

This is a peer-reviewed manuscript version. The final version is published in *The American Archivist*, Vol. 77, No. 1 (Spring/Summer 2014) 201–240.

## **Title: Data Curation and the University Archives**

### **INTRODUCTION:**

Walters and Skinner observe in *New Roles for New Times: Digital Curation for Preservation* that “Historically, librarians and archivists have been looked to as the custodians of physical cultural heritage. This focus on content curation—including selection, management, and preservation—is a defining component of research libraries’ ongoing work. A new role within this arena is emerging—that of digital curation....”<sup>1</sup> Over the past decade researchers and others in higher education have become more aware of the need for curation of research data. Lee and Tibbo recently noted, “Efforts to further articulate and define the characteristics of digital curation have come at the same time that the archival profession faces unprecedented opportunities and challenges related to electronic records... Digital curation is a recent (and we believe, useful) phrase to designate a set of opportunities for cross-institutional and cross-disciplinary engagement that have been evolving—with many valuable advances and contributions by archivists—for decades.”<sup>2</sup>

Data curation is a concern for researchers in many disciplines. As such, it is also part of the larger issues that surround the topic of e-science.<sup>3</sup> In 2009, the Association of Research Libraries conducted a survey to determine their member libraries’ engagement with e-Science

---

<sup>1</sup> Tyler Walters and Katherine Skinner, *New Roles for New Times: Digital Curation for Preservation* (March 2011), 26

[http://www.arl.org/bm~doc/nmt\\_digital\\_curation17mar11.pdf](http://www.arl.org/bm~doc/nmt_digital_curation17mar11.pdf), accessed 12 April 2012.

<sup>2</sup> Christopher A. Lee, and Helen Tibbo. “Where’s the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills” *Archivaria* 72 (Fall: 2011): 124-125.

<sup>3</sup> “e-Science is not a new scientific discipline in its own right: e-Science is shorthand for the set of tools and technologies required to support collaborative, networked science. The entire e-Science infrastructure is intended to empower scientists to do their research in faster, better and different ways.” Tony Hey and Jessie Hey. “e-Science and Its Implications for the Library Community” *Library Hi Tech* 24, no. 4 (2006): 515–28.

issues.<sup>4</sup> While nearly half the respondents reported a designated unit or units that provide data curation support, the report did not indicate whether the library or the institution's archives actually had a role in the process.<sup>5</sup>

In response to the increase of data—especially in technology dependent formats—being created and collected by researchers and a desire to provide access to funded research data, national granting agencies have begun to require researchers to submit a data management plan (DMP) along with their application for funds.<sup>6</sup> Depending upon the research, a DMP may articulate the need to maintain data long-term or permanently. This presents challenges for the libraries and archives that support these researchers. The role of the university archivist and the university archives as caretaker—preserving and making accessible, i.e. curation—of university institutional records and faculty (research) material could be an important complement to the new requirements. As such, it would only seem natural that archives may have a role in the data curation process. Our research project's goal was to determine what is occurring in university archives at research institutions in regards to data curation throughout the country, what the institution's archivist believes his/her role and that of the archives is, and whether or not the archivist has been an active participant in the discussion of the long-term preservation of data. This article will address the results of our investigation, contextualize these results within

---

<sup>4</sup> Association of Research Libraries, "Transforming Research Libraries, E-Science, 2009 E-Science Survey" <http://www.arl.org/rtl/eresearch/escien/esciensurvey/index.shtml> accessed 17 August 2011.

<sup>5</sup> Association of Research Libraries, "Notes on ARL E-science Working Group Survey" (September 2009) <http://www.arl.org/bm~doc/mm09esciencenotes.pdf> accessed 17 August 2011.

<sup>6</sup> In 2003, the National Institute of Health (NIH) began requiring a Data Sharing Plan. "Regardless of the mechanism used to share data, each dataset will require documentation. Proper documentation is needed to ensure that others can use the dataset and to prevent misuse, misinterpretation, and confusion. The precise content of documentation will vary by scientific area, study design, the type of data collected, and characteristics of the dataset." ([http://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_guidance.htm#doc](http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#doc) accessed 12 April 2012). In January 2011, the National Science Foundation (NSF) began requiring that "...proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (<http://www.nsf.gov/bfa/dias/policy/dmp.jsp> accessed 12 April 2012). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research result." As of the 2012 Institute of Museum and Library Studies' (IMLS) grant application, they encourages sharing of research data. "The purpose of this section is to help IMLS understand a grant applicant's research practices and plans for management of data that would be generated through a proposed research project. If the proposed project activities will generate datasets with the potential for future re-use or repurposing, answer the following questions." (<http://www.imls.gov/assets/1/AssetManager/DigitalProducts.pdf> accessed 12 April 2012).

archival theory and practice, and suggest additional steps that should be taken to further articulate the archivist's role in data curation.

## DEFINING DATA CURATION

The archival profession has yet to define or adopt a specific definition of data curation that is contextually grounded in the profession in its authoritative source, *A Glossary of Archival and Records Terminology*. However, several institutions and organizations have defined digital curation, data curation or both. For instance, the School of Library and Information Science at the University of Illinois defines data curation, a subset of digital curation, as "...the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain its quality, add value, and provide for re-use over time; this includes authentication, archiving, management, preservation, retrieval, and representation."<sup>7</sup> Lord and MacDonald define data curation in the 2003 *e-Science Curation Report* for JISC as, "The activity of, managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use."<sup>8</sup>

Some institutions include the concept of added value to their definitions. The Digital Curation Centre (DCC) states that digital curation involves "...maintaining, preserving and adding value to digital research data throughout its lifecycle."<sup>9</sup> While the Interuniversity Consortium for

---

<sup>7</sup> University of Illinois at Urbana-Champaign, The Graduate School of Library and Information Science, "Specialization in Data Curation | www.lis.illinois.edu" [http://www.lis.illinois.edu/academics/programs/specializations/data\\_curation](http://www.lis.illinois.edu/academics/programs/specializations/data_curation) accessed 5 April 2013

<sup>8</sup> Philip Lord and Alison MacDonald, *e-Science Curation Report: Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision* (London: JISC Committee for the Support of Research, 2003) [http://www.jisc.ac.uk/uploaded\\_documents/e-ScienceReportFinal.pdf](http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf) accessed 4 April 2013.

<sup>9</sup> Digital Curation Centre, "What is digital curation? | Digital Curation Centre" <http://www.dcc.ac.uk/digital-curation/what-digital-curation> accessed 4 April 2013.

Political and Social Research (ICPSR) suggests that data curation is the “...value added activities to make specific data as understandable and usable as possible to their community...”<sup>10</sup>

Perhaps the most comprehensive definition comes from the University of California San Diego’s Research Cyberinfrastructure Program. It defines data curation as “Managing data to ensure they are fit for contemporary use and available for discovery and reuse. Archiving and preservation are subsets of the larger curation process, which is a much broader, planned, and interactive process.” The program continues by defining archiving and preservation as “ensuring that data are properly selected, appraised, stored, and made accessible. The logical and physical integrity—including security and authenticity—of the data are maintained” and “ensuring that items or collections remain accessible and viable in subsequent technology environments” respectively.<sup>11</sup>

These definitions are similar to one another, in varying levels of detail; while all have merit, we suggest the following definition as a distillation of these ideas:

*Data Curation is the activity of managing data throughout its lifecycle, appropriately maintaining integrity and authenticity, ensuring that it is properly appraised, selected, securely stored, and made accessible, while remaining usable in subsequent technology environments.*

## **SURVEY SETUP:**

In order to find a sample of archivists at research institutions, we considered surveying several groups. We desired only one response from each institution, preferably from the university archivist or lead archivist. To achieve that goal, we chose to contact the archivists at

---

<sup>10</sup> Inter-university Consortium for Political and Social Research (ICPSR), “Principles and Good Practice for Preserving Data” *International Household Survey Network, IHSN Working Paper No 003* (December 2009): 2. <http://www.ihsn.org/home/download.php?file=IHSN-WP003.pdf> accessed 26 June 2012.

<sup>11</sup> UC San Diego, Research Cyberinfrastructure Program. “Definition of Data Curation” <http://rci.ucsd.edu/data-curation/definition.html> accessed 4 April 2013.

universities that are members of the Association of Research Libraries (ARL), a nonprofit organization of 126 research libraries at comprehensive, research institutions in the United States and Canada that share similar research missions, aspirations, and achievements and that “...operates as a forum for the exchange of ideas and as an agent for collective action.”<sup>12</sup>

Prior to the study, we received approval from The Ohio State University’s Institutional Review Board to conduct the primary survey and a follow-up survey with ARL archivists. Of the 126 ARL member institutions, 116 are universities. The participants for the study—the university archivist, or lead archivist of each institution—were identified by searching each institution’s web site. Participants were sent individual emails (see Appendix I) inviting them to participate in the online survey. Of the 116 institutions surveyed, we received 64 responses of which 55 completed the whole survey.<sup>13</sup> Those who responded remained anonymous unless he/she chose to include contact information at the end of the survey. The sixteen archivists who identified themselves and agreed to be contacted were emailed a follow-up survey designed with seven open-ended questions to expand upon their initial responses (see Appendix II). Of those sixteen participants, five completed the follow-up survey. While the number of responses to the secondary survey is not statistically significant, the responses do provide additional insight into the issues.

