. 2

CHAPTER

1. INTRODUCTION ....................................................................................... 3

2. LITERATURE REVIEW ............................................................................. 6

3. METHODOLOGY ....................................................................................... 18

4. RESULTS ................................................................................................... 24

5. DISCUSSION & CONCLUSION ................................................................. 38

BIBLIOGRAPHY ............................................................................................ 46

APPENDIX A: TABLES .................................................................................. 50

APPENDIX B: SURVEY PROTOCOL .............................................................. 57

APPENDIX C: E-MAIL INVITATIONS ............................................................ 74
LIST OF IN-TEXT TABLES AND FIGURES

TABLES

Table 4.1: Researcher familiarity with term “data management” .........................24
Table 4.2: Information professional familiarity with term "data management" ....25
Table 4.3: Typical storage space during active research .................................28
Table 4.5: How do you share your data? ..............................................................29
Table 4.6: Data management services offered by library/institution ..................30
Table 4.9: Resources available to learn about data management .......................32
Table 4.10: Researchers’ preference for topics to be covered in data management course ........................................................................................................33
Table 4.11: Information professionals' preference for topics to be covered by data management course .................................................................34

FIGURES

Figure 3.1: Information professionals subject specialties .................................22
Figure 4.1: What does 'data management' mean to information professionals? ....25
Figure 4.2: What does 'data management' mean to researchers? .........................26
Figure 4.3: Ranking order of online delivery methods preferences ......................36
Figure 4.4: Ranking order of online delivery methods preferences of information professionals .................................................................36
CHAPTER ONE

INTRODUCTION

Data storage and sharing, often referred to as data management, is an area of importance in the social science community. Funding agencies such as the National Science Foundation (NSF) and the National Institutes of Health (NIH) are requiring all current and future proposals to include data management plans. Within these data management plans there are requirements for a data sharing policy and long-term data storage. While these requirements are being reviewed and evaluated during the proposal selection process, the funding agencies are not offering training or education in data management for either grant applicants or recipients. This has led to a broad, but not deep, understanding of the issues involved with long-term storage and data sharing among the social science research communities.

Further, academic libraries and archives are tackling the job of assisting social science researchers in data management by providing training, education, long-term storage, and sometimes preservation. Information professional positions that may work with social science researchers such as: data librarians, GIS librarians, social science subject specialists, scholarly communication librarians, and digital initiatives librarians have started taking steps towards offering these services despite a limited understanding of the needs of the researcher communities and knowledge about data management.
Currently there are tools and resources available to academic researchers, information professionals, and the public regarding good practice in data management; however, most of these resources rely upon the individual’s initiative and self-learning. Online and face-to-face educational courses do exist for data management; however, they are limited in number and require interested individuals to seek them out. Social scientists, much like their colleagues in other disciplines rarely seek out data management training and thus there are gaps in their data management knowledge. Information professionals often exhibit very similar knowledge gaps concerning data management. In order to bridge these gaps and to increase awareness of available data management tools and resources, the CRADLE (Curating Research Assets and Data using Lifecycle Education) Project will develop an online course and other workshops and tools that will educate social science researchers and information professionals in data management best practices, tools, and resources.

The purpose of this study is to identify the educational needs of the social science researcher and information professional communities in regards to data management. This study will address three main questions:

Q1) How do researchers and the information professionals who serve them understand the term “data management”?

---

1 Tools such as the DMPTool, Databib, and DataOne aid researchers and information professionals in developing data management plans, locating data, and building frameworks and cyberinfrastructure for data archives and digital repositories. Additionally, courses such as MANTRA allow users to select topics and view short videos to learn about data management at their own pace.

2 MANTRA is an online educational tool for data management. Coursera offers a Data Management MOOC for Clinical Research. Face-to-face workshops have been offered at multiple institutions such as the workshops at the Digital Curation Center, University of Illinois at Urbana-Champaign, and CURATEcamp partnered with the Digital Library Federation.

3 More information about the CRADLE Project can be found at: http://cradle.web.unc.edu/
Q2) What are some of the gaps in data management knowledge of the social science researcher and information professional communities that need to be addressed?

Q3) What preferences do social science researchers and information professionals have in regards to data management training and education dissemination methods?

This study will take a quantitative research approach by analyzing responses from an online, e-mail needs assessment survey in an attempt to answer the above questions.

The outcome of this study will help the CRADLE Project and future data management courses to determine what curricula needs to be developed for the social science and information professional communities. Apart from bridging the gaps in data management knowledge, this curriculum will help to increase awareness of the data management tools and resources currently available and unite social science researchers and information professionals in an online data management network. Furthermore, this study will assist the CRADLE Project in selecting the appropriate educational platforms to disseminate the data management curricula to these communities.
CHAPTER TWO
LITERATURE REVIEW

Necessity of Education

Researchers

With funding agencies such as the NIH\textsuperscript{4} and the NSF\textsuperscript{5} requiring more transparency and encouraging data reuse, it is increasingly necessary for scientists to make their data publicly available. As more funders follow the NSF example, social scientists are finding they too must create data management plans and place their data in digital repositories. Underlying the notion of data repositories is data sharing, something in which researchers may not want to engage.

Vertesi and Dourish’s (2011) study followed two groups of scientists dealing with space exploration data in order to determine the researchers’ concepts of data sharing. They discovered that, “scientists do not, in principle, share their data with other scientists” (Vertesi and Dourish, 2011, p. 533). The study revealed that there is a hierarchy to data ownership. This hierarchy dictates who gets access to the data during its lifecycle. Attempting to access the data outside of the accepted norm usually resulted in long wait periods and a slew of obstacles.

\footnote{\textsuperscript{4} NIH Data Sharing Policy and Implementation Guidance (2010) states, “data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data” (para 3).

\textsuperscript{5} “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” (NSF Data Sharing Policy, 2012, para 1).}
This finding begs the question, if scientists are not interested in sharing data among their own closed community, why would they be interested in sharing data through repositories open to the general public? Vertesi and Dourish (2011), further explain that data production methods and established rules among the research teams tend to dictate the definitions and understanding of data management (p. 539). This concept is seen again in a variety of studies conducted with researchers (Gray et al., 2005; Scaramozzino, Ramirez, & McGaughey, 2012; Doty & Akers, 2013). In each of these studies researchers were asked questions targeted at determining their attitudes and understanding of data management and sharing. The results were similar across all studies in that the majority of respondents were unclear as to the meaning of data management and their roles in the process. According to Doty and Akers (2013), almost 40 percent of respondents were not familiar with data management plans and over 60 percent of respondents did not share their data at all (slides 4-5).

Christine Borgman (2012) discusses how researchers are being required to manage their data and provide access, yet they do not fully understand the necessity. She defines four clear models for the rationales behind data sharing and how researchers usually understand one or two of these models in relation to their data:

“(a) to reproduce or to verify research, (b) to make the results of publicly funded research available to the public, (c) to enable others to ask new questions of extant data, and (d) to advance the state of research and innovation.” (Borgman, 2012, p. 1072)

Given these constraints, researchers and stakeholders are able to determine the importance of data sharing and management on some level, however, Borgman (2012) makes clear that these models are not exhaustive and are merely created to generate discussion among research data stakeholders (p. 1067). Earlier research conducted by
Borgman (2010), also suggests that there are a number of disincentives keeping researchers from fully embracing data management and data sharing:

“Scholars are hired and promoted based on their publication record rather than on the quality of their metadata. Secondly, documenting data is a labor-intensive process even for local use. Documenting methods, instrumentation, and software, and producing metadata at a level that the data are interpretable by others, can require much more labor than documentation for use by oneself or one’s team. Thirdly, researchers are concerned about establishing the priority of their claims on research findings in the face of competition. Embargo periods, where they exist, protect the investigator by providing a period of time to analyze data and publish results prior to the public release of their data. Lastly is the set of concerns for intellectual property, both the ability to control one’s own resources and the ability to gain access to resources controlled by others.” (Borgman, 2010, p. 7)

Gray et al. (2005) highlight some of these disincentives in their research, also noting that while researchers are being required to make their data public, they lack the knowledge necessary to make their data accessible via metadata and careful documentation (Metadata enables data access, para 4). This is proven again in the study conducted at California Polytechnic State University by Scaramozzino, Ramirez, & McGaughey in April 2010. The goal of this study was to determine the data management needs of the researcher community at CPSU. After surveying over 300 participants, it was discovered that, “only 20 percent of faculty report[ed] being aware of criteria for the creation of descriptive information to aid in discovery and reuse of data” (Scaramozzino, Ramirez, & McGaughey, 2012, p. 359). Even fewer faculty members felt they were knowledgeable in the creation of metadata for data discovery.

In a recent study conducted by OCLC researchers Kroll and Forsman (2010), it was determined that, “researchers generally report that they have no time to take on the burden of uploading their work to an IR [Institutional Repository], devising metadata, and creating useful organization” (p. 11). A curriculum that covers the necessity of data
management, the best practices, and the means for data discovery is essential in providing researchers with the necessary knowledge and tools to make data management an integral part of their data’s lifecycle. Data management education for researchers through online courses and face-to-face workshops will hopefully facilitate confidence and alter researchers’ current mindset that data management is simply an additional burden with little value.

Information Professionals

While information professionals are currently not working with researchers’ data throughout the data lifecycle, they are helping researchers store their data in repositories and share it with other researchers. As mentioned in Gray et al.’s article (2005), researchers are increasingly expected to document and describe their data in order to make it accessible, however, Walton (2010), suggests that it would be better for a “professional group to take responsibility for managing the data effectively for the benefit of the wider community” (p. 2). While it is unlikely that there will be enough data curators to manage all research from cradle to grave any time soon, if more information professionals were knowledgeable in data management this could go a long way toward instructing researchers to prepare their data for storage and future reuse from trustworthy repositories.

Numerous research and academic libraries have started to implement research data services that “address the full data lifecycle, including the data management plan, data curation…and metadata creation and conversion” (Tenopir, 2012, p. 70). This may seem proactive, however, there is still minimal education among the information
professionals that work with research data services. Tenopir’s (2012) study involved questioning information professionals who deal with researchers in order to determine what services were being offered, what information professionals thought about the services, and their necessity (p. 71 – 76). The results of the study reflect that most librarians felt it was necessary and beneficial, but that additional training and increased involvement in research data services is necessary (Tenopir, 2012, p. 77).

