6-16-2016

Selection and Appraisal of Digital Research Datasets

Chris Eaker
ceaker@utk.edu

Recommended Citation

Eaker, C. (2016). Selection and appraisal of digital research datasets. In Kellam, L. & Thompson, K. (Eds.). Databrarianship: The Academic Data Librarian in Theory and Practice. American Library Association.

CHAPTER 15

Selection and Appraisal of Digital Research Datasets

Christopher Eaker

> *"Data is the currency of science …. To be able to exchange data, communicate it, mine it, reuse it, and review it is essential to scientific productivity, collaboration, and to discovery itself."*[1]
>
> ~ Anna Gold

ACADEMIC RESEARCH INSTITUTIONS are adding a new type of resource to their collections in today's data-intensive research environment. Historically, libraries have collected physical materials. More recently, libraries have purchased datasets for secondary use. And now, libraries are accepting digital research datasets for archiving. This move has become important as both public and private grant funding agencies are requiring that data from funded research be made publicly accessible and preserved. Emphasizing this, Paul Uhlir calls on academia to consider datasets as valuable assets that should be preserved rather than as disposable by-products of research.[2] Digital research datasets are different from the traditional items libraries historically have collected, and, therefore, have additional characteristics to consider when deciding whether to archive them.[3] How does an institution decide which datasets to accept into its collection? This chapter introduces selection and appraisal criteria and policies to help institutions answer that question.

The need for selection and appraisal policies lies upon the assumption that since scientific research is producing ever-increasing volumes of data, it is impossible to preserve it all.[4] Even if it were possible to preserve every dataset, some datasets do not need to be preserved.[5] For example, data generated by climate models are often discarded since they can be easily recreated by re-running the

model. Furthermore, accuracy improves from one generation of a model to the next, so the data generated by an earlier generation is less precise than data obtained from a more current version. On the other hand, climate data gathered during a one-time event, such as a hurricane, cannot be reproduced and is, therefore, more valuable. For this reason, institutions need policies and criteria governing which datasets will be preserved and how to appraise them against those criteria.

Selection and appraisal policies are commonplace in academic libraries for traditional materials and even other digital collections, such as photographs or audio/video materials, but few have complementary policies for datasets. This chapter will introduce and explain the range of selection criteria institutions may consider when developing selection policies, such as scientific or historical value, scarcity, relevance to institutional mission, and others. Lastly, it will discuss life cycle management of datasets, including the periodic refreshing of files and determining when to deaccession, or remove, datasets from the collection. Readers will find an overview of the issues surrounding selection and appraisal of digital research datasets and will be equipped with the knowledge and resources to develop such policies in their institutions.

Definitions

Data, as defined in the introduction, are electronic files of information that have been collected systematically and structured to serve as input for research. For purposes of this chapter, the term *dataset* is defined as one or more files containing multiple data observations. *Metadata* are additional information accompanying a dataset that provides contextual information. Metadata takes the form of descriptive information about the project and its data and is necessary for discovery and crucial for reuse. A *selection policy* is a policy containing specific criteria outlining which types of materials will be accepted into an institution's collection. Selection policies may specify preferences for materials of certain subject matters over others. It may also include criteria pertaining to author/creator; for example, materials created by authors affiliated with a specific institution could be given higher priority. An *appraisal policy* is a policy that outlines a set of processes and procedures to determine if a specific resource meets the selection criteria and will be archived. Appraisal procedures may include tools such as checklists or decision trees, which are discussed in more detail later. The selection policy is the "What?" while the appraisal policy is the "How?"