## RESULTS

### *Institutional size and reporting structure*

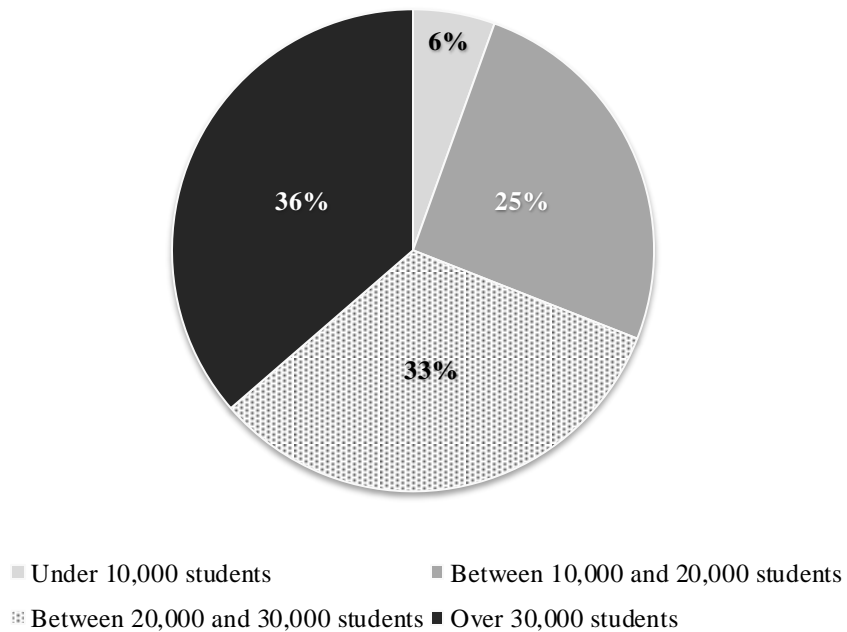
As one might expect, ARL institutions tend to be large research universities. 69% of the respondents answered that their institutions have more than 20,000 students, 36% have more

---

<sup>12</sup> Association of Research Libraries, “About ARL, Membership” <http://www.arl.org/arl/membership/members.shtml> accessed 12 April 2012.

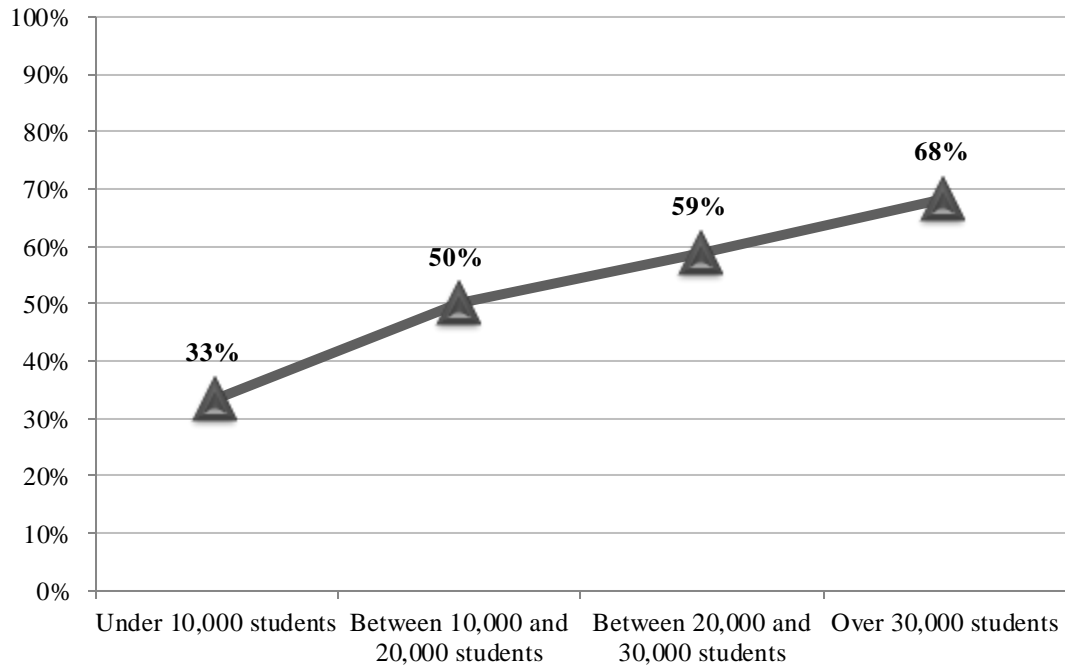
<sup>13</sup> The statistics throughout this article are calculated based upon the 55 completed surveys.

than 30,000 students (Figure 1). An overwhelming majority of the respondent's archives, 91%, are part of their university's library system, while the remaining 9% identified other reporting lines including the Office of the President, Chief Academic Officer, and the Office of the Provost or Vice Provost.



**Figure 1. Respondents' Institutional Size**

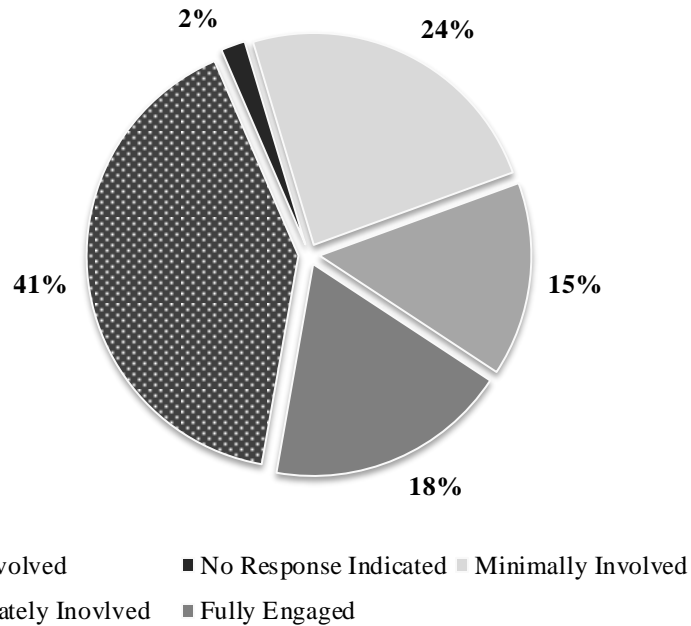
Regardless of their reporting line, the responses indicated approximately 60% of the archivists are involved in campus discussions pertaining to data curation. However, the size of the institution appears to be related to the involvement of the archivist in the ongoing data curation discussions. As can be seen in Figure 2, there is a linear progression from a 33% involvement at the smallest institutions to 68% participation at the largest.



**Figure 2. Archival Involvement with Data Curation**

### *Archivists Role in Data Curation*

When writing the survey, we did not expect a significant percentage of involvement in data curation. However, as illustrated in Figure 3, 57% of the respondents indicated that they are minimally involved, moderately involved or fully engaged in the discussion of data curation at their institutions. This is a much higher percentage than we anticipated. The archivists' responses ranged from "Very indirectly - informal, grapevine conversations" to "The University Archivist is a member of a data curation working group comprised of individuals from Libraries, IT, and Office of the VP for Research."



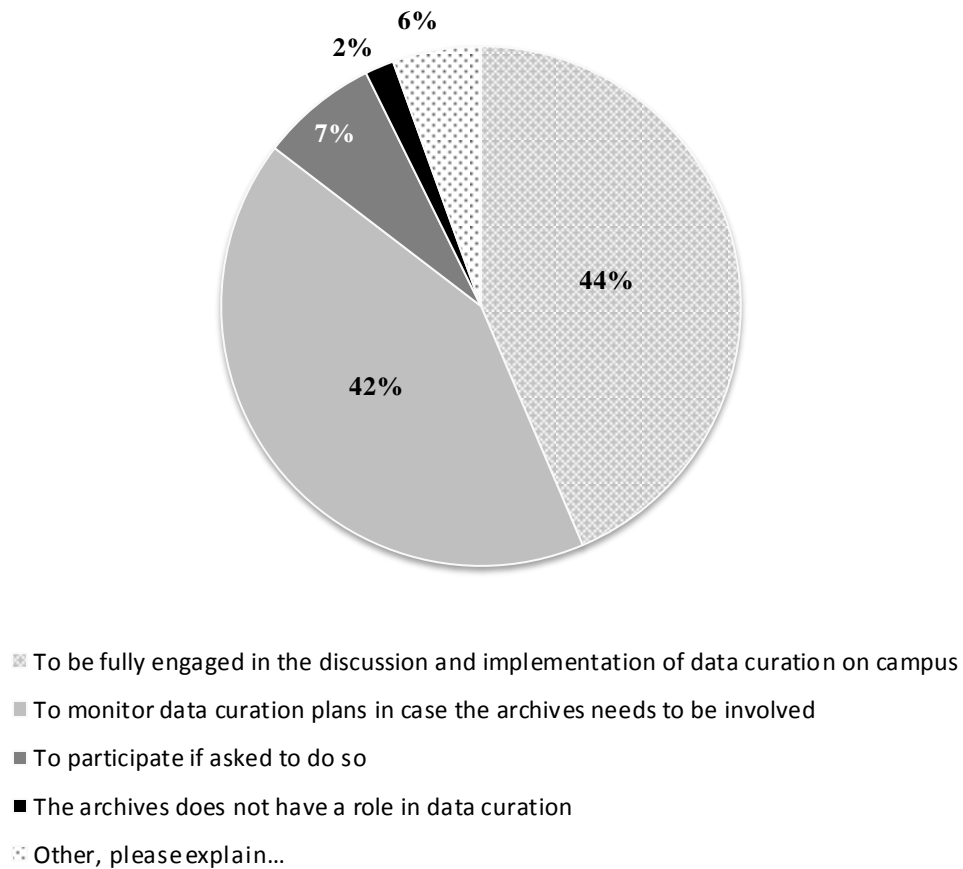
**Figure 3. Level of archival involvement in institutional discussion of data curation**

A slightly larger percentage of respondents (65%) indicated that they believed their institutions' administrations are actively concerned about data curation.<sup>14</sup> Irrespective of whether or not the archivists felt involved in data curation discussions, we asked them what they believe their role and the role of the archives should be. A clear majority of respondents (86%) feel that archivists should be fully engaged in the process or at least have the authority to monitor plans to determine whether or not the archivist needs to be involved on a case-by-case basis

<sup>14</sup> The respondents who completed the survey answered Yes (29), No (4), I do not know (12) and 10 provided expository answers, The expository answered were analyzed and categorized as 'Yes', 'No' or 'I do not know'. The revised totals were 36, 4 and 15 respectively.