Additional studies and reports by individuals such as Gabridge (2009) and Ogburn (2010) provide similar arguments that information professionals understand the benefits and necessity of data management, but that they are not confident in offering extensive data management services due to gaps in their own education. In some instances, this has resulted in librarians banding together to educate themselves in the best practices and tools for data management. Furthermore, librarians are pushing for education at the graduate school level in order to ensure that future librarians are competent in data management practices.

Heidorn (2011), takes Tenopir’s concepts one step further by remarking that data curation is, “within the libraries’ mission, and libraries are among the only institutions with the capacity to curate many data types” (p. 663). While Tenopir feels that information professionals should be able to assist with data management, Heidorn (2011) argues that information professionals should actually deal with the curation process and consult the researchers for specifics (p. 667). In either case, there is a need for a data

---

6 “A group of science and engineering liaison librarians, calling themselves the Data Initiatives Group (DIG), formed a study group to learn collaboratively about the needs of researchers and the current state-of-the-art in providing services to manage research data, and to identify the skills required to actively respond to their faculty’s data curation needs” (Gabridge, 2009, pg. 16).

7 “In order to grow effective future librarians, we must urge our professional graduate programs to incorporate data management into their curricula” (Ogburn, 2010, pg. 244).
management curricula to be developed that would help information professionals address the needs of these research communities.

Data curation and data management are not just simple plans; information professionals need a solid grounding in these concepts in order to build successful data curation programs. Opening discussions between information professionals and researchers is just one step in successful data management. Information professionals will need to delve into the realm of new technologies and current best practices; a task that can be quite daunting without a basic understanding of the data lifecycle and data curation. A data management curriculum aimed at information professionals, should build confidence and support for data management in university libraries and institutional repositories. If information professionals are confident in their knowledge and skills, then they will be confident in working with researchers to ensure that all of their data curation needs are being met.

Clifford Lynch (2008) discusses how scientists will need to work with information professionals in order to ensure the longevity of their data. Additionally, he states that, “the best stewardship of data will come from engagement with preservation institutions” (Lynch, 2008, p. 28). In order to provide data longevity, information professionals need to be informed about data management best practices and tools to ensure data is successfully stored and shared. There are various tools and resources online that offer data management education, however, there is no curriculum that unites data curation best practices and tools for social scientists and information professionals.
**Researcher Needs Assessment Surveys**

While there remains a strong need for more data management education for information professions, the last ten years has seen some increase in data management awareness among universities and research institutions. As such, some information professionals have been surveying their local communities in an effort to determine the needs of their patrons and staff. These surveys are usually conducted via an online service such as SurveyMonkey or Qualtrics and often emphasize anonymous responses in order to generate a higher response rate.

Some needs assessment surveys have targeted specific departments or library staff concerning their knowledge, confidence levels, and current practices in data management or corresponding services. The University of Illinois at Urbana-Champaign Libraries decided to focus their efforts on the needs of early career faculty in their agriculture department. The respondents of this survey attended an in-person presentation introducing data management and the services being offered by UIUC. Attendees were then invited to participate in an anonymous online survey that focused on gauging respondents’ confidence levels and needs in data management training and services (Williams, 2013, p. 44). Of the seven respondents, 71% deemed data management training to be important. These results are currently being implemented into the University Libraries’ Researcher Workshop series in order assist in educating researchers in data management practices and tools (Williams, 2013, p. 48).

In an effort to broaden the results, other institutions have conducted needs assessment surveys that utilize questions similar to the UIUC study, however, the survey samples are from a variety of departments or institutions. The Data Management Rollout
at Oxford Project (DaMaRO) conducted multiple surveys across various departments engaged in data collection. The results of their surveys have assisted in narrowing down the areas of greatest interest for data management training among researchers at the University of Oxford (Dally et al., 2012). Greg Janée and James Frew (2013) surveyed faculty and researchers at the University of California Santa Barbara on their data curation practices, roles, and services. Their results coincide with previous findings and are being used to, “argue for the establishment of a campus unit possessing data curation expertise” (p. 1).

Data curation groups have also worked to survey even broader samples to determine the needs of the global communities. Engelhardt, Strathmann, and McCadden (2012) implemented a multi-nation survey of training needs for researchers and information professionals. They received 454 responses with a range of 86% to 96% of respondents signifying a moderate to great need for training in data management with an emphasis on the tools and resources available (p.39). Additionally, they conducted a focus group as well as a job advertisement analysis for positions related to data management. The results fell in line with the findings of the online needs assessment survey and correspond with the findings of previously mentioned surveys.

Other surveys and research have focused on developing personas for data curation or are comprised of interviews and focus group sessions with individual patrons and/or information professionals. Despite these different methodologies, the results still offer a look into the needs of the researcher and information professional communities. Lage, Lossof, and Maness’s (2011) case study at the University of Colorado Boulder is one such example. It created a protocol for interviewing individual researchers to develop
personas that would, “embody the aggregated attributes of faculty and graduate student researchers” (p. 916). Jake Carlson and Marianne Stowell-Bracke (2013) conducted a similar study at the Water Quality Field Station, however, this study focused on the needs of graduate research assistants. Given the smaller sample, Carlson and Stowell-Bracke assessed the needs of this community through individual interviews using data management modules instead of an online survey. After analyzing the results of these interviews, Carlson and Stowell-Bracke were able to create data curation profiles to describe the various types of data management concerns (p. 347 – 349). These personas and profiles have assisted in creating “types” of researchers and graduate research assistants to aid information professionals in understanding the needs of their patrons. By utilizing these personas, information professionals are able to place their patrons within a corresponding persona, which will assist them in determining the researcher’s skill and knowledge levels in regards to data management.

While these techniques may not prove to be completely accurate, they are a starting point for building a working relationship and beginning a conversation between the information professional and the researcher. Before an institution can begin to offer data management services, it must first determine the needs of its patrons and their current data management practices. The results of these surveys will aid institutions in developing tools and resources that will educate and serve researchers in data management.
MOOCs as an Educational Platform

Massive open online courses (MOOCs) have been expanding across the internet for the past six years. The essential idea behind a MOOC is to offer an online course in an open education environment to a large population for free. Most MOOC platforms also have additional features that users can buy into such as a signature track or a certification program. The MOOC platform unites practicing professionals, educators, students, and interested individuals across the globe in an online forum dedicated to instruction and collaboration.

A typical MOOC course could be anywhere from four to twelve weeks, depending on the needs of the instructor and the curriculum being covered. In most cases, MOOCs are made up of video instruction, sometimes including PowerPoint slides that take the place of in-class instruction. These videos can be short or long and the, “videos [could] pause perhaps twice for a quiz to make sure you understand the material or, in computer programming, to let you write code” (Pappano, 2012, para 11). If advertised, the enrollments for MOOCs usually number in the thousands. One of the most appealing aspects of a MOOC is that it allows for a course and instructor to reach across thousands of miles; thus, allowing for interested individuals to participate from across the globe. The students of MOOCs could be undergraduates, high school students, stay-at-home mothers, businessmen, etc. Anyone can join and participate in a MOOC and complete it to receive a certificate of completion. In a report by McAuley, Stewart, Siemens, and Corimier on the potential of MOOCs, they remark that:

“Learning, however, requires a human, social element: MOOCs provide both peer-based support and interaction with subject area experts. It is in this consolidation of the many aspects of learning for a collaborative, participatory
This level of interaction helps to build a community of professionals and students that can assist each other in their coursework through the MOOC discussion forums. Additionally, MOOC forums can be networking venues to create study groups, location or language groups, tutoring, and resource hubs.

Since MOOCs are relatively new, there is still much to learn from them and how best to maximize their potential. The attrition rate can be incredibly high for MOOCs, with thousands dropping out of the course, not participating, or never completing assignments and homework. Given that the courses are usually free and not for credit, there could be steps instructors or MOOC moderators can take to encourage participation; however, successful methods have not yet been discovered. Another concern with MOOCs is over student connectivity with instructors and teaching assistants. In a recent survey of MOOC participants, Mackness, Mak, and Williams (2010) discovered that, “huge diversity, resulting from open access, meant that the support and moderation that would normally be expected in a course were not feasible” (p. 272). The result was that students who felt a lack in moderation and instructor interaction did not have as great an experience as other students. Unfortunately, with high attendance rates, it is difficult for instructors to communicate with each student. Though there have been steps to improve this issue by bringing in teaching assistants and forum moderators.

This type of platform is potentially useful in educating and uniting researchers and information professionals across the globe in data management education. There are
pros and cons to any instructional platform, but in order to reach a larger audience in these communities, a massive open online course could be more successful than past efforts have been. There is much to learn about MOOCs as they grow and develop, yet they are a resource that could be beneficial if utilized appropriately.
CHAPTER THREE

METHODOLOGY

Data Collection

This study implemented a quantitative approach with an online, e-mail survey created through Qualtrics. The survey questions were created after extensive deliberation with the CRADLE Project team and a review of various university needs assessment surveys as mentioned in the literature review. Each question was reviewed to determine if the responses would give adequate insight into one of the three research questions. After the basic questions were developed, two surveys were created: one for social science researchers and one for information professionals. The language in the questions was tailored to each of these communities in order to ensure that the questions were relevant to the participants’ professions. For example, the researchers’ survey contained phrases such as data management practices, data storage and data sharing; whereas, the information professionals’ survey used terms such as data management services and data lifecycle.

The Researcher Survey was designed to first gauge the level of data management knowledge of each participant by asking questions related to data storage and sharing practices. Participants were then asked to define data management or, if a participant was not familiar with the term, they were asked to give their best definition. Finally, participants were able to denote their preferences for data management education and
curriculum delivery methods. The Information Professional Survey was structured similarly, however, instead of asking about data management practices, the participants were asked about current data management services being offered by their institution.