A Brief History of Collection Development

Collection development of library resources, however rudimentary, has happened since libraries existed. Early libraries served mostly as places of preservation, not dissemination, of information. Today's libraries, however, exist not to keep information in a vault, but to share it with their user communities. These user communities, specifically their interests and characteristics, determine which resources libraries acquire. The term *selection* was originally applied to the process by which librarians determined which items to purchase for the library's collection. Later, in the 1960s, the term *collection development* emerged and implied more than simply selecting items to acquire, but also studying user communities and designing a collection to meet their needs. Even later, in the 1980s, the term *collection management* was introduced, which implied a life-cycle approach to managing a library's collection. Tasks ranged from studying the user community, selecting items to meet their needs, and periodically evaluating the items in the collection to determine whether they should continue to be supported or removed from the collection, a process known as *weeding*.[6] Appraisal and selection of digital materials is closely related to collection management in that files must be managed throughout their lifecycles, from ingest to disposition. It also contains aspects of collection development, as the choice of which datasets to obtain is, in part, related to how they will serve the user communities for which they are intended.

Developing Selection and Appraisal Policies

Development of selection and appraisal policies involves both practical and conceptual questions. Institutions may decide that digital datasets are covered under the traditional collection development policies already in place. However, digital datasets often have different requirements that need to be considered, such as intellectual property issues and the need for greater metadata about the projects that created the data, the data creators, and the data themselves. It is for this reason that Harvey says traditional selection and appraisal criteria cannot be applied to digital materials without some level of modification.[7] Thus, institutions may want to create separate policies regarding datasets because of the additional needs and different characteristics they have when compared to traditional materials. This determination must be based on what is best for the institution's operations.

The different intellectual property issues surrounding datasets often arise because the result of a research project may be a patentable product or procedure or a copyrightable work. The data generated in these projects support those copyrights and patents and are considered intellectual property. Intellectual property issues are discussed later in *Monetary Value*.

Additionally, datasets must have a high degree of metadata to make them useful for reuse. The data creators must provide information about the purpose of the project; the data collection, processing, and analysis techniques; the meaning of variable names; and the spatial and temporal coverages of the data. This information is crucial for data to be meaningful to researchers not originally involved in the research. Metadata are discussed in more detail later in *Level of Documentation*.

Selection Policy Criteria

Selection policies include criteria related to a dataset's content, authorship, value, and potential reuse to determine whether or not a dataset should be obtained for an institution's collection. Ideally, the institution evaluating the dataset for archiving should make this determination in coordination with the original data producer, who has the most intimate knowledge of the data. In their report "How to Appraise and Select Research Data for Curation," Whyte and Wilson[8] provide several criteria to consider when evaluating a potential dataset. Their criteria include relevance to institutional mission, value (scientific, monetary, and historical), scarcity, level of documentation, and readiness for redistribution. These criteria are also identified in Harvey's chapter "Appraisal and Selection" in the Digital Curation Centre's *Digital Curation Manual*, but Harvey goes further and adds vulnerability as an important criterion.[9] Institutions must evaluate each of these criteria and assess how each will be applied to archiving datasets at their institution.

Relevance to Institutional Mission

Relevance to institutional mission should be considered the primary primary criteria for judging which datasets are to be preserved. If a dataset does not serve an institution's mission, then its archiving should be considered a lower priority. For example, if an institution focuses heavily on marine sciences, then archiving a dataset about the effects of climate change on deepwater fish species would be a higher priority than archiving a dataset about deciduous trees in Kansas. Additionally, datasets produced by the institution's own researchers are relevant to the institution and may be considered a higher priority than those produced by researchers at another institution. Last, the institution's goals in archiving datasets is one more element of institutional relevance. In other words, if an institution's goal in archiving datasets is to support education of its students, then datasets that support educational opportunities are considered higher priorities than those that do not.

Institutional relevance also applies to data centers. Data centers archive datasets based on their mission, which is often dictated by their funders. For

example, the Inter-University Consortium for Political and Social Research primarily archives datasets useful to social science and political science researchers.[10] Likewise, the Oak Ridge National Laboratory Distributed Active Archive Center focuses primarily on archiving datasets related to NASA's Terrestrial Ecology Program.[11] These data centers' selection policies are discussed in more detail later.

Value

Once it is determined that a dataset supports the needs of an institution's constituents, a further determination must be made based on a dataset's value. Value can be defined in one or more ways, including *scientific*, *historical*, and *monetary*.