(Figure 4).



**Figure 4. Perceived archival role in data curation**

Despite this, only slightly over half (30 respondents or 54%) of the archivists feel that he/she has been able to fulfill his/her perceived role. One respondent noted, “The role I’d like to see is the first option [to be fully engaged in the discussion and implementation of data curation on campus], but it’s a constant struggle to have our voices heard.”

### *Collecting Habits*

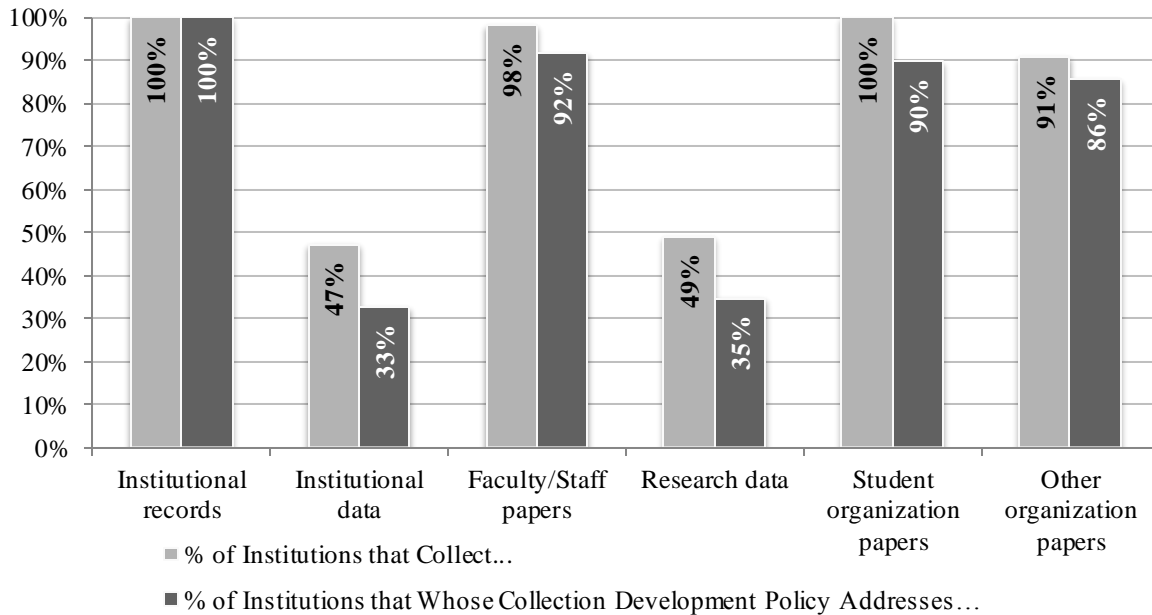
Archival best practices recommend that archives should have written collection development policies that address what is retained by the archives and, often, what is not

collected.<sup>15</sup> We wondered as more archivists consider collecting data whether collection development policies articulate a commitment to collect institutional and/or research data. Institutional data is the data maintained in enterprise systems that assists an institution in conducting its daily business. These systems typically manage the accounting, financial, and human resources functions, as well as student records and research grants administration in a higher education setting. Institutional data also includes analyses and reports generated by the system. Research data on the other hand is the information collected and generated in the course of faculty research that is used to support the outcomes of their research; whereas faculty papers would include the product (e.g. book, article, white paper, report) from an analysis of the data.

Of the institutions that responded 87% have a written collection development policy. Collection development policies do not, however, always reflect what the archives truly collect. All respondents accession institutional and student organization records, and nearly all collect faculty papers (Figure 5). While all of the collection development policies address their institutional records, only 90% and 92% address student organization records and faculty papers respectively. Similarly, while 47% of the respondent institutions are collecting institutional data and 49% research data, only 33% address institutional data in their collection development policies and 35% address research data. We will explore the implications of this disconnect later in the analysis section of this article.

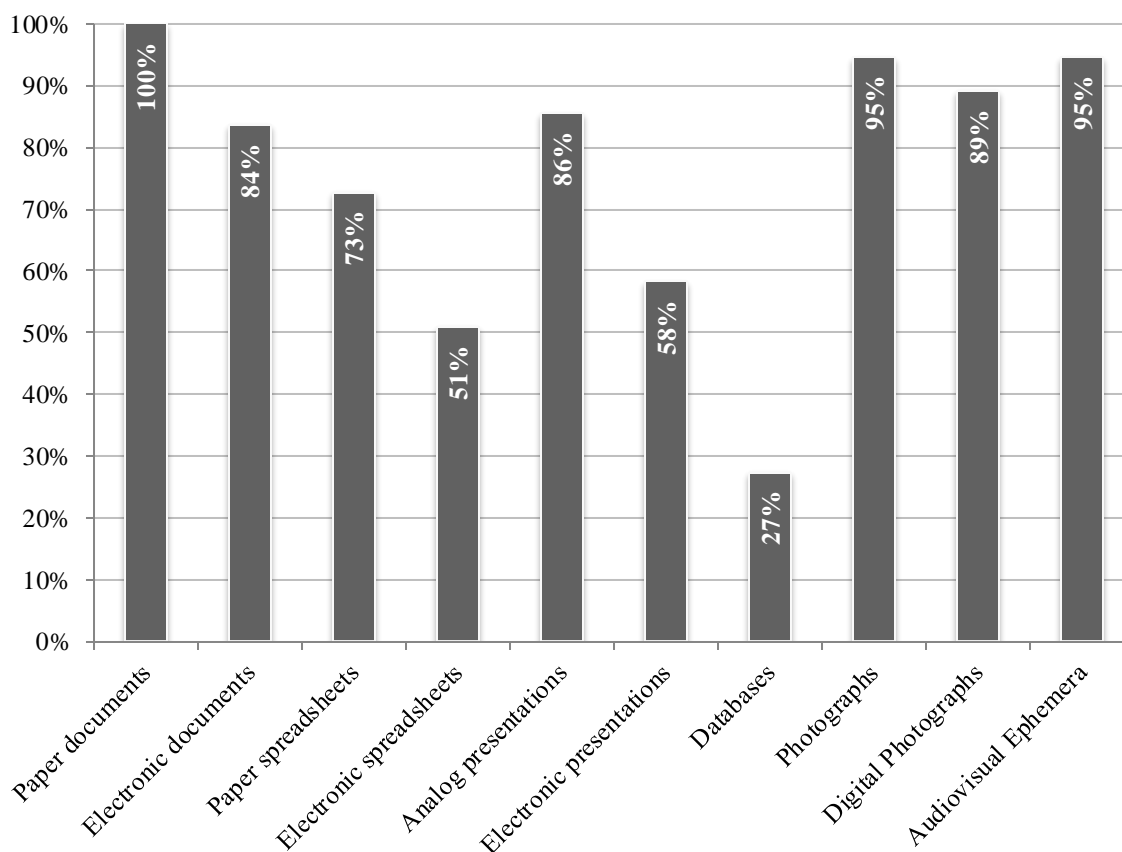
---

<sup>15</sup> In *Developing and Maintaining Practical Archives, Second Edition*, Gregory S. Hunter suggest “Policy statements for archival programs typically cover the following points...scope of items the archives or manuscript repository will collect...”(p44). James M. O’Toole and Richard J. Cox note in *Understanding Archives & Manuscripts*, “...the archives will need a clear statement about the scope of its holdings. What kind of records documenting what kinds of people and activities...What subject areas will be the special purview of this archives...The policy should be reviewed periodically as the archives extends its work into the future, accommodating changing institutional demands, research interests, and information technologies.”(p116). Frank Boles in *Selecting & Appraising Archives & Manuscripts* states, “...a well focused collecting policy will result in a body of complementary, interrelated collections that may become a cultural treasure...”(p. 66). Lastly, Laura Millar, declares in *Archives: Principles and Practices*, “The framework for establishing the logic behind archival acquisitions is the institutional acquisition policy. Building on core archives policy, the acquisition policy defines the scope of acquisitions: what will the archival agency acquire and what it will exclude from its holding.” (p. 116)



**Figure 5. Collecting Habits vs. Collection Development Policy**

In addition to inquiring about their collection development policies, we asked the archivists what type of record formats their repository collects. As shown in Figure 6, more than 80% of archivists tend to collect electronic documents and digital photographs. Whereas only 27% to 58% of the respondents are collecting more dynamic electronic files such as databases, spreadsheets and presentations.



**Figure 6. Types of Materials and the Percentage of Respondents Who Collect Them**

As noted above in Figure 5, of those surveyed, nearly 50% have been collecting either institutional or research data. 78% of those who indicated they are collecting institutional or research data have been doing this for more than two years.<sup>16</sup> The archivists that indicated that they are collecting data provided twenty-five different answers describing the types of data they collect. We found it difficult to categorize this disparate group of answers as some approached their responses in terms of file formats while others described the records in terms of institutional or research data. The responses ranged from “Primarily paper research data generated by the

<sup>16</sup> This represents 18 respondents (33%) of the 55 to complete the survey.

sciences and the medical school...” to “We do not actively try to collect data, but more deal with what people deem significant enough to bring to us.”

When asked, “Is collecting electronic data different than collecting paper-based data and/or records?” the definitive affirmative answers (55%) outweighed the definitive negative answers (18%) by more than 3 to 1, while there was a smaller percentage (11%) of answers that could be categorized as “maybe.”<sup>17</sup> This question generated open-ended responses such as:

- “No philosophically, but yes practically.”
- Addressing the notion of a practical difference, another respondent stated, “I believe that digital material is far more difficult, ephemeral and expensive to handle. Paper-based material may be voluminous but we have almost a century of experience in handling it and it has a history of lasting a long time. Electronic data also must be saved, transferred to current computer formats, to be viewed unless printed out.”
- “Different issues but similar themes – e.g. both require sufficient descriptive and preservation metadata and other documentation to be reusable.”
- “Papers are often self-explanatory...data requires interpretation (i.e. code books) and documentation of the context: normalization and calculations performed against it.”