The participant selection process was based on a number of factors and dependent on the individual’s professional community. First, I compiled a list of 139 universities in the United States that the Carnegie Classification identified as having very high or high research activity. The social science researcher e-mail list was comprised of social science faculty, research staff, and post-doctoral students working or studying in one of these research universities. The information professionals e-mail list was populated with librarians, archivists, and faculty members from the same group of research universities across the United States. In order to ensure that the information professional survey was relevant to the sample, the information professionals selected to participate were chosen based on their interaction regarding data with the social science community. This determination was made by looking for subject specialists, committee members, or data curation specialists at each university.

After the survey participants were selected, they were e-mailed an invitation to participate in the online survey. It was estimated that completing the survey would take up to 15 minutes to complete. The average completion time for the survey was 7 minutes, which was well within the estimation. At the end of each survey participants were also asked if they would like to be contacted at a later time for additional follow-up. Those

---

9 This classification is based on the Carnegie Foundation for the Advancement of Teaching Basic Classification Listing. (classifications.carnegiefoundation.org/lookup_lists/standard.php)
10 See Survey Protocol in Appendix B and E-mail Invitation in Appendix C
who responded with ‘yes’ were kept on an e-mail list for interviews to be conducted at a later time for the CRADLE project.

**Response Rate**

The e-mail survey was sent out on January 2nd, 2014 to 999 information professionals and 998 social science researchers. Seven e-mails bounced back in the information professionals’ sample, which dropped the number to 992. Of the researchers, four participants requested to be removed from the survey list, one e-mail failed to be sent, and eleven e-mails bounced back. Thus, the final participant sample for the researchers was 982. This accounted for a total sample size of 1,974 participants. The survey remained open until January 20th, 2014 with one reminder e-mail being sent out halfway through the survey period. After the survey closed, the final responses were: 156 from information professionals and 52 from social science researchers, for a total of 208 responses. The individual response rate for the information professionals was 15.7% with a response rate of 5.3% for the social science researchers. The combined response rate was 10.5% of the 1,974 participants surveyed.

The limited response rate could be indicative of the level of understanding across both participant samples. The small number of researcher participants could be due to the fact that, of the whole, this percentage is familiar enough with data management to feel comfortable contributing to the current dialogue. The same can be inferred by the information professional response rate. This could mean that a high percentage of researchers and information professionals in the field today have a very limited understanding of data management or have zero exposure to the concept in their profession.
Participants

Despite the low response rate, the information gleaned from this survey is still important for data management courses and the CRADLE Project to utilize when developing their final curriculum. The demographics of the respondents is also crucial in order to understand which areas of the social science research community and information professionals community are interested in data management education and are willing to contribute to this effort. Of the researcher sample, 48% of respondents are tenured faculty, 36% are tenure-track faculty, 6% are fixed-term faculty, 4% are post-doctoral researchers, 4% are university staff, and 2% are listed as other and classify themselves as research faculty\(^\text{11}\). The information professionals’ respondent sample identified as 46% non-tenure librarians, 20% tenure-track faculty librarians, 20% tenured faculty librarians, 6% university staff, 5% other (including technology specialists, non-tenure-track faculty librarians, and professional staff), 2% archivists, and 1% post-doctoral researchers\(^\text{12}\).

The subject specialties of the respondents is necessary to understand which departments are being referenced in the survey responses. Within the researchers, 33% identified as Psychology, 31% as Sociology, 12% as Anthropology, 8% as Political Science, 6% as Economics, 4% as Other (public health and higher education), 2% as Education, 2% as Communication, and 2% as Law\(^\text{13}\). Within the information professionals’ respondents, quite a few identified as multiple subject specialists. The responses of the information professionals have been broken down as follows in Figure

\(^{11}\) See Table 3.1 in Appendix A
\(^{12}\) See Table 3.2 in Appendix A
\(^{13}\) See Table 3.3 in Appendix A
3.1, below: 69% in the Social Sciences, 22% in the Humanities, 16% in Data/Maps/GIS, 10% in the Natural Sciences, 10% in Government, 8% in Business, and 8% in Other (scholarly communication, math, sports, and digital initiatives).

Figure 3.1. Information professionals subject specialties. Q2a of Information Professional Survey. Note: The numbers before the categories correspond to the codes given to each category when determining frequencies from the raw data.

This variety will assist in providing a well-rounded curriculum that will address the needs of multiple departments within the social science research and information professionals’ communities.

Data Analysis

The data collected from these surveys has been grouped into three categories related to the three research questions. As such, responses to Researcher Survey Q3-Q10a and Information Professional Survey Q3-Q6a have been used to develop answers for Question 1. Question 2 has been addressed by the responses to Researcher Survey Q11-Q12 and Q14, and Information Professional Survey Q7-Q8 and Q11. And finally,
Question 3 has been answered with responses from Researcher Survey Q12.1-Q13 and Q16, and Information Professional Survey Q9-Q10 and Q12.

All of the text entry responses have been coded based on frequency of terms and/or phrases. Any identifying information was removed to ensure the anonymity of the respondents. The findings have been relayed using descriptive statistics to show frequencies and percentages that correspond to the research questions. Corresponding tables and figures have been provided both in text and in the Appendices, depending on the importance of the information being communicated.
CHAPTER FOUR

RESULTS

Question 1:

Q1) How do researchers and the information professionals who serve them understand the term “data management”?

The results of the surveys for the researchers and information professionals provided a deeper look into these communities’ understanding of the term “data management”. Q8 of the Researcher Survey and Q3 of the Information Professionals Survey asked participants if they were familiar with the term “data management”. In the tables below, over 85% of both respondent samples were familiar with the term “data management”.

Table 4.1

Researcher familiarity with term “data management”

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>41</td>
<td>87%</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>6</td>
<td>13%</td>
</tr>
<tr>
<td>3</td>
<td>Other</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>47</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note. Question 8 of Researcher Survey*
The survey then asked respondents how they would define “data management” in an open-ended response box. This question was selected to determine if both researchers and information professionals have similar understandings of such a broad term. The results of this question have been coded based on the frequency of certain terms or phrases used in each definition. As Figure 4.1 shows, of the 121 responses to Q3a of the Information Professional Survey, 71% of respondents included storing/preserving data in their definition and 65% included sharing and accessibility of data.
There was not as much emphasis on categories such as metadata (only 38% included), cleaning data (only 6% included), and keeping data secure (only 8% included). Additionally, only 22 respondents (18%) referenced the data lifecycle in their definition, which will be a foundation for the CRADLE curriculum.

Of the respondents in the Researcher Survey, 35 answered Q8a, which asked researchers: “What does ‘data management’ mean to you?” The responses were similar to those of the information professionals, however, sharing/accessibility of data ranked higher with 24 respondents (69%) including the phrase in their definition. Figure 4.2 shows the emphasis on storing/preserving data, sharing/accessibility of data, and organizing data.

![Q8a - Researchers](image)

*Figure 4.2. What does 'data management' mean to researchers? Q8a of Researcher Survey.*

As with the information professionals, concepts such as data security, cleaning data, and metadata/documentation were not included in researchers’ definitions of data management. The phrase ‘data lifecycle’ was not mentioned in any responses and, therefore, did not make it into the selected codes for researchers.
Question 8b of the Researcher Survey asked respondents who were unsure of the term ‘data management’: “What do you think ‘data management’ means?” Of the five responses, 80% of respondents offered a definition that referenced storing/sharing data and working with data. Only one respondent answered with ‘Don’t Know’. Only one information professional responded to Q3b, which asked the same question as Q8b of the Researcher Survey, however, their answer was not related to the term ‘data management’.

Since the data is being created and produced by researchers and data management hinges upon researcher cooperation, it is important to determine researchers’ attitudes toward data management. During the survey design, I decided that using a more generic phrase such as ‘data storage and sharing’ instead of ‘data management’ would decrease any possible alienation factor for questions targeted at researchers. This decision was based upon the recommendations of the CRADLE project team since the term ‘data management’ can sometimes hold negative connotations for researchers. As such, Q6 of the Researcher Survey asked respondents: “Do you find data storage and sharing beneficial?” Responses to these questions would then prompt participants to answer why they found it beneficial/non-beneficial. Overall, 89% of the 47 participants who responded felt that data storage and sharing is beneficial. One respondent took it a step further and stated that it is, “important, but it’s tedious and time-consuming”. This aligns with the author’s personal experiences with researchers and highlights the need for education and training to aid in making data management less “tedious and time-consuming”.

Expanding upon these current understandings of data management will assist both the researcher and information professional communities in integrating data management
tasks into their current research routines and data management services. Additionally, understanding what these communities are currently doing in terms of data management practices will also aid in training and education. The surveys for both communities asked questions about data management practices and tools being used by researchers and information professionals.

Within the Researcher Survey, Q3 asked participants: “Where do you typically store your data during the active phase of your research projects?” Respondents were then able to select the best choice that described their storage activities. As referenced in Table 4.3, of the choices available, 29% of respondents selected ‘Work computer hard drive’, followed by 18% selecting ‘Cloud drive’.

Table 4.3

Typical storage space during active research

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Personal computer hard drive</td>
<td>7</td>
<td>14%</td>
</tr>
<tr>
<td>2</td>
<td>Work computer hard drive</td>
<td>14</td>
<td>29%</td>
</tr>
<tr>
<td>3</td>
<td>USB drive or DVD/CD-ROM</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>4</td>
<td>Cloud drive</td>
<td>9</td>
<td>18%</td>
</tr>
<tr>
<td>5</td>
<td>Department server</td>
<td>4</td>
<td>8%</td>
</tr>
<tr>
<td>6</td>
<td>Digital repository</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>7</td>
<td>Other</td>
<td>8</td>
<td>16%</td>
</tr>
<tr>
<td>8</td>
<td>Conference campus server</td>
<td>4</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>49</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Note. Q3 of Researcher Survey

The ‘Other’ category responses consisted of 16% of researchers stating that they usually used more than one of the above choices during the active phase of their research projects, so selecting just one was not an option. One respondent selected ‘Digital
repository’ and answered Q3a of the Researcher Survey which asked: “Does the digital repository you use to store your data provide free, public access to your data?”

Interestingly, the response was ‘No’. So, while the researcher is taking the initiative to actively store their data in a digital repository, the digital repository is closed to the general public. This could be due to security concerns or privacy issues, but at least the researcher is making a step in the right direction in terms of funding agency requirements\(^{14}\).