Scientific Value. One primary purpose of archiving a dataset is to foster reuse in scientific research. Reuse of datasets serves two purposes: 1) supporting the replicability of the initial study and 2) supporting the creation of new knowledge beyond the data creator's original intent. Reproducibility of research allows for greater transparency and accountability, thereby increasing research integrity.[12] However, reusing research data can be challenging. For example, locating suitable data to answer a potential user's specific research question may be difficult, and even if a potential dataset is located, it may lack enough documentation to be useful.[13] Even so, Paul Uhlir indicates that the value of a dataset increases as it is reused for new research.[14] In other words, as more research is conducted with a dataset and new findings are discovered, its inherent value increases. In determining if a dataset should be archived, institutions must decide its potential and readiness to serve future, even unintended needs. This concept, called *analytic potential,* is determined by two main components: a dataset's fit for serving research outside the original field and its readiness for preservation.[15]

A dataset's potential to serve researchers in communities other than the originally intended community must be evaluated at the point of archiving. Which potential user communities may be able to reuse this data? Which potential research questions might this dataset be able to answer? The answers to these questions may be difficult to anticipate, but should be evaluated by the institution and the original data creator. Furthermore, to support future reuse, datasets must be fit for this purpose, which means the dataset is verified as high quality and is accompanied by adequate descriptive metadata. Quality assurance builds trust for the dataset and encourages its reuse in new research.[16]

Historical Value. Similarly to acquiring physical resources with scientific value, datasets having historical value may be considered a high priority for preservation.[17] Datasets with historical value may include those from research projects that were significant in scope or were especially groundbreaking. The Data Center is an example of how archiving historical data is central to an institution's mission. The

Data Center collects and maintains data related to southeastern Louisiana. After Hurricane Katrina in 2005, The Data Center began a special collection of data related to the storm's damages and the region's subsequent recovery.[18] Archiving of these datasets is part of preserving the record of a historical event, and thus their preservation might be considered a higher priority.

Monetary Value. Some datasets have monetary or commercial value and may support intellectual property, such as copyrights or patents. These datasets might be considered a high priority for preservation. Indeed, many institutions have policies governing the length of time a dataset must be preserved if it supports intellectual property. These datasets should be preserved for at least the life of the patent or copyright they support. However, in situations where data support intellectual property with monetary value, it may not be feasible to make them publicly available. Thus, the goal in these cases is simply preservation of the datasets rather than sharing.

Scarcity and Irreplaceability

Datasets that record a one-time occurrence are more valuable since the collection is unique and cannot be reproduced.[19] One example of irreplaceable data is data collected during the Deepwater Horizon oil spill in the Gulf of Mexico in 2010. These data were collected during an event that occurred once (we hope!) and can never be collected again. A dataset's scarcity also affects its historical value, thereby adding to the priority in the archiving decision. More commonplace situations where the scarcity principle governs are in the daily recording of weather conditions. Today's weather will not be repeated. In these cases, the institution should preserve the datasets and be careful to maintain accessibility.

Level of Documentation

As mentioned earlier, clearly described datasets are better suited to serving future research, both in the original discipline and in other disciplines.[20] The question must be asked: Will another researcher to be able to make sense of the dataset to reuse it for his or her research? If the answer is no, the institution should determine if it should supplement the metadata on its own. This may be possible in limited cases when the dataset meets other important criteria for acceptance. For example, if the dataset is valuable and cannot be replicated easily or at all, then the library may decide to accept the dataset and create the necessary metadata on its own. For example, when processing a submitted dataset for archiving, Oak Ridge National Laboratory Distributed Active Archive Center creates a document detailing the dataset. The level of documentation provided by the data creator varies, and when that is found lacking, DAAC staff will read

publications that used the data to extract information important to reusing the dataset, such as procedures and variable used, which they will describe in the documentation. However, staffing levels and time constraints may prevent the institution from augmenting the metadata. Datasets initially provided with a high level of metadata should be considered higher priority than those without it.