The respondents’ answers support the theory that it is not as much about whether or not archivists should collect research data, but if we can. Philosophically, if the materials—whether electronic or paper-based—fall within an institution’s mission and/or collection development

---

<sup>17</sup> 16% of the 55 respondents who completed the survey did not indicate an answer to this question.

policy, archivists should collect, preserve and provide access to it. The more pertinent question to ask then is whether or not archivists have the resources to do so.

## **DISCUSSION AND IMPLICATIONS OF STUDY**

In contextualizing the results of our survey with archival theory and practice we have identified areas that warrant a deeper discussion: the decision to collect research data, the articulation of data curation in collection development policy, and the implications of collecting research data.

### *Deciding Whether or Not to Collect*

Archivists must decide whether or not they and their repositories have a role in collecting and curating research data. One archivist in the follow-up survey stated, “When we engage with faculty about donating their papers, we are including research data in the discussion. We do have space constraints, like all archives, so we try focus on original research in the disciplines identified in our collecting policy.” When archivists choose to involve themselves in the data curation process, they need to make certain it is compatible with their repository’s mission and collecting policies and determine whether they have the capacity and skills to do it effectively. If archivists choose not to collect and curate data, does the profession abrogate its right to archive other types of materials? Not necessarily. The repository’s mission statement and collection development policy should govern its participation in the data curation process and the collection of materials. Through collection development policies archivists have always limited what their repositories collect; therefore if the archivist chooses not to participate at this time, because it does not fit with their repository’s mission or policies, archivists do not necessarily remove their

authority to collect research data in the future. Furthermore, participating in the data curation process does not necessarily mean that archivists have to collect research data. If the archivist determines that his/her repository does not currently have the capacity for collecting and curating data, he/she may still participate by collaborating with researchers to identify appropriate repositories and curation best practices.

One respondent expressed the sentiment that data curation will ultimately be addressed by scientists and their organizations, rather than universities. When an archives does not have the capacity to collect and accession particular research data, or it falls outside of the scope of their collection development policy, the archivist may still play a role by encouraging the use of discipline specific repository, in essence a form of documentation strategy.<sup>18</sup> One such example is the Inter-university Consortium for Political and Social Research (ICPSR), which maintains a data archive of more than 500,000 files of research in the social sciences is an example of this for social science data.<sup>19</sup> Similarly, the National Snow and Ice Data Center (NSIDC) collects, manages, and distributes scientific data, creates tools for data access, supports data users, performs scientific research, and educates the public about the cryosphere.<sup>20</sup> Both of these repositories are multi-institutional consortia, and although both are based at academic institutions—University of Michigan and University of Colorado respectively—they are not part of the university archives.

---

<sup>18</sup> A methodology that guides selection and assures retention of adequate information about a specific geographic area, a topic, a process, or an event that has been dispersed throughout society. Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* Society of American Archivists (Chicago: Society of American Archivists, 2005) [http://www.archivists.org/glossary/term\\_details.asp?DefinitionKey=225](http://www.archivists.org/glossary/term_details.asp?DefinitionKey=225) accessed 26 June 2012.

<sup>19</sup> Inter-university Consortium for Political and Social Research, "About ICPSR" <http://www.icpsr.umich.edu/icpsrweb/ICPSR/org/index.jsp> accessed 26 June 2012.

<sup>20</sup> National Snow and Ice Data Center, "About the National Snow and Ice Data Center" <http://nsidc.org/about/> accessed 26 June 2012.

### *Collection development policy articulation*

Is collecting research data really any different from collecting faculty papers? One archivist noted, “Faculty papers involve correspondence, authored pieces, minutes of meetings, syllabi, course notes, etc. Faculty research data is data that faculty gather as part of a dedicated research project over an extended period of time.” In other words, faculty papers that archivists typically collect include the synthesized results of a faculty member’s research, while research data is the raw and transformed data that is used as the basis for analysis and synthesis.

Despite the hesitation to collect data, there are many reasons that it may be desirable to do so. Research data may allow future researchers to:

- enhance the understanding of published research by making the actual data available for inspection
- test further assumptions based upon the data and findings
- conduct longitudinal studies
- preserve one of a kind observational data and/or costly reproducible experimental data for future analysis

The reality is that archivists have likely collected research data over the years whether it is actually identified it as such or not.<sup>21</sup> Several respondents reinforced this notion:

---

<sup>21</sup> A great example of research data “hidden” amongst archival collections is the Zooniverse’s Old Weather project (<http://www.oldweather.org/> accessed 17 May 2012) that is harnessing the resources of “citizen science” to transcribe British ship logs from the eighteenth, nineteenth and early twentieth century to ascertain climate data. While not originally collected for their climate data, they are now a wealth of irreproducible, observational data. What is Citizen Science? Research often involves teams of scientists collaborating across continents. Now, using the power of the Internet, non-specialists are participating, too. Citizen Science falls into many categories. A pioneering project was SETI@Home, which has harnessed the idle computing time of millions of participants in the search for extraterrestrial life. Citizen scientists also act as volunteer classifiers of heavenly objects, such as in Galaxy Zoo. They make observations of the natural world, as in The Great Sunflower Project. And they even solve puzzles to design proteins, such as FoldIt. (Scientific American, “Citizen Science” <http://www.scientificamerican.com/citizen-science/> accessed 17 May 2012).



- “We have historically received faculty research data in hard copy and research files. We also have received this kind of data in electronic format as part of faculty donation.”
- “We have never singled out data as something to collect, however datasets do occasionally come as part of collections.”
- “I see no difference between data and traditional research materials, other than size, format, and issues for long-term electronic preservation.”
- “At this time, we have received data rather haphazardly, for the most part, with faculty papers as part of the University Archives.”

Despite this apparent reality, the survey results indicate that there exists inconsistency between archivists’ collecting practices and their collection development policies. As noted previously in Figure 5, 49% of the respondents are collecting research data, while only 35% address it in their collection development policies. The collection development policies at the University of North Carolina at Chapel Hill and Iowa State University are useful examples of policies at archives that collect data. UNC’s policy states:

Types of papers and information we collect: Observational research data—data that is time bound and difficult to recover, repeat, or reconstruct. Types of papers and information we do not collect: Experimental research data—data that is generated through a repeatable process; the documents accumulated during literary and historical research are included in this category as they can be reassembled.<sup>22</sup>

Similarly, Iowa State’s policy includes:

---

<sup>22</sup> University of North Carolina at Chapel Hill – University Libraries, “University Archives and Records Service-University Archives, UNC-Chapel Hill, About - Faculty Papers Collecting Policy” <http://www.lib.unc.edu/mss/uars/faculty-papers.html> accessed 22 June 2012.

Materials sought by the archives: Research files which may include outlines, research designs, raw data, notes, analyses, report of findings.<sup>23</sup>

Whether or not archivists plan to collect data, doing so should be addressed in their collection development policy.

### *Implications of collecting research data*

A very practical issue archivists face if they embrace a role in the data curation process is the potential to be inundated with massive amounts of data. Do repositories have the capacity—technological, personnel, and fiscal—to effectively process these collections? If archivists do not have the capacity, is it wise to collect and accession these records? Without the appropriate technological resources—and the human resources to implement and utilize the technology, as well as the fiscal resources to fund and sustain the effort—archival repositories may not be able to effectively manage collections. This leaves the collection inaccessible or, potentially worse, makes collections accessible in such a way that possibly exposes private and/or restricted data. One respondent indicated, “We do have medical research data in the collections which require us to ‘scrub’ personal identification before making [it] accessible.” If archivists cannot safely provide access, is it ethical to collect the materials?

One strategy to address this is suggested by the AIMS (Born Digital Collections: An Inter-Institutional Model for Stewardship) project. The project’s authors articulate the need for archivists to more effectively engage donors in the identification of digital materials and associated rights to those materials prior to the actual donation and accessioning of the

---

<sup>23</sup> Iowa State University – University Archives, “Guidelines for Transfers and Donations - ISU Library Special” Collections <http://www.lib.iastate.edu/spcl/arch/guidelines/faculty.html> accessed 22 June 2012.

materials.<sup>24</sup> The report further indicates the need to utilize forensic tools in the accessioning and processing of born digital objects. One of the respondents to our survey echoed this sentiment, “Archivists need to be more involved with negotiation of the gift so that agreements are signed, copyrights are secured if possible, privacy is protected and context is preserved. All of this appears to be new territory for the digital repository folks. Faculty tend to think of this in terms of IRB [Institutional Review Board] requirements, agreements and restrictions. It seems the IRB releases are often very narrow and no one is advocating and negotiating for public access and/or secondary uses of the data that enable research replication or creative reuse.”

Archivists who collect data must also decide whether to take physical custody of data. The definition of accessioning states: “To take legal and physical custody of a group of records or other materials and to formally document their receipt.”<sup>25</sup> What happens when an archivist accessions research data without taking “physical custody?” This may happen when research data is accessioned and then stored in an institutional repository (IR) that is not necessarily hierarchically part of the archives, thereby creating an issue of custody and/or conflict with the archives collection development policy. One purpose of placing research data into an IR or other digital preservation environment is to provide at a minimum some amount of preservation activity, at least preservation of the bit stream. However, the most practical way to maintain the research data may be to leave it in-situ, thereby maintaining pertinent context and functionality, as opposed to taking physical custody and potentially providing a more suitable preservation environment, albeit with loss of key linkages to contextual information. In either scenario, archivists should develop policy and procedures to address these issues.