In regards to sharing data, Q5 of the Researcher Survey asked researchers to select from a list the methods they use to share their own research data. We can see below in Table 4.5 that over half (57\%) of the respondents prefer to share via e-mail.

Table 4.5

*How do you share your data?*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Data requests via e-mail</td>
<td>28</td>
<td>57%</td>
</tr>
<tr>
<td>2</td>
<td>Digital repository</td>
<td>11</td>
<td>22%</td>
</tr>
<tr>
<td>3</td>
<td>University library</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>Website</td>
<td>7</td>
<td>14%</td>
</tr>
<tr>
<td>5</td>
<td>Hard copy available at research facility</td>
<td>4</td>
<td>8%</td>
</tr>
<tr>
<td>6</td>
<td>I do not share my data</td>
<td>14</td>
<td>29%</td>
</tr>
<tr>
<td>7</td>
<td>Other (Please specify):</td>
<td>8</td>
<td>16%</td>
</tr>
</tbody>
</table>

*Note. Q5 of Researcher Survey*

Out of the 49 responses, 29\% selected ‘I do not share my data’. It is unclear as to whether this is due to security, privacy, or personal reasons. Eleven respondents selected that they share their data through a digital repository; however, these respondents did not select

\(^{14}\) See Table 4.4 in Appendix A: This table shows responses to Q4a, which asked: “Which of the following requires you to store and share your data?”.
digital repositories as their typical storage space for their data when asked in Q3. It is possible that these researchers may use digital repositories after the initial data collection and analysis have been completed.

Since most information professionals are not involved with the actual research process itself, the above question was rephrased to assess the data management services currently being offered by the institutions in which the information professionals work. Question 4 asked respondents if their library or institution offers data management services. Of the 129 responses to this question, 71% answered with ‘Yes’. In order to dig deeper, the respondents were prompted to select from a list of services that their library or institute currently provide.

Table 4.6

**Data management services offered by library/institution**

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Long-term data storage</td>
<td>65</td>
<td>71</td>
</tr>
<tr>
<td>2</td>
<td>Data sharing</td>
<td>53</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>Data preservation</td>
<td>60</td>
<td>66</td>
</tr>
<tr>
<td>4</td>
<td>Data collection</td>
<td>37</td>
<td>41</td>
</tr>
<tr>
<td>5</td>
<td>Processing data (including description of data using metadata)</td>
<td>52</td>
<td>57</td>
</tr>
<tr>
<td>6</td>
<td>Library instruction about data management</td>
<td>69</td>
<td>76</td>
</tr>
<tr>
<td>7</td>
<td>Creating data management plans</td>
<td>68</td>
<td>75</td>
</tr>
<tr>
<td>8</td>
<td>Other</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>Cleaning sensitive data</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

Note. Q4a of Information Professional Survey

As Table 4.6 shows, a high percentage of respondents selected long-term data storage (71%), library instruction about data management (76%), and creating data management plans (75%) as part of their current data management services. The responses to this question could indicate a skewed respondent pool since only individuals working in libraries offering data management services may have answered.
The survey protocols also included a set of questions for both researchers and information professionals about seeking data for secondary analysis or patrons. In Q10a of the Researcher Survey, respondents were asked to identify locations or services they used to locate existing data online. The purpose of this question was to determine if digital repositories or libraries even factor into researchers’ search processes. Out of the 40 responses, 83% selected online search engines, 68% online databases, and 63% selected online digital repositories. Information professionals were asked the same question, it was targeted at locating existing data for patrons online. Of the 92 responses, 91% go to online databases, 87% use online search engines, and 80% use digital repositories.

**Question 2:**

*Q2) What are some of the gaps in data management knowledge of the social science researcher and information professional communities that need to be addressed?*

The next section of the surveys was formatted to determine what gaps in data management knowledge need to be addressed by the CRADLE curriculum. Since a majority of the respondents were somewhat familiar with data management, it was decided that specifics were needed to target these gaps. As such, Q11 of the Researcher Survey and Q7 of the Information Professional Survey asked respondents to select the resources available in regards to data management education. Participants were able to select all choices that applied.

---

15 See Table 4.7 in Appendix A.
16 See Table 4.8 in Appendix A.
Within the information professional respondents, 126 answered this question. Table 4.9 reveals that 89% (112 respondents) have university librarians available for data management education, 61% offer library instruction, and 58% have various data management tools and training guides. Additionally, those respondents who answered with ‘Other’ (12%), referred to websites, online tutorials, as well as student assistants.

Table 4.9

*Resources available to learn about data management*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>University librarians</td>
<td>112</td>
<td>89%</td>
</tr>
<tr>
<td>2</td>
<td>Library instruction</td>
<td>77</td>
<td>61%</td>
</tr>
<tr>
<td>3</td>
<td>Various data management tools</td>
<td>73</td>
<td>58%</td>
</tr>
<tr>
<td>4</td>
<td>Training guides</td>
<td>73</td>
<td>58%</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
<td>15</td>
<td>12%</td>
</tr>
</tbody>
</table>

*Note. Q7 of Information Professional Survey*

Forty-one researchers answered Q11, which asked: “What resources do you already have available in regards to learning about data management”? The results showed that 49% of the respondents have various data management tools available, while 39% refer to university librarians. Additionally, of the 34% that responded with ‘Other,’ 2 respondents refer to their colleagues and 7 respondents have experience or refer to various departments within their institution.

Question 12 of the Researcher Survey and Q8 of the Information Professional Survey asked participants if they would be interested in receiving training in data storage and sharing. This question was added to the survey to determine motivation and willingness to participate in future education/training. Within the researchers’ sample, 46 responded with 57% stating they would be interested in training and 43% stating they
would not be interested in training. In the information professional’s sample, 125 responded with 66% stating they would like additional training, 22% stating they were not interested in additional training, and 12% uncertain or not needing additional training.

Finally, Q14 of the Researcher Survey and Q11 of the Information Professional Survey asked respondents to identify the topics they would like covered in a data management course. Respondents were able to select all that applied and also enter their own suggestions in the ‘Other’ text entry. The responses to these questions will be used by the CRADLE project to ensure highly-requested topics are fully addressed in the CRADLE curriculum.

Out of the 52 researchers, 42 responded to this question. Table 4.10 shows the responses with 90% of respondents wanting training on the tools available for data management and 71% wanting training on long-term data storage best practices.

Table 4.10

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Long-term data storage best practices</td>
<td>30</td>
<td>71%</td>
</tr>
<tr>
<td>2</td>
<td>Dealing with sensitive data</td>
<td>21</td>
<td>50%</td>
</tr>
<tr>
<td>3</td>
<td>Sharing data efficiently</td>
<td>29</td>
<td>69%</td>
</tr>
<tr>
<td>4</td>
<td>Developing data storage and sharing policies</td>
<td>23</td>
<td>55%</td>
</tr>
<tr>
<td>5</td>
<td>Tools available for data management</td>
<td>38</td>
<td>90%</td>
</tr>
<tr>
<td>6</td>
<td>Who to contact about data management</td>
<td>16</td>
<td>38%</td>
</tr>
<tr>
<td>7</td>
<td>Data curation throughout the lifecycle of the research project</td>
<td>24</td>
<td>57%</td>
</tr>
<tr>
<td>8</td>
<td>What to do with my data after the life of the research project</td>
<td>20</td>
<td>48%</td>
</tr>
<tr>
<td>9</td>
<td>Other</td>
<td>3</td>
<td>7%</td>
</tr>
</tbody>
</table>

*Note. Q14 of the Researcher Survey*
Table 4.11

Information professionals’ preference for topics to be covered by data management course

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Long-term data storage best practices</td>
<td>95</td>
<td>81%</td>
</tr>
<tr>
<td>2</td>
<td>Dealing with sensitive variables</td>
<td>81</td>
<td>52%</td>
</tr>
<tr>
<td>3</td>
<td>Sharing data efficiently</td>
<td>84</td>
<td>71%</td>
</tr>
<tr>
<td>4</td>
<td>Developing data storage and sharing policies</td>
<td>79</td>
<td>67%</td>
</tr>
<tr>
<td>5</td>
<td>Tools available for data management</td>
<td>97</td>
<td>82%</td>
</tr>
<tr>
<td>6</td>
<td>Who to contact about data management</td>
<td>47</td>
<td>40%</td>
</tr>
<tr>
<td>7</td>
<td>Data curation throughout the lifecycle of the project</td>
<td>88</td>
<td>75%</td>
</tr>
<tr>
<td>8</td>
<td>What do I do with my data after the project?</td>
<td>88</td>
<td>58%</td>
</tr>
<tr>
<td>9</td>
<td>Other:</td>
<td>11</td>
<td>9%</td>
</tr>
</tbody>
</table>

Note. Q11 of the Information Professional Survey

Out of the 156 information professionals that participated in this survey, 118 responded to this question with 82% also wanting training on tools available in data management and 81% wanting training on long-term data storage best practices. Table 4.11 shows the rest of the responses and also the emphasis placed on lifecycle education and effective data sharing.

**Question 3:**

*Q3) What preferences do researchers and information professionals have in regards to data management training and education dissemination methods?*

The final question is necessary for the CRADLE team and future data management courses to create a well-received curriculum using educational platforms and methods that appeal to both the researcher and information professional communities. The following survey questions addressed interests in MOOCs, preferred
training/educational platforms, and preferred delivery methods for online education. Participants were also asked how much time they would be willing to dedicate per week to education/training and how long they would be able to remain active in an online course.

In Q12.1 and Q12.1a, researchers were asked what types of educational delivery platforms appealed to them most. If respondents selected MOOCs, they were then asked how long they would prefer a MOOC to run. Out of the 43 respondents, 65% prefer short-course videos, 53% workshops or seminars, and 33% prefer a MOOC. Of the 33% who prefer a MOOC, 79% prefer that the MOOC be less than 4 weeks long\(^{17}\). Q9 and Q9a of the Information Professional Survey asked the same question about preferences for education delivery platforms and MOOC course length. Of the information professional respondents, 120 answered Q9 with 78% preferring workshops and seminars, 70% preferring short-course videos, and 35% preferring a MOOC. Additionally, of the 35% who preferred MOOCs, 48% would prefer the MOOC to be less than 4 weeks and 24% would prefer it to be 4 weeks\(^{18}\).