Readiness for Redistribution

Datasets that contain sensitive information, such as medical information or locations of sensitive species, must be given extra attention to determine if this information has been properly and thoroughly removed.[21] Datasets that have been properly prepared for sharing may be given higher priority than those that have not. However, even if a dataset has been properly anonymized, it should not be made publicly available if the human participants in the research were not informed of this possibility when they agreed to participate in the research. This dataset must be embargoed unless other agreements or situations supersede the informed consent.

Vulnerability

Harvey adds vulnerability to the list of criteria to assess when making the determination of whether to archive a dataset.[22] Vulnerability is determined based on special requirements to read or access the data or the condition and age of the media on which they reside. For example, if the dataset requires special hardware or software to be accessed, it is considered vulnerable. The institution must determine if the additional cost to provide special access to the dataset outweighs the financial benefits of providing that access. Similarly, if the data are on aging media (over 15 years old) or on obsolete media, such as a 5.25-inch floppy disk, they are considered vulnerable. In situations where the data reside on vulnerable media, the institution may decide it is important to transfer the data from that media onto more stable media in order to provide continuing access.

Economic Viability

One final criterion for determining whether or not a dataset should be archived is the costs associated with preparation, archiving, and maintaining those data. Maintaining data accessibility over the long term is not free. Thus, institutions should estimate all associated costs and weigh them against the potential value of maintaining the dataset for future reuse. Costs should be estimated for items such as the storage space for hosting the data, the storage space for providing geograph-

ically dispersed backup copies, labor costs for periodic reappraisal and refreshing, and costs to serve up the data when requested.[23] The UK Data Service has produced a useful tool for estimating costs associating with managing and archiving research. It outlines eighteen activities with questions to consider and suggestions about how those activities add cost.[24]

Appraisal Policies

The selection policy must be accompanied by an appraisal policy that identifies procedures to follow each time a dataset is considered for preservation. These procedures reduce redundant activities, eliminate subjectivity, and improve the efficiency of the appraisal process. Appraisal policies should also include lifecycle management tasks, such as periodic refreshing, reappraisal, and deaccessioning. Those processes, as well as two helpful appraisal tools, are discussed below.