---

<sup>24</sup> The University of Virginia Library, in partnership with Stanford University, the University of Hull, and Yale University have been funded by The Andrew W. Mellon Foundation for “Born Digital Collections: An Inter-Institutional Model for Stewardship (AIMS). Over a two-year project period (October 2009 – October 2011), project partners will create an inter-institutional framework for stewarding born-digital content (<http://www2.lib.virginia.edu/aims/> accessed 17 May 2012).

<sup>25</sup> Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Chicago: Society of American Archivists, 2005) [http://www.archivists.org/glossary/term\\_details.asp?DefinitionKey=115](http://www.archivists.org/glossary/term_details.asp?DefinitionKey=115) accessed 17 May 2012.

## FURTHER CONSIDERATIONS OF CURATING DATA

In our review of the literature, professional resources, and the responses supplied in the survey, we have identified additional areas of consideration for action or further research, an investigation of what skills archivists bring to the table, specifically in the area of appraisal and an investigation of the fiscal considerations of curating research data

### *Archival Skills*

One respondent noted, “There is a disconnect in how higher administrators view the role and competencies of archivists... it’s very difficult for them to focus on the principles of data curation as fundamentally the same as all other records curation, rather than focusing on the technology. Very frustrating for the archivists.” If, as the DCC and ICPSR definitions of data curation noted earlier suggest that data curation is a value-added proposition, what value or skills do archivists add? While our survey did not ask the archivists to identify what skills they feel the profession can contribute, this is certainly an avenue for further detailed research. Even so, it is not presumptuous to suggest that the value-added skills—core to the archival enterprise—that archivists provide may include the appraising and selecting of materials, as well as the administration and management of the materials by assisting in describing, organizing, and providing intellectual control.<sup>26</sup> These are activities that archivists can offer to either actually perform on collections or at a minimum to provide advice and guidance to researchers who conduct these activities.<sup>27</sup>

Of these skills, one of the most distinctive that archivists can provide to researchers is appraisal. The core archival function of appraisal provides a framework for assessing the value

---

<sup>26</sup> Christopher A. Lee, and Helen Tibbo. “Where’s the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills” *Archivaria* 72 (Fall: 2011): 160.

<sup>27</sup> Kalpana Shankar. “Towards a Framework for Managing Records in Scientific Research” *Archival Issues* 24 no. 1 (1999):27.

of maintaining records and/or data. Researchers clearly understand that their data has primary value to them. In forming partnerships and working with researchers, archivists can provide training and guidance regarding the identification of the research data's potential evidential and informational secondary values.<sup>28</sup>

InterPARES1 suggest that "...information about the technological context of electronic records comes into play at two vital stages of selection. It is needed when assessing records' authenticity, and when determining the feasibility of preserving authentic electronic records..."<sup>29</sup> Conscientious appraisal considerations result from the deliberate examination of content, context and structure of the data, while heeding the interests expressed by the researcher. This may lead to the development of prudent criteria for the classification, storage, preservation, and retrieval of data.<sup>30</sup> Because of this, it may be of value to researchers to have archivists guiding researchers in "black box" appraisal decisions regarding the preservation of their research and data.<sup>31</sup> By considering the value of the information—including its circumstances of creation, uniqueness, and potential re-use—the cost of retention, and the political implications of the appraisal decision, researchers should be able to come to a more informed decision regarding the data's long term value and whether they need to preserve and provide long term access to their research data.

This approach is consistent with what some of the respondents suggested is the archivist's role:

---

<sup>28</sup> T.R. Schellenberg, *Modern archives: principles and techniques*. (Chicago: University of Chicago Press, 1956), 133-160.

<sup>29</sup> The InterPARES Project. "Part Two — Choosing To Preserve: The Selection of Electronic Records - Appraisal Task Force Report," *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (2002): 5.

[http://www.interpares.org/ip1/ip1\\_documents.cfm?cat=aprf](http://www.interpares.org/ip1/ip1_documents.cfm?cat=aprf) accessed 3 July 2013

<sup>30</sup> Chauncey Bell. "Re-Membering the Future: Organizational Change, Technology, and the Role of the Archivist" *Archival Issues* 25 no. 1-2 (2000): 28.

<sup>31</sup> Frank Boles and Julia Marks Young, "Exploring the Black Box: The Appraisal of University Administrative Records" *American Archivist* 48 no. 2 (Spring: 1985): 121-140.

- [Our role is] More evolving than changing. [We s]till need to provide space (digital) and services to help creators manage their digital data. However, the archives needs to be involved earlier in the records lifecycle and be part of the discussion when...systems are being implemented. Archivists should be at the table from the beginning when institutions embark upon data curation activities, regardless of the types of data being brought in. They can offer good insight on appraisal, access, acquisition, and digital preservation and complement the digital initiatives librarians and IT staff that generally are part of the process.
- Archivists need to be involved from the very beginning in terms of discussions, planning, and implementation. Archival work will be shifted to seeking out and acquiring a different form of archival record, as well as guiding faculty in uploading material, and then embrace use of this data by students and scholars as a normal part of reference work.

### *Fiscal implications of research data*

As we saw in Figure 3, there was a relationship between the size of the institution and whether or not the archives was involved in data curation. Further research should investigate if there is a correlation between collecting research data and institutional funding. It is costly to do data curation correctly. Francine Berman, in the *Chronicle of Higher Education* notes:

“Retention of valued research collections past the project completion date poses a serious problem. Universities and their libraries [and by extension archives] need financial support for data stewardship and preservation. They cannot absorb the bulk of retention-worthy research

data without additional money.” Where will that money come from? Berman suggests, “Universities, and in particular their libraries, can improve and expand their repository infrastructure for research data through institutional investment, user fees, federal investment, philanthropy, or some combination thereof.”<sup>32</sup>

Federal and philanthropic investment, while helpful, may only have a limited impact. Since these resources typically provide seed money for these efforts, but not indefinite funding, they are therefore not the basis of a sustainable model. The allocation of resources to digital curation activities needs to be a continuous effort, but all too often is not part of the project planning and scoping.<sup>33</sup> User fees, while not unheard of, are less desirable as they provide a barrier to access and use, which is contrary to the archival profession’s Core Values.<sup>34</sup> Instead, archivists must recognize that funding the housing and preservation of born digital assets is theoretically no different than constructing bricks and mortar repositories, funding annual plant operation and maintenance, and the inevitable funding of renovations and expansions.

---

<sup>32</sup> Berman, Francine “Who’s Paying the Data Bill?” *The Chronicle of Higher Education* (April 29, 2012) <http://chronicle.com/article/article-content/131670/> accessed 3 May 2012.

<sup>33</sup> The 2002 RLG-OCLC report, “Trusted Digital Repositories: Attributes and Responsibilities” ((May 2002), <http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf?urlm=161690> accessed 29 August 2013) defines the attributes of a trusted digital repository including financial sustainability. There has been a variety of literature since that report that address concerns of sustainability. Brian Lavoie and Lorcan Dempsey remarked, “Two key economic challenges plague efforts to preserve digital materials. First, allocation of funds to digital preservation has been insufficient. . . . Second, funds that are made available are usually provided on a temporary basis, often as grants to support one-off undertakings or special projects.” (“Thirteen Ways of Looking at . . . Digital Preservation” *D-Lib Magazine* 10, no 7/8 (July/August 2004) <http://dlib.org/dlib/july04/lavoie07lavoie.html> accessed 8 August 2013). Helen Hockx-Yu observed in “Digital preservation in the context of institutional repositories” *Program: electronic library and information systems* 40 no. 3 (2006): 232-243 “The fact that many institutional repositories rely on short-term funding makes it hard to make long-term stewardship commitments. Concern of sustainability contributes to the lack of active engagement from many researchers. . . . Since, the core funding for institutions does not grow in line with information growth, this makes it difficult to justify and make a case for digital preservation to institutional management.” Kevin Bradley noted in “Defining Digital Sustainability” (*Library Trends*, 56, no. 1 (Summer 2007): 148–163) “. . . digital preservation is as much an economic issue as a technical one. The requirements of ongoing sustainability demand a source of reliable funding, necessary to ensure that the constant, albeit potentially low-level support for the sustainability of the digital content—and its supporting repositories, technologies, and systems—can be maintained for as long as necessary. It is not too strong to say the biggest single risk to sustained access to digital information is economic.” More recently, Ricky Erway and Brian Lavoie in “The Economics of Data Integrity” Curating for Quality: Ensuring Data Quality to Enable New Science (Pre-print available online at: <http://www.oclc.org/content/dam/research/publications/library/2012/erway-dataintegrity.pdf> accessed 12 April 2013) suggest “Many times funding is available to establish a data archiving service, but is not available for ongoing operations. Likewise, funding may be included in a grant proposal to cover the costs of preparing a particular dataset for deposit in the archive, but not for its long-term care. Another challenge for service providers is that technology and preservation practices change over time, requiring acquisition of new hardware and software and reworking of processes.”

<sup>34</sup> The Access and Use value in the SAA’s Core Values of Archivists states, “Archivists promote and provide the widest possible accessibility of materials, consistent with any mandatory access restrictions, such as public statute, donor contract, business/institutional privacy, or personal privacy. Although access may be limited in some instances, archivists seek to promote open access and use when possible. Access to records is essential in personal, academic, business, and government settings, and use of records should be both welcomed and actively promoted. . . .” <http://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics> accessed 5 April 2013

## CONCLUSION

While acknowledging institutional and research data pose complex curation and preservation challenges, archivists should not view this complexity as a barrier to engaging in the process of data curation. Archivists should address data curation by determining what type of role they choose to play—active participant, consultant, both, neither—and highlighting the expertise that the profession can provide. Through this research endeavor, we have a glimpse of what some archivists in the United States perceive to be their role in the process of data curation and description of some of their actual involvement. It is clear from the survey that archivists—at least at research-intensive institutions of higher education—are engaged in curating institutional and/or research data at a greater rate than one might expect. Despite this, in some instances collection development policies have not caught up with this activity.