Participants were then asked to rank online education delivery methods in order of preference. For Q13 of the Researcher Survey, 40 respondents answered this question. Figure 4.3 shows their ranked preferences based on the number of responses for each item and their most selected rankings. For Q10 of the Information Professional Survey, 111 participants answered the question. The results have been ranked and are shown below in Figure 4.4. It is interesting to note that these preferences did not deviate much

\(^{17}\) See Table 4.12a and Table 4.12b in Appendix A

\(^{18}\) See Table 4.13a and Table 4.13b in Appendix A
from the original order of preferences. The final order could be a result of the actual type of question, rather than the actual preferences of the respondents.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3-5 minute short videos</td>
</tr>
<tr>
<td>2</td>
<td>6-10 minute lecture videos</td>
</tr>
<tr>
<td>3</td>
<td>Discussion questions in forum</td>
</tr>
<tr>
<td>4</td>
<td>Hands-on activities</td>
</tr>
<tr>
<td>5</td>
<td>Networking activities</td>
</tr>
<tr>
<td>6</td>
<td>Guest lectures</td>
</tr>
<tr>
<td>7</td>
<td>Short answer questions</td>
</tr>
<tr>
<td>8</td>
<td>Multiple choice questions</td>
</tr>
<tr>
<td>9</td>
<td>PowerPoint/Slide presentations</td>
</tr>
<tr>
<td>10</td>
<td>Other</td>
</tr>
</tbody>
</table>

*Figure 4.3. Ranking order of online delivery methods preferences. Q13 of Researcher Survey.*

<table>
<thead>
<tr>
<th>Rank</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3-5 minute short videos</td>
</tr>
<tr>
<td>2</td>
<td>6-10 minute lecture videos</td>
</tr>
<tr>
<td>3</td>
<td>Hands-on activities</td>
</tr>
<tr>
<td>4</td>
<td>Discussion questions in forum</td>
</tr>
<tr>
<td>5</td>
<td>Guest lectures</td>
</tr>
<tr>
<td>6</td>
<td>Networking activities</td>
</tr>
<tr>
<td>7</td>
<td>Short answer questions</td>
</tr>
<tr>
<td>8</td>
<td>Multiple choice questions</td>
</tr>
<tr>
<td>9</td>
<td>PowerPoint/Slide presentations</td>
</tr>
<tr>
<td>10</td>
<td>Other</td>
</tr>
</tbody>
</table>

*Figure 4.4. Ranking order of online delivery methods preferences of Information Professionals. Q10 of Information Professional Survey.*

Finally, it was necessary to gauge researcher and information professionals’ available time commitments to data management training. This information will be used during the curriculum development for homework, quiz, and assignment design. Question 16 of the Researcher Survey asked 43 respondents how much time they would be able to dedicate to additional education/training. Of those respondents, 77% would be able to
dédicate less than two hours a week, with 19% selecting ‘Other’ and emphasizing very limited time available around less than an hour or two per semester\textsuperscript{19}. Question 12 of the Information Professional Survey asked the same question of 121 respondents and received somewhat similar results with 71% being able to dedicate less than two hours a week to additional education/training, however, 20% of respondents stated they would be able to dedicate 3 to 6 hours a week to additional education/training\textsuperscript{20}.

\textsuperscript{19} See Table 4.14 in Appendix A
\textsuperscript{20} See Table 4.15 in Appendix A
CHAPTER FIVE
DISCUSSION & CONCLUSION

The purpose of this study is to assess the data management education needs and educational delivery preferences of the social science research community and the information professional community. This assessment will allow the CRADLE Project team to create a well-balanced and informative educational curriculum for data management that will target key areas that have been deemed important and/or by these communities. The surveys used in this study address three main research questions:

Q1) How do researchers and the information professionals who serve them understand the term “data management”? 

Q2) What are some of the gaps in data management knowledge of the social science researcher and information professional communities that need to be addressed? 

Q3) What preferences do social science researchers and information professionals have in regards to data management training and education dissemination methods? 

The quantitative data and text responses from these surveys provide the CRADLE Project team valuable insight into these questions. The findings will be useful for development of the curriculum and implementation of the educational delivery methods.
In order to keep the findings clear and concise, they will be broken down by question in this discussion section.

**Q1 Findings**

The responses that answer the first research question offer insight into the current understanding of the term “data management” and the practices of the researcher and information professional communities. From both survey’s results, it is evident that a majority of participants have a broad understanding of the term “data management.” On the researcher side, more than half of the respondents claim familiarity with the term data management, however, their understanding mainly includes data storage and sharing as part of the definition for data management. The same can be said for the information professionals. Concepts such as keeping data secure, cleaning data, and making data discoverable via documentation and metadata are not as highly referenced by both communities. This finding aligns with previous research conducted by Christine Borgman (2012) and highlights how researchers and information professionals are being tasked by funding agencies to manage data, however they still lack a deeper understanding of the term and seem to focus on the four rationales that Borgman introduces (p. 1072). The main idea behind these questions is to use the results to inform the CRADLE curriculum and future data management courses. As such, concepts such as data security, metadata, and cleaning data will be emphasized in the hopes that it will fill some of the gaps in these communities’ definitions of “data management.”

Questions about the benefits of data management are also necessary to understand the mindset of researchers when approaching this topic. The results show that most
respondents find data storage and sharing beneficial. Respondents were able to elaborate on why they found it beneficial and a majority of the responses emphasized personal re-use and collaboration. This gives a better understanding of how researchers perceive data management. For the most part, it seems that the respondents are thinking of data storage and sharing in terms of collaborative efforts and personal use only. The idea that others may wish to access their data is not as important, or rather, not in the forefront of their minds when thinking of potential benefits. Christine Borgman’s (2012) research also touched upon this concept stating that, “researchers…cannot imagine who might use them [data]” (p 1073). It is important to use this data to ensure that an emphasis on secondary analysis, collaboration between disciplines, and even educational re-use are made evident as benefits and motivators for data storage and sharing.

Since the CRADLE Project curriculum will educate researchers and information professionals on best practices in data management, the survey protocols were tailored to ask respondents where they store their data during research, how they share their data, and where they go to find existing data. Researchers’ responses indicate that they usually store their data on their work or personal computer hard drive. Very few actually use digital repositories and even fewer share their data unless it is for collaboration or requested directly. These findings align with the author’s personal experience with researchers during data archiving and with the findings of previous studies (Vertesi and Dourish, 2011; Doty and Akers, 2013).

Interestingly, when both communities were asked about finding existing data, the majority of respondents seem to use online databases and search engines before even considering digital repositories. It is unclear if this is because there is a lack in knowledge
of available repositories or if online search engines and databases are just easier to navigate, however, it will be useful in developing a list of resources for participants of the CRADLE educational course. It may also be useful to conduct further investigation into the usability of digital repositories versus online search engines and databases. This information could aid CRADLE in developing educational tools for using digital repositories, databases, and online search engines.

**Q2 Findings**

The training and education available to researchers and information professionals ranges from online tutorials and websites to subject specialists and library instruction. For researchers trying to learn more about data management, this could seem overwhelming and thus result in hesitancy or minimal investigation. For information professionals, especially those not as familiar with data sciences, it may also seem like a large undertaking. The purpose of getting answers to Q2 is to determine where these communities go to find or provide information about data management. From the responses to these survey questions, it is clear that most respondents of the Researcher Survey utilize statistical packages as data management tools for education before considering information professionals. Information professionals emphasize university librarians as instructors for data management education. This could be a result in a lack of communication between the two communities. It is possible that researchers do not know there are university librarians available for consultation and education in data management. This could also mean that university librarians are not trained in data management services at a researcher’s institution. Either way, it is important to note this
information when developing networking activities for the CRADLE courses. Creating an interconnected network of researchers and information professionals will help bridge any communication gaps and foster stronger relationships during the data lifecycle.

Additionally, questions within the surveys asked respondents to specify what topics in data management they would like to have included in training and educational courses. From the results, the Top 5 Topics that need to be emphasized are:

- Data management tools
- Long-term data storage best practices
- Sharing data efficiently
- Data curation throughout the lifecycle of the research project
- Developing data storage and sharing policies

The responses to this question from both surveys are aligned and remarkable in that both tables emphasize the same categories in the same order. These topics are important for both researchers and information professionals and they give each item the same priority. This is also true for the rest of the results from this question. Each category is ranked the same between both surveys. With this in mind, these topics can be building blocks for the CRADLE project and future data management curricula to ensure that each is fully developed.

**Q3 Findings**

The third and final question is addressed by the last section of the surveys. The responses from these questions relate to MOOCs, educational platforms, and educational delivery methods. The CRADLE Project, based on the results of this survey, will be
creating webinars and short online instructional videos as part of their curriculum. Additionally, a MOOC will be developed to disseminate data management education across a larger population. Since this is a major product of the CRADLE Project, it is important that we gauge researcher and information professionals’ preferences towards educational platforms. The findings show that most respondents’ educational preferences lean toward short video courses or workshops and seminars. Since MOOCs are still relatively new, it is interesting to note that there is still a good amount of interest in a MOOC course.

A preference for short video courses can be used by the CRADLE Project when developing the videos for the MOOC. For instance, instead of longer videos per unit, the CRADLE team can use shorter, targeted videos to address specific topics of interest in smaller amounts of time. This will appeal to the majority of participants based on the findings of this study. Also, using the rankings created from the surveys, the CRADLE team can emphasize these educational delivery preferences in assignments, lectures, and presentations. Most respondents prefer shorter videos with more hands-on activities and discussions within forums. This is a great opportunity to develop stronger networks across communities and gain practical experience with data management exercises.

Furthermore, in order to plan the length of a MOOC and the time required to successfully complete the MOOC, the CRADLE team will be able to refer to the responses to the later questions in these surveys. The majority of participants will not be able to dedicate a lot of time to additional training; therefore, it is important that assignments be designed so that completion times do not exceed more than an hour. Also, since this MOOC will cater to students, hobbyists, and working professionals, the length
of the course will need to reflect the lifestyles of the majority of its participants. Most researchers and information professionals cannot participate in a course that is 6-8 weeks long. A shorter course may be beneficial for the professional workforce, or perhaps even holding the MOOC in-between semesters or during the summer would be even convenient.