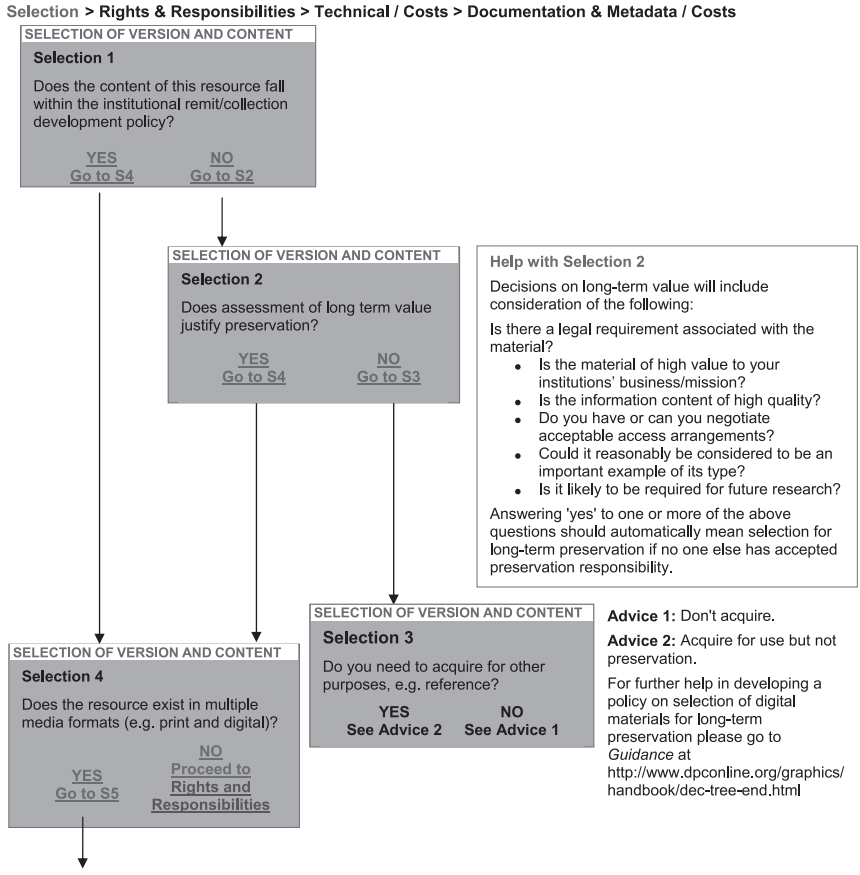
Checklists

Checklists help determine whether or not a dataset should be preserved by providing a clear outline of selection criteria. Checklists also reduce oversight by providing a list of all tasks that must be completed before an archival determination is made. The checklist should include all important items, and the institution's personnel must go through the list systematically to verify each item. Once the checklist has been completed, a determination can then be made whether or not to preserve the dataset based on the information the checklist provides. Does the dataset meet enough of the criteria on the checklist to warrant preservation? If so, it can be preserved. If not, the dataset can be rejected or sent back to the data producer for more information, if necessary.

Decision Trees

Decision trees are another helpful way of determining whether or not to preserve a dataset. Decision trees provide a graphical representation of a logical progression of thought. Questions are presented with possible answers. Each answer then leads to another possible question and set of answers. Once a path of questions and answers is exhausted, the end result is an answer of "YES, this item should be preserved," "NO, this item should not be preserved," or "MAYBE, this item may be preserved, but more information is needed." Decision trees take the guesswork out of appraisal and provide an objective answer rather than relying on sometimes subjective criteria. An example of a decision tree developed by the Digital Preservation Coalition is shown in Figure 15.1.

Figure 15.1. Example of an Appraisal Decision Tree



(Used with permission from Neil Beagrie, Maggie Jones, and Digital Preservation Coalition. *Preservation Management of Digital Materials: The Handbook*, 2008, http://www.dpconline.org/pages/handbook/)

Periodic Refreshing, Reappraisal, & Deaccessioning

Periodic reappraisal of preserved files serves two purposes: 1) to determine if the datasets are still accessible; and 2) to determine if continuing preservation is warranted. Long-term accessibility of digital files can be a challenge if steps are not taken to periodically check the files for file format obsolescence. If the file format is in danger of becoming obsolete, it should be refreshed to a newer format as necessary. However, in cases where file formats are periodically refreshed, it is important to remember that some special formatting may be lost in the process. In these cases, it must be determined whether or not saving the content is the most

important goal, to the detriment of the formatting. File format obsolescence happens when the file formats in which the data were created are no longer supported by modern computer software and hardware. In these cases where file formats may become obsolete over time, institutions must migrate these data to newer file formats to maintain accessibility. To avoid format obsolescence, ideally the data should be initially archived in a preservation-friendly file format, such as .TXT for documents with textual information and Comma Separated Values (.CSV) for documents with tabular numerical information.

Likewise, files should be periodically checked for data degradation, also known as bit rot. Over time, physical storage media degrade, thus preventing access to the files stored on them. This degradation is a breakdown of the electrical, optical, or magnetic properties of the storage media, thus causing them to lose their ability to hold the digital information. To avoid this potentiality, institutions must continually check the integrity of the physical media and upgrade to new media as necessary.

In addition to maintaining accessibility, institutions should periodically reappraise to determine whether to continue preserving a dataset. Similar to a collections librarian weeding his or her collection based on changing usage patterns and user needs, libraries may find that they no longer need to maintain access to a particular dataset. Further, it might be discovered that there is a breach of confidentiality or a legal issue surrounding its continued archiving. This process of removing datasets from a collection is known as deaccessioning. There should be a process by which a dataset's need for continuing preservation is assessed. If it is determined that it is no longer necessary to maintain, it can be deaccessioned from the collection. Even in those situations, institutions may consider simply hiding the dataset from public view instead of permanently deleting it from the server, as a situation may arise when the dataset needs to be re-accessioned. In any case of deaccessioning, the archive should provide a note at the dataset's usual location explaining why it was deaccessioned and who to contact for more information or access.