Further avenues of research that emerged from this project include broadening the investigation to a larger sample of the archival profession, identifying skills that archivists can provide to add value for researchers and their data, investigating the correlation between collecting research data and institutional funding, and developing sustainable funding models.



## APPENDIX I: INITIAL SURVEY INSTRUMENT ADMINISTERED VIA ZOOMERANG

### Email Invitation

As you are likely aware, data curation has become an important topic for researchers, libraries and librarians, and IT units in research institutions. However, one voice not often heard from yet on the topic is that of the archivist. To get a better understanding of how archivists are addressing or involved in data curation, we are conducting a survey of archivists from Association of Research Libraries institutions. The survey consists of 20 questions and should take no more than 10 minutes. Those taking the survey can stop participating at any time. We will not know your identity unless you choose to provide it. If you choose to provide your contact information, all records will be kept confidential.

Before taking the survey, you will be asked to give your consent to collect the data. If you would like to take the survey, please follow this link:

Survey will be available from today through December 2.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1.800.678.6251.

If you have any questions or concerns, please contact us using the information below.

Sincerely,

Tamar Chute, Co-Investigator  
University Archivist  
The Ohio State University  
2700 Kenny Road  
Columbus, OH 43210  
614.292.3271  
[Chute.6@osu.edu](mailto:Chute.6@osu.edu)

Daniel Noonan, Co-Investigator  
Digital Resources Archivist  
The Ohio State University  
221D Thompson Library  
Columbus, OH 43210  
614.247.2425  
[Noonan.37@osu.edu](mailto:Noonan.37@osu.edu)

## Survey Instrument with responses

Our Institutional Review Board requires that respondent identifiable information not be included, therefore where this appears in the survey results below, we have replaced it with [redacted].

**1. Informed Consent: The survey consists of 20 questions and should take no more than 10 minutes. You may stop participating at any time. We will not know your identity unless you choose to provide it in the last question. If you do not provide your name, your answers will be kept anonymous. If you choose to provide your information, we may contact you to ask a few follow-up questions. You may or may not choose to participate in the follow-up questions. In either case, all records will be kept confidential. We will destroy any identifying information such as name, institutional name, and/or non-generic job titles. For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1.800.678.6251. If you choose to continue participating in this survey, please click "Yes" below; if you choose not to continue, please click "No" below and this survey will end:**

1 Yes	64
2 No	0
<b>Total</b>	<b>64</b>

### **2. What is the size of your institution?**

1 Under 10,000 students	4
2 Between 10,000 and 20,000 students	16
3 Between 20,000 and 30,000 students	20
4 Over 30,000 students	24
<b>Total</b>	<b>64</b>

### **3. Is the institution's Archives a unit of the Libraries?**

1 Yes	56
2 No	7
<b>Total</b>	<b>63</b>

### **4. If "No" to Q3, who does the archives report to?**

1. Office of the Provost
2. Office of the President
3. Business Affairs
4. Vice Provost, Libraries and Cultural Resources
5. Office of the Provost
6. Chief academic officer
7. Vice Provost of Libraries, Computing & Technology

**5. Does the archives have a written collection development policy?**

1 Yes	55
2 No	8
<b>Total</b>	<b>63</b>

**6. If “Yes” to Q5, does the collection development policy address (check all that apply):**

1 Institutional records	55
2 Institutional data	18
3 Faculty/Staff papers	51
4 Research data	19
5 Student organization papers	50
6 Other organization papers	49

**7. Does the archives collect (check all that apply):**

1 Institutional records	62
2 Institutional data	32
3 Faculty/Staff papers	60
4 Research data	33
5 Student organization papers	62
6 Other organization papers	56

**8. Do you collect (check all that apply):**

1 Paper documents	62
2 Electronic documents	53
3 Paper spreadsheets	46
4 Electronic spreadsheets	33
5 Analog presentations	52
6 Electronic presentations	36
7 Databases	18
8 Photographs	59
9 Digital Photographs	56
10 Audiovisual Ephemera	58

**9. If you are collecting data, please list/describe the type(s) of data collected.**

1. We do not actively try to collect data, but more deal with what people deem significant enough to bring to us. With that in mind, we do have a number of faculty research collections containing original data, ranging in age and subject from fairly recent sociological studies to 120 year old daily meteorological reports.

2. We are only in the very beginning of collecting a variety of external drives, diskettes, floppies and other forms of digital forms
3. We have historically received faculty research data in hard copy and research files. We also have received this kind of data in electronic format as part of faculty donation.
4. We have never singled out data as something to collect, however datasets do occasionally come as part of collections.
5. Databases, spreadsheets, audio and video files, image files, moving image files, text files
6. Statistical data reflecting development of institution: finances, students, holdings, staff, faculty, programs, organizations
7. polling data
8. Primarily paper research data generated by the sciences and the medical school. The Library, but not specifically the archives is involved in working on a repository for electronic data.
9. For example, GIS data files.
10. Electronic research data
11. Mostly faculty research data in paper form. Some electronic files are coming to us but we haven't evaluated them.
12. data sets, grant research data, institutional data
13. Data may form the basis of institutional records or may be contained within the records of private "manuscript" collections. Types are as varied as record groups or collections.
14. Research Data Administrative Business Datasets
15. We have always collected field notes and lab notes in paper form.
16. Spreadsheets, documents (pdfs, doc, txt) images, audio, video
17. Storage Media
18. There is some incidental data held in professor's papers, but no electronic data held in databases. There are some electronic records of music.
19. institutional data and research data
20. Personal digital archives (word processing, email); research data (spreadsheets, images, other tabular data); institutional records and data
21. We don't have what I would consider much research data. We do have some lab notebooks, and have some scheduled.
22. Very rarely, when included in a collection, but seldom as separate materials.
23. Laboratory and field notes. I am defining "data" rather broadly.
24. So far, we don't have an actual plan for which types of data we do or don't collect. We're just taking everything we're offered. Currently, that includes Word, TXT, PDF, PowerPoint, Bitmap, JPEG, and maps (SHP, SHX, GPX, DBF).
25. Electronic data is governed by the same collection policy for paper-based documents.

**10. If you are collecting data, when did you begin doing so?**

1 In the past 6 months	1
2 6-12 months ago	0
3 Between and year and two years ago	4
4 Over two years ago	18
<b>Total</b>	<b>23</b>

**11. Do you feel that collecting electronic data is different from collecting paper-based data?**

1 Yes 15

2 No 5

Please explain your answer... [up to 945 characters]

27

**Total** 47

1. Preservation of electronic data is much more challenging. We know how to deal with paper. Maintaining data integrity and accessibility are huge issues with no easy answers.

2. Yes, only because it requires different skill sets and requires a different kind of "storage space" as well as occasional migration to newer technology.

3. No, the content and the intent for collecting it remain the same; however, the format change from paper to electronic presents new challenges for preservation.

4. Collecting electronic data requires staffing and hardware from entities outside of the archives.

5. No (I can't both click "no" and explain!), but... the difference is in the long-term caretaking. Print materials we know can be read in 50 years, while an Access database pretty much has to be kept in that format (printing a complex one would be nearly impossible) and probably will be unaccessible in 10 years, much less 50, and more obscure programs have already proven so dated as to make them unretrievable without a lot of work.

6. Yes, I believe that digital material is far more difficult, ephemeral and expensive to handle. Paper-based material may be voluminous but we have almost a century of experience in handling it and it has a history of lasting a long time. Electronic data also must be saved, transferred to current computer formats, to be viewed. unless printed out. As yet, we are not configured for such access.

7. Yes, the sheer volume and difficulty in providing access.

8. Data itself is the same. If you collect one, you should collect the other.

9. Neither yes nor no: there are points of similarity and points of obvious difference, practically and theoretically.

10. Takes a different skill set to provide preservation and maintenance

11. You need IT/Digital infrastructure. We either bag it and tag it (which implies limited or no access because of lack of description) or we index, describe and arrange (very time consuming and costly).

12. Electronic data has greater potential for re-use but presents greater preservation challenges.

13. It is different because of the distributed nature of the materials and difficulty of keeping the associated.

14. It differs only in that we have yet to devise an effective means of collecting and preserving it.

15. Different issues but similar themes - e.g. both require sufficient descriptive and preservation metadata and other documentation to be reusable. Obviously digital files have some affordances in terms of reuse that paper doesn't have.

16. mostly the preservation issues surrounding the electronic data

17. have to take obsolescence into account

18. No, it's data, but I think the possible size of the data (files) can pose a problem.

19. Sustainable data and codebook format are important issues to be discussed with donors. Data analysis tool versions might impact replicable data processing.
20. Obviously it differs in medium, but that is only the beginning. Institutional data is a fundamental responsibility for the archivist; scientific research data is, in my opinion, a quite different problem...one that I think ultimately will be addressed by the scientists and their professional organizations, rather than institutions within the academy.
21. It shouldn't be but maintaining access and usability has proven to be a challenge.
22. The collecting of electronic records requires a closer connection to creation than analog records.
23. More work involved, creator is more involved with the acquisition, new skills necessary for archival staff
24. No philosophically, but yes practically.
25. They have some different requirements for ingest and preservation, but info is info.
26. It's not different intellectually, but it's different in terms of preservation and access/handling.