Limitations

The low response rate is a limitation of this study. With almost 2,000 participants invited to partake in this survey, it would be beneficial to conduct a similar survey to see if a higher response rate can be achieved. Furthermore, due to the time constraints, individual follow-up interviews could not be conducted. These interviews may possibly aid in addressing questions regarding ranking order and educational preferences. Why did a researcher or information professional emphasize education on long-term data storage over data security? Obtaining more detailed information could aid the CRADLE team and any future data management courses in understanding respondents’ reasoning. Additionally, the interviews could provide insights that were not addressed by the questions in the surveys. Interviewees could give more personalized responses on the topic of data management. This feedback could potentially influence the final curriculum of the new courses.

Initially this study was also going to conduct focus groups with researchers and information professionals. Due to scheduling issues and the winter break, the focus groups could not be completed, however, the CRADLE Project will be using the results from this study to influence focus group protocols and will conduct these groups at a later
date. It will be interesting to see if the results from the focus groups correlate with the results of the needs assessment survey.

**Conclusion**

The results from the Researcher Survey and the Information Professional Survey provide insight into the research questions proposed at the beginning of this study. Before developing a curriculum, it is necessary to determine where researchers and information professionals are currently at in terms of understanding data management. Moreover, understanding data management practices of these communities will aid the CRADLE Project team in creating resources and tools that will educate students in the best practices related to their current data management practices. It will also fill the gaps within current practices to increase awareness of all of the necessary aspects of data management.

While MOOCs may not have ranked the highest choice for educational platforms, the high preference towards online short course videos indicates that MOOCs could utilize this format to reach a wider audience of participants. Uniting short videos with guest lectures, hands-on activities, and discussions in the MOOC forum will create a unique learning experience for researchers and information professionals. And knowing what types of educational delivery methods to avoid, such as PowerPoint presentations, will make sure that the CRADLE MOOC retains more of its students by keeping them engaged in the videos and activities.
BIBLIOGRAPHY


Dally, K., et al. (2012). The data management rollout at Oxford project (DaMaRO). Retrieved from http://damaro.oucs.ox.ac.uk/outputs.xml


Janée, G., & Frew, J. (2013). Faculty/researcher survey on data curation. doi:10.5062/F4PN93K4


## APPENDIX A

### TABLES

Table 3.1

*Researcher professions*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tenure-track Faculty</td>
<td>18</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>Tenured Faculty</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>Fixed-term Faculty</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Adjunct Faculty</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Post Doctoral Researcher</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Doctoral Student</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Master's Student</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>University Staff</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>State Agency Staff</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Private Agency Staff</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Other</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

*Note.* Q1 of Researcher Survey
Table 3.2

*Information professional professions*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tenure-track Faculty Librarian</td>
<td>29</td>
<td>20%</td>
</tr>
<tr>
<td>2</td>
<td>Tenured Faculty Librarian</td>
<td>28</td>
<td>20%</td>
</tr>
<tr>
<td>3</td>
<td>Non-Tenure Librarian</td>
<td>65</td>
<td>46%</td>
</tr>
<tr>
<td>4</td>
<td>Archivist</td>
<td>3</td>
<td>2%</td>
</tr>
<tr>
<td>5</td>
<td>Post Doctoral Researcher</td>
<td>1</td>
<td>1%</td>
</tr>
<tr>
<td>6</td>
<td>Doctoral Student</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>7</td>
<td>Master's Student</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>8</td>
<td>University Staff</td>
<td>9</td>
<td>6%</td>
</tr>
<tr>
<td>9</td>
<td>State Agency Librarian</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>10</td>
<td>Private Agency Librarian</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>11</td>
<td>Other:</td>
<td>7</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>142</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note.* Q1 of Information Professional Survey
### Table 3.3

**Researcher subject specialty**

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sociology</td>
<td>15</td>
<td>31%</td>
</tr>
<tr>
<td>2</td>
<td>Political Science</td>
<td>4</td>
<td>8%</td>
</tr>
<tr>
<td>3</td>
<td>Library Science</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>Information Science</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>Psychology</td>
<td>16</td>
<td>33%</td>
</tr>
<tr>
<td>6</td>
<td>Anthropology</td>
<td>6</td>
<td>12%</td>
</tr>
<tr>
<td>7</td>
<td>Economics</td>
<td>3</td>
<td>6%</td>
</tr>
<tr>
<td>8</td>
<td>Education</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>9</td>
<td>Communication</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>10</td>
<td>Law</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>11</td>
<td>History</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>12</td>
<td>Linguistics</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>13</td>
<td>Other</td>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>49</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note. Q2 of Researcher Survey*

### Table 4.4

**Who requires you to share your data?**

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Funding agency</td>
<td>11</td>
<td>85%</td>
</tr>
<tr>
<td>2</td>
<td>Journal publisher</td>
<td>4</td>
<td>31%</td>
</tr>
<tr>
<td>3</td>
<td>Department</td>
<td>1</td>
<td>8%</td>
</tr>
<tr>
<td>4</td>
<td>Supervisor</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>Institution</td>
<td>3</td>
<td>23%</td>
</tr>
<tr>
<td>6</td>
<td>Other (Please specify)</td>
<td>2</td>
<td>15%</td>
</tr>
</tbody>
</table>

*Note. Q4a of Researcher Survey*
Table 4.7

Where do you search for existing data?

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Online search engine (Google, Bing, Yahoo, etc.)</td>
<td>33</td>
<td>83%</td>
</tr>
<tr>
<td>2</td>
<td>University library website</td>
<td>15</td>
<td>38%</td>
</tr>
<tr>
<td>3</td>
<td>Public library website</td>
<td>4</td>
<td>10%</td>
</tr>
<tr>
<td>4</td>
<td>Online database</td>
<td>27</td>
<td>68%</td>
</tr>
<tr>
<td>5</td>
<td>Online digital repository</td>
<td>25</td>
<td>63%</td>
</tr>
<tr>
<td>6</td>
<td>Other</td>
<td>5</td>
<td>13%</td>
</tr>
</tbody>
</table>

*Note. Q10a of Researcher Survey*

Table 4.8

Where do you search for existing data for patrons?

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Online search engine (Google, Bing, Yahoo, etc.)</td>
<td>80</td>
<td>87%</td>
</tr>
<tr>
<td>2</td>
<td>University library</td>
<td>71</td>
<td>77%</td>
</tr>
<tr>
<td>3</td>
<td>Public library</td>
<td>9</td>
<td>10%</td>
</tr>
<tr>
<td>4</td>
<td>Online database</td>
<td>84</td>
<td>91%</td>
</tr>
<tr>
<td>5</td>
<td>Digital repository</td>
<td>74</td>
<td>80%</td>
</tr>
<tr>
<td>6</td>
<td>Other</td>
<td>19</td>
<td>21%</td>
</tr>
</tbody>
</table>

*Note. Q6a of Information Professional Survey*
Table 4.12a

*Researchers' preference for educational delivery platforms*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MOOC (Massive Open Online Course)</td>
<td>14</td>
<td>33%</td>
</tr>
<tr>
<td>2</td>
<td>Workshops or Seminars</td>
<td>23</td>
<td>53%</td>
</tr>
<tr>
<td>3</td>
<td>Training pamphlets</td>
<td>15</td>
<td>35%</td>
</tr>
<tr>
<td>4</td>
<td>Short-Course Videos</td>
<td>28</td>
<td>65%</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
<td>5</td>
<td>12%</td>
</tr>
</tbody>
</table>

*Note. Q12.1 of Researcher Survey*

Table 4.12b

*Researchers' preference for length of MOOC*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 weeks</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>6 weeks</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>8 weeks</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>10 weeks</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>12 weeks</td>
<td>1</td>
<td>7%</td>
</tr>
<tr>
<td>6</td>
<td>Other</td>
<td>2</td>
<td>14%</td>
</tr>
<tr>
<td>7</td>
<td>Fewer than 4 weeks</td>
<td>11</td>
<td>79%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>14</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note. Q12.1a of Researcher Survey*
Table 4.13a

*Information professionals' preference for educational delivery platforms*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MOOC (Massive Open Online Course)</td>
<td>42</td>
<td>35%</td>
</tr>
<tr>
<td>2</td>
<td>Workshops or Seminars</td>
<td>94</td>
<td>78%</td>
</tr>
<tr>
<td>3</td>
<td>Training pamphlets</td>
<td>30</td>
<td>25%</td>
</tr>
<tr>
<td>4</td>
<td>Short-Course videos</td>
<td>84</td>
<td>70%</td>
</tr>
<tr>
<td>5</td>
<td>Other</td>
<td>10</td>
<td>8%</td>
</tr>
</tbody>
</table>

*Note. Q9 of Information Professional Survey*

Table 4.13b

*Information professionals' preference for length of MOOC*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 weeks</td>
<td>10</td>
<td>24%</td>
</tr>
<tr>
<td>2</td>
<td>6 weeks</td>
<td>8</td>
<td>19%</td>
</tr>
<tr>
<td>3</td>
<td>8 weeks</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td>10 weeks</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>12 weeks</td>
<td>1</td>
<td>2%</td>
</tr>
<tr>
<td>6</td>
<td>Other</td>
<td>3</td>
<td>7%</td>
</tr>
<tr>
<td>7</td>
<td>Fewer than 4 weeks</td>
<td>20</td>
<td>48%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>42</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note. Q9a of Information Professional Survey*
Table 4.14

*Amount of time researchers can dedicate to training/education*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>less than two hours a week</em></td>
<td>33</td>
<td>77%</td>
</tr>
<tr>
<td>2</td>
<td>3 - 6 hours a week</td>
<td>2</td>
<td>5%</td>
</tr>
<tr>
<td>3</td>
<td><em>More than 6 hours a week</em></td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td><em>Other</em></td>
<td>8</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>43</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note. Q16 of Researcher Survey*

Table 4.15

*Amount of time information professionals can dedicate to training/education*

<table>
<thead>
<tr>
<th>#</th>
<th>Answer</th>
<th>Response</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>less than two hours a week</em></td>
<td>86</td>
<td>71%</td>
</tr>
<tr>
<td>2</td>
<td>3 - 6 hours a week</td>
<td>24</td>
<td>20%</td>
</tr>
<tr>
<td>3</td>
<td><em>More than 6 hours a week</em></td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>4</td>
<td><em>Other</em></td>
<td>11</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>121</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Note. Q12 of Information Professional Survey*
APPENDIX B
SURVEY PROTOCOL

CRADLE - Researchers Survey

Thank you for participating in the CRADLE (Curating Research Assets and Data using Lifecycle Education) survey.