Examples of Selection and Appraisal Policies

To demonstrate the types of selection and appraisal policies that currently exist and how they utilize the criteria previously explained, the following are descriptions of policies at data centers and academic institutions. While these examples are primarily from data repositories or data centers, they are meant to provide an example of the types of policies an academic institution might implement.

*Oak Ridge National Laboratory Distributed Active Archive Center (*http://daac.ornl.gov/*)*

The Oak Ridge National Laboratory Distributed Active Archive Center (DAAC) is a NASA-funded data archive that archives datasets about the interactions between the biological, geological, and chemical components of the Earth's environment from NASA's Terrestrial Ecology programs and projects.[25] Datasets in good condition when they are submitted to the DAAC, meaning they do not require much effort to prepare them for archiving, are given a higher priority than those that require extensive work or for which the data producers are less responsive to the DAAC's requests for information.

In addition to terrestrial ecology data, the DAAC will archive datasets associated with manuscripts. Many publishers now require data associated with their publications to be archived in a publicly accessible place. The DAAC provides a place for datasets related to its mission to be publicly accessible and, once archived, provides a Digital Object Identifier to the publisher for inclusion in the article. Last, the DAAC will also archive datasets recommended to it by its User Working Group, which is made up of researchers and scientists from other data centers and from universities across the United States. These datasets are given the lowest priority for archiving.

*Inter-university Consortium for Political and Social Research (*http://www.icpsr.umich.edu/*)*

The Inter-university Consortium for Political and Social Research (ICPSR) is a data archive that seeks to preserve and provide access to social science research data. In particular, datasets that show "demonstrated importance" to the research community are given high priority.[26] Emphasis areas may change over time, but currently include datasets that are interdisciplinary, complex, and focused on cultural diversity. These datasets are checked against the selection criteria, which include availability of data, confidentiality concerns, data quality, data documentation, and data format. Datasets that meet its selection criteria are curated and made available to the research community.

*The Odum Institute (*http://www.odum.unc.edu/*)*

The Odum Institute hosts a data archive based at the University of North Carolina at Chapel Hill. Similarly to ICPSR, the Odum Institute hosts significant social science datasets and provides access to them via the Dataverse network of repositories. The archive solicits datasets and then determines if they meet their selection criteria by using an appraisal checklist.[27] The Odum Institute's criteria include in-

stitutional mission and different aspects of the dataset's value, namely its scientific value, historical value, and monetary value. In addition to value, the archive looks at the level of documentation and whether it is complete and readable.

Based on the results of the appraisal checklist and the dataset's current level of documentation and preparation, the dataset undergoes varying levels of processing. Minimal processing is conducted on datasets that come to the archive with a high level of documentation already completed and in the preferred data format (SPSS or Stata). Datasets that have a lower level of documentation may undergo a high level of processing called "routing processing." Datasets considered especially important or valuable or are part of multi-site or multi-year studies may undergo a high level of processing called "intensive processing."[28] In these cases, the archive has determined the extra processing required to prepare the dataset for ingest is warranted due to its high value.

*University of Minnesota Libraries (*https://www.lib.umn.edu/*)*

The University of Minnesota (UM) Libraries' data repository requires that at least one of the dataset's producers be a researcher at that university.[29] It also requires data depositors to make sure their datasets are prepared properly before submitting them to the repo7sitory. Proper preparation includes providing files in a preservation-friendly file format, providing an adequate level of metadata, and ensuring any sensitive information has been removed. Datasets are expected to be open access once deposited, and all go through a curatorial review process before submission to ensure compliance with the selection criteria.