**12. Are administrators (i.e. Directors, Deans, Vice Presidents, Provosts, Chancellor, President) at your institution concerned about data curation?**

1 Yes	30
2 No	4
3 I do not know	12
Additional comments [up to 955 characters]:	10
<b>Total</b>	<b>56</b>

1. Only minimally, and the concern is not widespread.
2. While I know the Library Dean is aware of the issue, I'm not sure anyone else is!
3. They are now, due to the NSF data curation requirements (and NEH open access requirements as well, I'm told)
4. We are at the "hot potato" stage - everyone knows we have to Do Something, but no one wants to take leadership
5. There is a concern as it relates to possible litigation and an abstract notion of value. But the commitment of funds on a local or systemwide scale is only now being investigated.
6. yes, because profs are all starting to collect digital or electronic archives on their own with no backup, etc.
7. Recent institutional research survey that was distributed about data management practices; high-level taskforce created by provost
8. there are several campus groups looking at data curation
9. They say they are but nothing has been done on a university-wide basis yet
10. Partially, but mostly they are concerned about born-digital published material produced by faculty, staff, and students or purchased by the library (theses and dissertations, articles, exhibits, other publications). There is also a movement to build an OAIS repository for research data sets.

**13. If your institution is discussing data curation, has anyone from the archives been involved?**

1 Yes	31
2 No	23
<b>Total</b>	<b>54</b>

**14. If “Yes” to Q13, how is your archives involved?**

1. We have had some initial discussions with administrators, and we led an Electronic Records Working group several years ago. Sadly the report from that group never left the Libraries and made its way to upper-level administrators.
2. We serve on the committee exploring an institutional repository, and as we are responsible for records management, we are developing guidelines for electronic records maintenance and preservation.
3. The University Archivist is a member of a data curation working group comprised of individuals from Libraries, IT, and Office of the VP for Research.
4. There will be a webinar this month (November) for the university system regarding the new data curation system.
5. As yet, the desire to support archival initiatives seeking the ability to archive email and computer data have yet to advance beyond the exploratory phase.
6. Only marginally
7. We are represented on the library's digital strategies committee and have reps on the working groups for metadata, digital preservation, and data.
8. We are at the table providing input in arrangement and description and also often times our donors or potential donor are providing the content or funding.
9. Very indirectly - informal, grapevine conversations.
10. The Science Library has created a data curation librarian position and are offering consulting services and partnering to store and preserve data on the sciences campus. the overall organization of the library units and the institutional archives are collaborative, so each unit does it's own collecting and managing and there is a central storage location.
11. As part of a general committee related to electronic records.
12. Our digital repository is led by two co-directors, myself and our "research services librarian" whose chief area of interest and expertise is in data. We work closely on collecting scope, metadata guidelines, training for faculty, etc, including data management workshops.
13. very minimally we have a digital archivist in university archives. the library is hiring a digital data outreach librarian who will report to the coordinator of subject librarians. we have a librarian in Digital Library Services that will work with an IR when established. we will try to work across these lines of divide.
14. Archivist on Data Curation Committee; assist in creation of process for data management plans.
15. Through discussions re. institutional repository and data security, primarily.
16. The digital collection archivist is on the data curation research committee
17. Working with the digital institutional repository for collecting and preserving data
18. In the past four months our library has created a Digital Policy committee which includes archive's personnel.
19. Chairs, Records Management Advisory Committee Member, Data Categorization Committee Member, Administrative Information Systems Management Committee

20. We have a digital content committee; we run open access week symposiums, we currently operate a digital asset management system; we have a dedicated position to digital materials called the "Curator of Digital Collections"
21. The involvement comes from the Institutional Archive Division within Special Collections. The Manuscript Division and Multi-Media Division are not that involved.
22. Involved at planning level within the Libraries, but not actively undertaking work now.
23. I serve on a committee that provides a digital scholarship service for faculty. One service is to provide consultations on good data management practices.
24. We are talking about the Libraries (and Archives) being the lead about holding at least digital archives for other units, but not data sets.
25. In trying to persuade the university to hire an electronic records manager
26. University Archivist is on campus-wide research data task force and [redacted] data storage work group. University Archives led a year long, university-wide Digital Curation Planning Project. Electronic Records Archivist is on a Library committee focused on digital information. Archivists co-teach a faculty seminar on research data curation. University Archivist is on enterprise business systems steering committee focused on university finance and human resources business systems.
27. Represented on committees, part of repository development process, data management plan consultations
28. Either on committees or kept in loop through discussion with people. The main campus group looking at data curation is a joint venture between the library and the computing center.
29. On a committee
30. The implications of managing data at the "end" of its life cycle.
31. I'm new to the job, so I don't know if before I came (before Sept. 2011), anyone from the Archives was involved in the OASIS or institutional repository conversations. I have not been involved so far, and, in fact, there is an entirely separate division of the libraries (Digital Library Services) that has been involved.

**15. If "No" to Q13, do you see this as a problem?**

1 Yes	14
2 No	0
Please explain...[up to 960 characters]	9
<b>Total</b>	<b>23</b>

1. At the moment here, data curation is being treated no differently than any other digital preservation issues. Given our budget issues, this is not likely to change anytime soon, and until there's more demand/concern on this issue, I'm kind of fine with that. Well, not fine, but its rare enough here that other priorities have to take, well, priority.
2. Yes. The longer that administrations and libraries wait to make decisions regarding the archiving of electronic materials, the more likely it is that data and records will be lost.
3. Faculty research is considered their own property, not University property, so research data wouldn't be transferred to me even if it were on paper.
4. At this stage it is not a problem, but we will want to be involved in the discussion.
5. If the Archives is responsible for the long term care of the data it should be involved.



6. Archivists can help interpret sustainability and ownership issues with donors. Often these collections are hybrid electronic and hardcopy materials and archivists can help coordinate access to the related hardcopy.
7. There's a disconnect in how higher administrators view the role and competencies of archivists ... it's very difficult for them to focus on the principles of data curation as fundamentally the same as all other records curation, rather than focusing on the technology. Very frustrating for the archivists. And demeaning.
8. On a systemwide level, archivists are involved. On the campus level, we seem trapped by the costs of developing and maintaining something on site.
9. Yes it's a problem, since the University Archives is the repository of record for all permanent university records.

**16. What do you feel is the role of your archives in data curation?**

1 To be fully engaged in the discussion and implementation of data curation on campus	24
2 To monitor data curation plans in case the archives needs to be involved	23
3 To participate if asked to do so	4
4 The archives does not have a role in data curation	1
5 Other, please explain...	3
<b>Total</b>	<b>55</b>

1. Somewhere between "fully engaged" and "monitor." There are issues of long-term preservation (file format migration), interoperability, and access which aren't always raised in campus discussions. We don't have the time/personnel/training to develop and implement the needed technology, but we need to be involved in the development of the system itself, if for no other reason than to raise questions others have not thought of.
2. The role I'd like to see is the first option, but it's a constant struggle to have our voices heard.
3. All of the above. There is not a linear progression, the result is not foreordained.

**17. Is your archives currently fulfilling the role you have defined in Q16?**

1 Yes	25
2 No	30
<b>Total</b>	<b>55</b>

**18. If "No" to Q17, please explain why not.**

1. So far the Archives has had only minor involvement in discussions, and there has been no implementation to date.
2. Not seen as problem yet, but am working with records management on campus to help increase awareness on campus

3. For this campus, we are early on in the process.
4. Despite multiple requests, we have not been invited to be part of the process -- if indeed it is even happening.
5. Dta Curation has been established as a separate department within the library.
6. As noted, this is not something really yet happening on our campus.
7. This university does not have a thorough records management system, has not designated an official archives, has no policy regarding the archiving of, and allowing access to, its public email records. While the university has begun to cease publishing many former paper records, schedules, directories, and so on, there is no effort ongoing, of which I am aware, to insure the duration or archiving of its web-based information and data. While this facility collects university materials above the dean level, we only refer to ourselves as "a repository" of university records. Pretending to be a university archives in the past merely prolonged the university's disinterest in establishing, funding, staffing, and housing an officially designated university archives.
8. Have not been invited much to participate.
9. we don't have the infrastructure or staff to do so.
10. No authority
11. The host institution does not yet possess the necessary IT infrastructure to handle the task.
12. University is decentralized - it would be easier to get to the table if the Library leadership where we are housed helped us get there.
13. At the policy level, responsibility has been delegated to each PI/department. At the implementation level, there is no campus-wide discussion of data management yet. We're still at the stage where each PI or, at best, department is crafting its own solutions. Some concerns about the risks and burdens of this decentralized approach are beginning to be articulated.
14. The archives has not been asked to participate in the data curation discussions.
15. Staffing and resource constraints.
16. It's never enough - I need to be better versed in the issues and to be constantly communicating and connecting with faculty - so, I think we are doing well but it's too big an issue to feel that we're comfortably fulfilling our role.
17. Archives is involved in process of setting up data management system run by both computer operations department in Libraries and Information Services for entire university.
18. Only in early stages of curating electronic documents; not ready yet to undertake curation of data sets.
19. I have not been asked by the current administration. A past AUL asked me to get involved a few years back. He left.
20. We need to have more campus involvement and support.
21. Limited staff and resources, other current priorities, and a lack of awareness of the Archives elsewhere in the University community.
22. Probably because our budgetary resources are inadequate to actually fulfill the role we should play. Without the actual capacity to be a player, it's hard to be taken seriously.
23. Clear responsibility has yet to be assigned within the University and there is no clear mandate (yet).
24. Administration is not interested in electronic records management
25. We're not as involved as we'd like to be. We seem to have to keep reminding groups that Archives has an interest/responsibility in data curation.

26. no resources

27. Lack of staff to push for involvement--there are other high priorities that need to be addressed, first, by the limited staff we have.