Introduction

The CRADLE Project is led by Dr. Helen Tibbo, Alumni Distinguished Professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill. This survey seeks to identify current attitudes and practices in data storage and sharing. Additionally, this survey will assist in developing a curriculum tailored to educating information professionals about data management best practices, tools, and available resources.

More information about the CRADLE Project is available at: http://cradle.web.unc.edu

All responses will be kept anonymous to ensure your privacy.

If you have any questions about your rights as a research participant, you may contact the UNC Institutional Review Board at IRB_Subjects@unc.edu; please mention study number 13-3347.

Data

For the purpose of this survey, the term data will be defined as, "[digital assets that] are collected, observed, or created, for the purposes of analysis to produce and validate original research results" (Ekmekcioglu, Rice, & Macdonald, 2013).

Citation


Q1 What is your profession? (Please select the response that is the most applicable)

- Tenure-track Faculty (1)
- Tenured Faculty (2)
- Fixed-term Faculty (3)
Q1 Which of the following best describes your role?

- Adjunct Faculty (4)
- Post Doctoral Researcher (5)
- Doctoral Student (6)
- Master's Student (7)
- University Staff (8)
- State Agency Staff (9)
- Private Agency Staff (10)
- Other: (11) ____________________

Q2 Which of the following best describes your area of research?

- Sociology (1)
- Political Science (2)
- Library Science (3)
- Information Science (4)
- Psychology (5)
- Anthropology (6)
- Economics (7)
- Education (8)
- Communication (9)
- Law (10)
- History (11)
- Linguistics (12)
- Other: (13) ____________________
Q3 Where do you typically store your data during the active phase of your research projects?

- Personal computer hard drive (1)
- Work computer hard drive (2)
- USB drive or DVD/CD-ROM (3)
- Cloud drive (4)
- Department server (5)
- Digital repository (6)
- Other: (7) _________________
- Central campus server (8)

Answer If “Where do you store your data during the active phase of your research?” “Digital repository” Is Selected

Q3a Does the digital repository you use to store your data provide free, public access to your data?

- Yes (1)
- No (2)
- Other: (3) _________________

Q4 Are you required to store and share your data?

- Yes (1)
- No (2)
- Other: (3) _________________
Answer If “Are you required to store and share your data?” “Yes” Is Selected

Q4a Which of the following requires you to store and share your data? (Please select all that apply)

- Funding agency (1)
- Journal publisher (2)
- Department (3)
- Supervisor (4)
- Institution (5)
- Other (Please specify): (6) ______________________

Q5 How do you share your data? (Please select all that apply)

- Data requests via e-mail (1)
- Digital repository (2)
- University library (3)
- Website (4)
- Hard copy available at research facility (5)
- I do not share my data (6)
- Other (Please specify): (7) ______________________

Q6 Do you find data storage and sharing beneficial?

- Yes (1)
- No (2)
- Other: (3) ______________________
Answer If “Do you find data storage and sharing beneficial?” “Yes” Is Selected

Q6a Why do you find it beneficial?

Answer If “Do you find data storage and sharing beneficial?” “No” Is Selected

Q6b Why do you find it non-beneficial?

Q7 What typically happens to your data after the project is completed?

- Stored on a USB drive or DVD/CD-ROM (1)
- Stored on Cloud drive (2)
- Stored on personal computer hard drive (3)
- Stored on a department server (4)
- Made available in digital repository (5)
- Data is destroyed (6)
- I do not know (7)
- Other: (8) ____________________
- Stored on a central campus server (9)

Q8 Are you familiar with the term “data management”?

- Yes (1)
- No (2)
- Other: (3) ____________________

Answer If “Are you familiar with the term ‘data management’?” “Yes” Is Selected

Q8a What does "data management" mean to you?

Answer If “Are you familiar with the term ‘data management’?” “No” Is Selected
Q8b What do you think "data management" means?

For the purposes of this survey, we define data management as those tasks involved in the collection, cleaning, analysis, storage, sharing, disposal, and/or archiving of research data. For the remainder of the survey, please refer to this definition.

Q9 Are you required by your institution or funding agency to include data management plans as part of your research proposals?

☐ Yes (1)
☐ No (2)

Q10 Have you ever tried to find existing data online?

☐ Yes (1)
☐ No (2)

Answer If “Have you ever tried to find existing data online?” “Yes” Is Selected

Q10a Where did you go to locate this existing data? (Please select all that apply)

☐ Online search engine (Google, Bing, Yahoo, etc.) (1)
☐ University library website (2)
☐ Public library website (3)
☐ Online database (4)
☐ Online digital repository (5)
☐ Other: (6) ____________________

Q11 What resources do you already have available in regards to learning about data management? (Please select all that apply)

☐ University Librarian (1)
☐ Various data management tools (2)
☐ Training guides (3)
☐ Library instruction (4)
☐ Other: (5) ____________________

*Answer If “What resources do you already have available in regards to learning about data management? (Please select all that apply)” “Various data management tools” Is Selected*

Q11a Please enter the names of the data management tools that are available to you.

*Answer If “What resources do you already have available in regards to learning about data management? (Please select all that apply)” “Training guides” Is Selected*

Q11b Please enter the names of the training guides that are available to you.

Q12 Would you be interested in receiving training in data storage and data sharing (i.e., data management)?

☐ Yes (1)
☐ No (2)

Q12.1 Which of the following type(s) of educational delivery platform appeals to you? (Please select all that apply)

☐ MOOC (Massive Open Online Course) (1)
☐ Workshops or Seminars (2)
☐ Training pamphlets (3)
☐ Short-Course Videos (4)
☐ Other: (5) ____________________
Answer If “What type of educational delivery platform appeals to you? (Please select all that apply)” “MOOC (Massive Open Online Course)” Is Selected

Q12.1a How long would you prefer the MOOC to run?

- 4 weeks (1)
- 6 weeks (2)
- 8 weeks (3)
- 10 weeks (4)
- 12 weeks (5)
- Other: (6) _________________
- Fewer than 4 weeks (7)

Q13 Please rank the following online education delivery methods in order of preference (most preferred method at the top):

- 3-5 minute short videos (1)
- 6-10 minute lecture videos (2)
- Discussion questions in forum (3)
- Hands-on activities (4)
- Networking activities (5)
- Guest lectures (6)
- Short answer questions (7)
- Multiple choice questions (8)
- PowerPoint/slide presentations (9)
- Other: (10)

Q14 What topics would you like to be covered in the data management course? (Please select all that apply)
☐ Long-term data storage best practices (1)
☐ Dealing with sensitive data (2)
☐ Sharing data efficiently (3)
☐ Developing data storage and sharing policies (4)
☐ Tools available for data management (5)
☐ Who to contact about data management (6)
☐ Data curation throughout the lifecycle of the research project (7)
☐ What to do with my data after the life of the research project (8)
☐ Other: (9) ____________________

Q16 How much time would you be able to dedicate to additional education/training?
☐ less than two hours a week (1)
☐ 3 - 6 hours a week (2)
☐ More than 6 hours a week (3)
☐ Other: (4) ____________________

Q17 Is there anything else you would like to discuss about this topic?

Q18 Would you be available for a possible follow-up interview in the future?
☐ Yes (1)
☐ No (2)

Answer If “Would you be available for a possible follow-up interview in the future?” “Yes” Is Selected

Q18a What is the best e-mail to reach you at?:

Answer: ____________________
Thank you for participating in the CRADLE (Curating Research Assets and Data using Lifecycle Education) survey.

Introduction

The CRADLE Project is led by Dr. Helen Tibbo, Alumni Distinguished Professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill. This survey seeks to identify current attitudes and practices in data storage and sharing. Additionally, this survey will assist in developing a curriculum tailored to educating information professionals about data management best practices, tools, and available resources.

More information about the CRADLE Project is available at: http://cradle.web.unc.edu

All responses will be kept anonymous to ensure your privacy.

If you have any questions about your rights as a research participant, you may contact the UNC Institutional Review Board at IRB_Subjects@unc.edu; please mention study number 13-3347.

Data

For the purpose of this survey, the term data will be defined as, "[digital assets that] are collected, observed, or created, for the purposes of analysis to produce and validate original research results" (Ekmeckioglu, Rice, & Macdonald, 2013).

Citation

Q1 What is your profession? (Please select the response that is the most applicable)

- Tenure-track Faculty Librarian (1)
- Tenured Faculty Librarian (2)
- Non-Tenure Librarian (3)
- Archivist (4)
- Post Doctoral Researcher (5)
- Doctoral Student (6)
- Master's Student (7)
- University Staff (8)
- State Agency Librarian (9)
- Private Agency Librarian (10)
- Other: (11) ____________________

Q2 Are you considered a subject specialist or discipline-specific librarian?

- Yes (1)
- No (2)

Answer If “Are you considered a subject specialist or discipline-specific librarian? “
“Yes” Is Selected

Q2a What is your subject specialty(ies)?

Q3 Are you familiar with the term "data management"?

- Yes (1)
- No (2)
- Other: (3) ____________________
Answer If “Are you familiar with the term ‘data management’?” “Yes” Is Selected

Q3a What does "data management" mean to you?

Answer If “Are you familiar with the term ‘data management’?” “No” Is Selected

Q3b What do you think "data management" means?

For the purposes of this survey, we define data management as those tasks involved in the collection, cleaning, analysis, storage, sharing, disposal, and/or archiving of research data. For the remainder of the survey, please refer to this definition.

Q4 Does your library/institute currently offer data management services?

☐ Yes (1)
☐ No (2)
☐ Other: (3) ____________________

Answer If “Does your library/institute currently offer data management services?” “Yes” Is Selected

Q4a What services does it offer? (Please select all that apply)

☐ Long-term data storage (1)
☐ Data sharing (2)
☐ Data preservation (3)
☐ Data collection (4)
☐ Processing data (including description of data using metadata) (5)
☐ Library instruction about data management (6)
☐ Creating data management plans (7)
☐ Other: (8) ____________________
Cleaning sensitive data (9)

Q5 Does your institution currently have policies in place for data management?