Conclusion

After an extensive search for samples of selection and appraisal policies and requests for written policies from colleagues at academic libraries, it became clear that it is far more common for data centers to have publicly accessible, written policies than it is for academic libraries. Academic libraries have only recently begun archiving datasets as a part of their collection, and many still do not have data repositories in place. Those having data repositories may be processing ingest of datasets on a case-by-case basis without written policies. Even so, it is important to be proactive rather than reactive. Having a policy in place before researchers require services would help eliminate confusion and uncertainty. An additional benefit of establishing selection criteria and appraisal procedures is that academic institutions, namely academic libraries, will have a clearer understanding of the types of datasets they want in their collections. This clarity will help when determining how to promote their repository services by identifying where to focus outreach efforts. Especially now that almost all federal granting agencies and many

private granting agencies require the results of research they fund to be publicly accessible, data archiving services are becoming increasingly important and valuable services. Having well-described, trusted datasets in its institutional repository will improve not only an institution's reputation, but also the reputations of its researchers.[30]

---

1. Anna K. Gold, "Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians," *D-Lib Magazine* 13, no. 9/10 (2007), doi: 10.1045/september20september-gold-pt1.
2. Paul F. Uhlir, "Information Gulags, Intellectual Straightjackets, and Memory Holes: Three Principles to Guide the Preservation of Scientific Data," *Data Science Journal* 9 (2010):ES1-ES5, doi: 10.2481/dsj.Essay-001-Uhlir.
3. Ross Harvey, "Appraisal and Selection," *DCC Digital Curation Manual*, January 2007, http://www.dcc.ac.uk/resources/curation-reference-manual.
4. Angus Whyte and Andrew Wilson. "How to Appraise and Select Research Data for Curation," *DCC How-to-Guides* (Edinburgh: Digital Curation Centre, 2010), http://www.dcc.ac.uk/resources/how-guides/appraise-select-data.
5. Ross Harvey, "Appraisal and Selection."
6. Peggy Johnson, *Fundamentals of Collection Development and Management* (Chicago: ALA Editions, 2014).
7. Ross Harvey, "Appraisal and Selection."
8. Angus Whyte and Andrew Wilson. "How to Appraise and Select Research Data for Curation."
9. Ross Harvey, "Appraisal and Selection."
10. Inter-university Consortium for Political and Social Research, "ICPSR Collection Development Policy," accessed April 27, 2015, http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/policies/colldev.html.
11. Les Hook, personal communication regarding ORNL DAAC selection policies, 2014.
12. Louise Corti, Veerle Van den Eynden, Libby Bishop, and Matthew Woollard, *Managing and Sharing Research Data: A Guide to Good Practice* (London: Sage Publications Ltd. 2014).
13. Ibid.
14. Paul F. Uhlir, "Information Gulags, Intellectual Straightjackets, and Memory Holes."
15. Carole L. Palmer, Nicholas M Weber, and Melissa H Cragin, "The Analytic Potential of Scientific Data: Understanding Re-Use Value," *Proceedings of the American Society for Information Science and Technology* 48, no. 1 (2011):1–10, doi: 10.1002/meet.2011.14504801174.
16. Ibid.
17. Angus Whyte and Andrew Wilson. "How to Appraise and Select Research Data for Curation."
18. The Data Center, "Katrina-Related Data," 2015, http://www.datacenterresearch.org/data-resources/katrina/.
19. Ibid.
20. Ibid.
21. Ibid.
22. Ross Harvey, "Appraisal and Selection."
23. Angus Whyte and Andrew Wilson. "How to Appraise and Select Research Data for Curation."
24. UK Data Service, "Data Management Costing Tool," 2013, September 2015, http://www.data-archive.ac.uk/create-manage/planning-for-sharing/costing.
25. Les Hook, personal communication regarding ORNL DAAC selection policies, 2014.
26. Inter-university Consortium for Political and Social Research, "ICPSR Collection Development Policy," accessed April 27, 2015, http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/policies/colldev.html.

27. The Odum Institute, "Appraisal, Content Selection, Acquisition and Processing Policies," accessed April 27, 2015, http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=627.
28. Ibid.
29. University of Minnesota Libraries, "Data Repository for U of M: About the Data Repository." 2014, https://conservancy.umn.edu/pages/drum/.
30. University of California Santa Barbara, "Data Curation and Management," accessed April 27, 2015, http://www.library.ucsb.edu/scholarly-communication/data-curation-management.