**19. Are there other special collections in your institution that collect research data?**

1 Yes	18
2 No	35
<b>Total</b>	<b>53</b>

**20. If "Yes" to Q19, please list the types of collections that include research data?**

1. Special Collections (literary photographic research) [redacted] (music research)
2. Research-centered special collections.
3. The answer should probably be no, depending on what you mean by "collect," but there are many places where data are aggregated. We are more decentralized than centralized and I would not see that changing.
4. Insofar as the material is included in a particular professor's donated material, it would end up in our Manuscripts department.
5. Manuscript and Library collections probably. my guess is different institutes, departments maybe even the research office are collecting data sets
6. If you mean Special Collections in the Libraries sense, yes, I think in a sort of haphazard way. If you mean in a general sense, yes, we have for example the [redacted].
7. earth sciences unit collects research data
8. social science data
9. I answered "No" because I do not believe any other units have real long-term curation strategies. I'd say they're temporarily "storing" data, no curating it for permanent research use.
10. Most of the data curation occurs in IT offices all over campus. This is short-term, generally supported by a single individual at each site, and likely to fold when that person leaves. There is no formal coordination.
11. There are a number of semi-autonomous research teams/groups across campus that are retaining data.
12. several research centers based in colleges on campus collect research data; the Library collects some research data
13. [redacted] Historical Collection, [redacted] Folklife Collection
14. There is a separate data library, although much of their holdings are purchased data.
15. Projects of various sorts run in cooperation between academic depts and library
16. The Digital Research and Curation center collects faculty data for NSF-funded projects.
17. Personal papers; organizational records
18. [redacted] Medical Center has its own archives.

**21. If you are willing to have the researchers contact you if they have any follow-up questions, please give us your contact information below:**

## **APPENDIX II: FOLLOW-UP SURVEY ADMINISTERED VIA EMAIL**

Dear:

Thank you for participating in our “Data Curation and Archivists” survey, and agreeing to answer some additional questions. These questions below allow you to elaborate on several of the questions in the initial survey. You may either include your thoughts in a response email or attach a document with your replies. We would greatly appreciate your response by Wednesday, February 1.

1. Would you be willing to share your collection development policy with the researchers?  
If so please provide link or attach to response email.
2. If you are collecting data, how have you done so, in what format?
3. If you are collecting research data, how are you appraising it?
4. What are the differences between collecting faculty papers and faculty research data?
5. Do you feel your archives is effectively engaged in the discussion about data curation?
6. What direction do you think your archives’ participation in data curation will take?
7. Do you see the role of the archives changing in relationship to data curation? At what point do you think archivists might need to be involved?

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1.800.678.6251.

If you have any questions or concerns, please contact us using the information below.

Sincerely,

## Follow-up survey Instrument with responses

Our Institutional Review Board requires that respondent identifiable information not be included, therefore where this appears in the survey results below, we have replaced it with [redacted].

### 1. Would you be willing to share your collection development policy with the researchers? If so please provide link or attach to response email:

#### *Respondent 1*

Link to the University Archives Faculty Papers Collecting Policy: [redacted] Work is being done on a collecting policy for the Repository.

#### *Respondent 2*

[redacted] We wrote this Collection Development Policy in 2009, it should be updated again.

#### *Respondent 3*

[redacted]  
[redacted]  
[redacted]  
[redacted]

#### *Respondent 4*

University Archives collection Development policy is online at the bottom of the “Plan Your Visit” box here: [redacted]

#### *Respondent 5*

Our collection development policy is available here:  
[redacted]  
At this time, the policy still reflects an analog-focus.

#### Donation of Faculty Papers:

[redacted]

We do specifically request the following be considered for donation, but do not mention what format they may come in:

Research files:

Outlines

Research designs

Raw data  
Notes  
Analyses  
Report of findings

**2. If you are collecting data, how have you done so, in what format?**

*Respondent 1*

Email, digital photos, MSOffice files, PDF's.

*Respondent 2*

We have begun collecting business and administrative data from university offices. The first is four electronic record series created by the Registrar's Office (RO) and now managed in large databases. We worked with RO staff to identify the data fields to be archived and now have a scheduled data transfer. For more details please see [redacted]. We are now working with other business offices to identify other data and reports to be archived using a similar process.

*Respondent 3*

Scattered, usually through counters and interviews.

*Respondent 4*

We are mainly engaged in collecting faculty research data, but we also have some data from [redacted]. For Archives most of this is hardcopy raw data survey forms, some in microfilm. In cases where faculty have electronic data we are referring those folks to our digital repository operation. Archives is not really involved in development of the repository, although we have recently been invited to upload and describe materials from our collections.

*Respondent 5*

At this time, we have received data rather haphazardly, for the most part, with faculty papers as part of the University Archives. The data is normally on flashdrives, external hard drives, CD or in printed form.

**3. If you are collecting research data, how are you appraising it?**

*Respondent 1*

See the faculty papers collecting policy. The Repository will accept digital research data if the faculty, center, institute, etc. requests support for the management and preservation of the data.

*Respondent 2*

We have collected research data for years in analog formats, spreadsheets, photos, slides, publications, etc. We appraise research data in terms of the reputation and impact of the researcher/research and the potential historical value and future research value. Our challenge is to adequately describe and catalog the research data and “market” to the research community.

*Respondent 3*

Establishing the need for specific subject collections.

*Respondent 4*

Mostly by faculty person recommendation. We haven't been able to evaluate much of this material.

*Respondent 5*

At this time, we simply follow our traditional guidelines for the review of faculty papers. At this point in time, we do not store it on a server, nor provide electronic access.

**4. What are the differences between collecting faculty papers and faculty research data?**

*Respondent 1*

Two different stewards. Special Collections is the steward for faculty papers. Faculty Research Data falls under the aegis of the Repository (both are ingested into the [redacted]). Different appraisal criteria (see #3).

*Respondent 2*

This is not too big a distinction at [redacted] Archives. When we engage with faculty about donating their papers, we are including research data in the discussion. We do have space constraints, like all archives, so we try focus on original research in the disciplines identified in our collecting policy. We do have medical research data in the collections which require us to “scrub” personal identification before making accessible.

*Respondent 3*

Faculty papers involve correspondence, authored pieces, minutes of meetings, syllabi, course notes, etc. Faculty research data is that data faculty gather as part of a dedicated research project over an extended period of time.

*Respondent 4*

Papers are often self explanatory but the data requires interpretation (i.e. code books) and documentation of the context: normalization and calculations performed against it.

*Respondent 5*

I see no difference between data and traditional research materials, other than size, format, and issues for long-term electronic preservation.

**5. Do you feel your archives is effectively engaged in the discussion about data curation?**

*Respondent 1*

Yes. We serve on the University-wide Institutional Data Coordinating Committee and have input on faculty research data through the [redacted].

*Respondent 2*

Definitely. The University Archives is separate from the Libraries and reports up through central IT and the university "CIO." This gives our unit high visibility in the data curation discussion. Our recent [redacted] grant to archive vital university data (records) has also helped. The University Archives director is part of the university-wide Research Data Task Force convened by the Provost's office. One of our electronic archivists is on a library-based Research Data working group and presents frequently to faculty and graduate students on research data management. The University Archives and [redacted] Libraries are also working together to help faculty preparing funding proposals that require data management plans (NSF, NEH, etc.).

*Respondent 3*

Yes, absolutely.

*Respondent 4*

No. The digital repository folks have grant funded collaborations with certain faculty in Anthropology but we're not part of that work.

*Respondent 5*

No, we have been left out of the discussion entirely.

**6. What direction do you think your archives' participation in data curation will take?**

*Respondent 1*

Continue to be involved in the development of the systems and tools necessary to manage born-digital materials, assisting with differentiating University records and faculty personal papers from faculty research data, advising on appraisal, continue providing practical experience in data curation for archives students in the School of Library and Information Science, and increasing our outreach for and acquisition of born-digital materials.

*Respondent 2*



I think the University Archives will continue to take a leadership role in data curation at [redacted]. We have the skills and experience (and the need!) to address data archiving and contribute to the overall curation of data. At [redacted] the University Archives is part of an overall reorganization of central IT services and will become even more embedded in the IT structure of the university.

*Respondent 3*

Establishment of an institutional repository.

*Respondent 4*

I think we will eventually play in an important role in advocating for public access and secondary uses of data.

*Respondent 5*

I think we will expand our collection development policy to include various data sets as a format. We have also tried to serve as a resource for those who have been assigned responsibility for data curation.

**7. Do you see the role of the archives changing in relationship to data curation? At what point do you think archivists might need to be involved?**

*Respondent 1*

More evolving than changing. Still need to provide space (digital) and services to help creators manage their digital data. However, the archives needs to be involved earlier in the records lifecycle and be part of the discussion when new enterprise systems are being implemented.

Archivists should be at the table from the beginning when institutions embark upon data curation activities, regardless of the types of data being brought in. They can offer good insight on appraisal, access, acquisition, and digital preservation and complement the digital initiatives librarians and IT staff that generally are part of the process.

*Respondent 2*

Yes, definitely! Personally, I hope more archivists start taking a leadership role in data curation.

As archivists we need to be involved in the curation of data so we can help identify data that should be archived upfront and ensure that the data is managed so that it CAN be archived. If we don't have a voice in data curation, we'll have more difficulty archiving digital records and marginalize our profession even further. Archivists can learn from the other information professions focusing on data curation and have skills and experience to add to the discussion.

*Respondent 3*

Archivists need to be involved from the very beginning in terms of discussions, planning, and implementation. Archival work will be shifted to seeking out and acquiring a different form of archival record, as well as guiding faculty in uploading material, and then embrace use of this data by students and scholars as a normal part of reference work.

*Respondent 4*

Archivists need to be more involved with negotiation of the gift so that agreements are signed, copyrights are secured if possible, privacy is protected and context is preserved. All of this appears to be new territory for the digital repository folks. Faculty tend to think of this in terms of IRB requirements, agreements and restrictions. It seems the IRB releases are often very narrow and no one is advocating and negotiating for public access and/or secondary uses of the data that enable research replication or creative reuse.

*Respondent 5*

I think archivists need to be involved right now. We have had several librarians assigned to data curation here in our Library, and one of them was amazed that we were not included, since faculty papers often include this kind of material.