- Yes (1)
- No (2)
- Other: (3) ____________________

*Answer If “Does your institution currently have policies in place for data management?” “Yes” Is Selected*

Q5a Are you familiar with these policies?

- Yes (1)
- No (2)
- Other: (3) ____________________

Q6 Have you ever tried to find existing data online for a patron?

- Yes (1)
- No (2)

*Answer If “Have you ever tried to find existing data online?” “Yes” Is Selected*

Q6a Where did you go to locate this existing data? (Please select all that apply)

- Online search engine (Google, Bing, Yahoo, etc.) (1)
- University library (2)
- Public library (3)
- Online database (4)
- Digital repository (5)
Q7 What resources do you already have available to patrons interested in learning more about data management? (Please select all that apply)

☐ University librarians (1)
☐ Library instruction (2)
☐ Various data management tools (3)
☐ Training guides (4)
☐ Other: (5) ____________________

Answer If “What resources do you already have available to patrons interested in learning more about data management? (Please select all that apply)” “Various data management tools” Is Selected

Q7a Please enter the names of the data management tools that are available to your patrons.

Answer If “What resources do you already have available to patrons interested in learning more about data management? (Please select all that apply)” “Training guides” Is Selected

Q7b Please enter the names of the training guides that are available to your patrons.

Q8 Would you like education in data storage and data sharing (data management)?

☐ Yes (1)
☐ No (2)
☐ Other: (3) ____________________
Q9 What type of educational delivery platform appeals to you? (Please select all that apply)

- MOOC (Massive Open Online Course) (1)
- Workshops or Seminars (2)
- Training pamphlets (3)
- Short-Course Videos (4)
- Other: (5) ____________________

Answer If “What type of educational delivery platform appeals to you? (Please select all that apply)” “MOOC (Massive Open Online Course)” Is Selected

Q9a How long would you like the MOOC to run?

- 4 weeks (1)
- 6 weeks (2)
- 8 weeks (3)
- 10 weeks (4)
- 12 weeks (5)
- Other: (6) ____________________
- Fewer than 4 weeks (7)

Q10 Please rank the following online delivery methods in order of preference (most preferred method at the top):

_____ 3-5 minute short videos (1)
_____ 6-10 minute lecture videos (2)
_____ Discussion questions in forum (3)
_____ Hands-on activities (4)
_____ Networking activities (5)
Q11 What topics would you like to be covered in the data management course? (Please select all that apply)

☐ Long-term data storage best practices (1)

☐ Dealing with sensitive variables (2)

☐ Sharing data efficiently (3)

☐ Developing data storage and sharing policies (4)

☐ Tools available for data management (5)

☐ Who to contact about data management (6)

☐ Data curation throughout the lifecycle of the project (7)

☐ What do I do with my data after the project? (8)

☐ Other: (9) ____________________

Q12 How much time would you be able to dedicate to additional education/training?

☐ less than two hours a week (1)

☐ 3 - 6 hours a week (2)

☐ More than 6 hours a week (3)

☐ Other: (4) ____________________
Q13 Is there anything else you would like to discuss about this topic?

Q14 Would you be available for a possible follow-up interview in the future?

☐ Yes (1)
☒ No (2)

*Answer If “Would you be available for a possible follow-up interview in the future?” “Yes” Is Selected*

Q14a What is the best e-mail to reach you at?:

APPENDIX C

EMAIL INVITATIONS

RESEARCHER INVITATION

Good Morning,

The Curating Research Assets and Data using Lifecycle Education (CRADLE) Project led by Dr. Helen Tibbo, Alumni Distinguished Professor at the University of North Carolina at Chapel Hill's School of Information and Library Science, is seeking social science researchers to participate in a short, 15-minute online survey. This survey seeks to identify current attitudes and practices in data storage and sharing. Additionally, this survey will assist in developing a curriculum tailored to educating social science researchers about data management best practices, tools, and available resources.

The CRADLE project will use this information to inform the development of data management tools and services for members of the social science research community and the information professionals who serve them. The survey is also being conducted in partial fulfillment of the UNC School of Information and Library Science degree requirements.

Results of the survey will never include individually identifiable information; your name will never be associated with your data.

The last day to submit the survey is January 20th, 2014 at 9am EST.

If you have any questions about the survey, please contact Mandy Gooch, CRADLE Fellow, at [email] or Dr. Helen Tibbo, SILS advisor, at [email]. If you have any questions about your rights as a research participant, you may contact the UNC Institutional Review Board at IRB_Subjects@unc.edu; please mention study number 13-3347.

Thank you for your time and contribution to this project.

Follow this link to the Survey:
${l://SurveyLink?d=Take the Survey}$

Or copy and paste the URL below into your internet browser:
${l://SurveyURL}$

Sincerely,

Mandy Gooch
MSLS Candidate 2014, UNC Chapel Hill
Follow the link to opt out of future emails:
${l://OptOutLink?d=Click here to unsubscribe}

INFORMATION PROFESSIONAL INVITATION

Good Morning,

The Curating Research Assets and Data using Lifecycle Education (CRADLE) Project led by Dr. Helen Tibbo, Alumni Distinguished Professor at the University of North Carolina at Chapel Hill's School of Information and Library Science, is seeking information professionals to participate in a short, 15-minute online survey. This survey seeks to identify current attitudes and practices in data storage and sharing. Additionally, this survey will assist in developing a curriculum tailored to educating information professionals about data management best practices, tools, and available resources. The CRADLE project will use this information to inform the development of data management tools and services for members of the social science research community and the information professionals who serve them. The survey is also being conducted in partial fulfillment of the UNC School of Information and Library Science degree requirements.

Results of the survey will never include individually identifiable information; your name will never be associated with your data.

The last day to submit the survey is January 20th, 2014 at 9am EST.

If you have any questions about the survey, please contact Mandy Gooch, CRADLE Fellow, at [email] or Dr. Helen Tibbo, SILS advisor, at [email]. If you have any questions about your rights as a research participant, you may contact the UNC Institutional Review Board at IRB_Subjects@unc.edu; please mention study number 13-3347.

Thank you for your time and contribution to this project.

Follow this link to the Survey:
${l://SurveyLink?d=Take the Survey}
Or copy and paste the URL below into your internet browser:
${l://SurveyURL}

Thank you,
Mandy Gooch
MSLS Candidate 2014, UNC at Chapel Hill
Follow the link to opt out of future emails:
${l://OptOutLink?d=Click here to unsubscribe}

**RESEARCHER REMINDER INVITATION**

Good Morning,

I recently contacted you to participate in the CRADLE Project survey. If you would be willing to complete this short, 15-minute survey, it would greatly assist the CRADLE Team in developing a data management curriculum targeted to the needs of your research community.

I've added the original invitation here:
The Curating Research Assets and Data using Lifecycle Education (CRADLE) Project led by Dr. Helen Tibbo, Alumni Distinguished Professor at the University of North Carolina at Chapel Hill's School of Information and Library Science, is seeking social science researchers to participate in a short, 15-minute online survey. This survey seeks to identify current attitudes and practices in data storage and sharing. Additionally, this survey will assist in developing a curriculum tailored to educating social science researchers about data management best practices, tools, and available resources. The CRADLE project will use this information to inform the development of data management tools and services for members of the social science research community and the information professionals who serve them. The survey is also being conducted in partial fulfillment of the UNC School of Information and Library Science degree requirements.

Results of the survey will never include individually identifiable information; your name will never be associated with your data.

The last day to submit the survey is January 20th, 2014 at 9am EST.

If you have any questions about the survey, please contact Mandy Gooch, CRADLE Fellow, at [email] or Dr. Helen Tibbo, SILS advisor, at [email]. If you have any questions about your rights as a research participant, you may contact the UNC Institutional Review Board at IRB_Subjects@unc.edu; please mention study number 13-3347.

Thank you for your time and contribution to this project.

Follow this link to the Survey:
${l://SurveyLink?d=Take the Survey}
Or copy and paste the URL below into your internet browser:
${l://SurveyURL}
Sincerely,
Mandy Gooch
MSLS Candidate 2014, UNC Chapel Hill

Follow the link to opt out of future emails:
${l://OptOutLink?d=Click here to unsubscribe}

**INFORMATION PROFESSIONAL REMINDER INVITATION**

Good Morning,

I recently contacted you to participate in the CRADLE Project survey. If you would be willing to complete this short, 15-minute survey, it would greatly assist the CRADLE Team in developing a data management curriculum targeted to the needs of your research community.

I've added the original invitation here:

The Curating Research Assets and Data using Lifecycle Education (CRADLE) Project led by Dr. Helen Tibbo, Alumni Distinguished Professor at the University of North Carolina at Chapel Hill's School of Information and Library Science, is seeking information professionals to participate in a short, 15-minute online survey. This survey seeks to identify current attitudes and practices in data storage and sharing. Additionally, this survey will assist in developing a curriculum tailored to educating information professionals about data management best practices, tools, and available resources.

The CRADLE project will use this information to inform the development of data management tools and services for members of the social science research community and the information professionals who serve them. The survey is also being conducted in partial fulfillment of the UNC School of Information and Library Science degree requirements.

Results of the survey will never include individually identifiable information; your name will never be associated with your data.

The last day to submit the survey is January 20th, 2014 at 9am EST.

If you have any questions about the survey, please contact Mandy Gooch, CRADLE Fellow, at [email] or Dr. Helen Tibbo, SILS advisor, at [email]. If you have any questions about your rights as a research participant, you may contact the UNC Institutional Review Board at IRB_Subjects@unc.edu; please mention study number 13-3347.

Thank you for your time and contribution to this project.
Follow this link to the Survey:
${l://SurveyLink?d=Take the Survey}
Or copy and paste the URL below into your internet browser:
${l://SurveyURL}

Thank you,
Mandy Gooch
MSLS Candidate 2014, UNC at Chapel Hill

Follow the link to opt out of future emails:
${l://OptOutLink?d=Click here to unsubscribe}