

Open Exit: Reaching the End of the Data Lifecycle

Andrea Ogier, Natsuko Nicholls, and Ryan Speer

Introduction

Scientific research data often have a longer lifespan than the project that creates them. This is particularly true when good research data management practice is put into place. Good data management throughout the data lifecycle is essential for successful long-term preservation and sharing, ensuring a long lifespan of use for research data. In addition to good data management, many of us would agree that the importance, impact, and relevance of one's research data often influences the potential long-term value of it—that is that relevance, value assessment, and retention are all closely linked. Yet it remains uncertain whether or not the retention of data increases their inherent value. More fundamentally, do data in the lifecycle smoothly progress from one stage to another without a gap or an exit? Should review, assessment, and evaluation functions for scientific records and data be included at every stage prior to reaching the end of the data lifecycle? These questions and similar inquiries about the lifespan (and 'death' of data¹) have motivated us to investigate a variety of actions involved in curation decisions for data retention or deletion.

In this chapter, we suggest that potential use or retention should be considered by researchers and/or data curators in every phase of the data lifecycle, particularly in the lifecycle “potholes” where the cycle could naturally slow, stall, or end. We argue that identifying and preparing for these points is a vital part of data curation in which long-term value is of central

importance. From a curation practitioner's perspective, value assessment of all kinds of records, including datasets, is a crucial part of appraisal and selection for records management, curation, and collection development. These appraisal and selection activities are the iterative, responsive, and active processes of re-appraisal, weeding, deselection, deaccession, and disposition. These actions are backed-up by a variety of technical, legal, and institutional policy. And appraisal activities should occur throughout the research process and data lifecycle and should be based on criteria rather than on the assumption that the very act of long-term preservation implies value.

In order to advance our understanding of the actions and decisions that adequately safeguard data for future use, we examine a variety of technical, legal, and institutional responses, controls, and resources that influence actions (and the actors involved in these actions) to retain or not retain the data. Three areas provide context for discussion: university records & information management, library collections management, and data curation. University records & information management, hereafter “records management,” has grown out of a concern for records as corporate assets which must be managed according to a specific set of practices set by a local regulatory environment.² Similarly, library collections management (or “collection development”) is understood as a set of routines aimed at adding materials, removing materials, and efficiently finding materials in a library's collection. We believe this comparative exploration, bringing the discourse and practices developed by well-articulated records management and library collections philosophies alongside the formative practices of data curation, will help us identify points in the data lifecycle where curation would (or should) come to an end.

Comparative Exploration

In her article exploring a selection and appraisal framework for digital curation, Jinfang Niu³ adopts a comparative approach based on the processes and theories from the archives and records management communities. We take a similar approach; however, as Niu draws from methodologies aimed solely at selecting digital objects for preservation, we broaden our focus by exploring methodologies aimed at deletion, disposition, and rejection of materials that exist as part of a collection. The distinction is slight, but important; we want to shed light on the diverse interpretation and understanding of how data should progress throughout the lifecycle.

As we approach disposition and end-of-lifecycle issues from the three perspectives (university records & information management, library collections, and data curation), we focus on the following five areas:

- 1) Terminology (usage and interpretation);
- 2) Scope (types, formats, and uses of objects);
- 3) Authority (actors and directives);
- 4) Appraisal Criteria (actions and factors that influence those actions); and,
- 5) Resources (human, financial, and physical space).

Although some existing studies suggest as many as ten criteria for disposition (as it appears in routines of selection and appraisal),⁴ we focus on these five elements not as criteria themselves, but as a basis for comparison in order to determine how items are excluded or removed from collections and archives. Tables 1 through 5, following a brief discussion, will showcase our comparative observation across the three areas.

“End of Lifecycle” Terminology

Beginning with terminology allows us to draw out conceptual similarities and differences across the three areas to get a better sense of accepted definitions. As shown in Table 1, the term ‘disposition,’ which is a key term in records management, refers to a strictly-bounded and regularly scheduled decision-making process where an item is either archived or destroyed.⁵ The term ‘weeding’ used in library collection management, for example, creates the mental image of a gardener removing weeds so that the carefully planted seeds can get more sunlight and rain, aligning these decisions with natural processes.⁶ ‘Selection’ and ‘de-selection,’ link the additive and subtractive collections decisions, just as using ‘appraisal’ and ‘reappraisal’ creates a cyclic decision narrative in the realm of data curation. In this chapter we use the terminology native to the discipline considered in order to tie it more closely to the source material.

Table 1: Comparison in End of Lifecycle Terminology

	University records & information management	Library Collection management	Data Curation
Terminology	Official record; Active/inactive records; Disposition: retention or destruction	Collection; Maintenance; Weeding; Deaccessioning; “Data-driven” deaccession; Deselection	Digital content; Retention; Appraisal/Re-appraisal; Selection/Acquisition; Data transfer/migration; Disposition; Destruction

Scope

The second element refers to the types, influence, and use of objects (whether physical, digital, or combined). Within the domain of records management, library collections, and data

curation we have explored the extent to which disposition decisions (or the lack thereof) about digital objects are based on the methodologies developed for physical objects as shown in Table 2. A comparison of scope across these three areas demonstrates important distinctions between ideal scope and the reality of implementation. For example, records management encompasses all the documentation generated by an organization, but various factors, such as local policies and confidence in the disposition of secure documents, affect the ability of a records program to manage secure digital records in the same way as secure paper records. Library collections management distinguishes between the object (either physical or digital) and the metadata representing that object; discarding the object and discarding the metadata in the library catalog are often two entirely separate processes. For data curation this distinction between ideal scope and factors that limit implementation may also be important where raw data contains sensitive information or is too large to be easily stored. In these situations the metadata may be an important representation for the data itself. What is interesting about data curation is that its scope is expanding at record speed, given the diverse formats of data and types of digital content that even includes research project Web sites, audio and video files, and geospatial information systems.⁷ By contrast, some institutions like the National Oceanic and Atmospheric Administration (NOAA) narrowly defines specific categories of scientific records (original data, synthesized products, and experimental products) as subject to the appraisal and disposition procedure.⁸

Table 2: Comparison in Scope

	University records & information management	Library Collection management	Data Curation
--	--	--------------------------------------	----------------------

Scope	Theoretically embraces all information created by organization; includes any information created in support of the organization's mission or in fulfillment of its legal obligations. Disposition of paper records is often more effective than electronic records management.	Everything the library or archive subscribes to or collects, including provisions for gift and legacy materials: books, journals, digital resources, media, hardware, software, etc., There is a difference between discarding the object and discarding the metadata.	Theoretically everything that researchers generate out of research projects—recorded factual material commonly accepted in the scientific community as necessary to validate research findings.⁹ Decisions are often influenced by types of data, state of data (e.g., raw, primary, analyzed, published) and the sensitivity of data.
--------------	--	--	---

Authority

The third element broadly covers actors who have some control or power over decisions made about the object (see Table 3). These actors may be in the form of people or positions within a larger organization (Records Manager, Collections Librarian) or in the form of policies, mandates, or laws. As an example, distinct characteristics of traditional records and information management approaches frame the death of data as more dependent on human factors than on the analysis of legal requirements. Where official records retention schedules are incomplete, long-term records appraisal must rely on professional judgment. In the world of research data management and curation, where policies are still being formed, this acknowledgement of human decisions above legal agency could legitimize evidence-based data exit strategies.

Table 3: Comparison in Authority

	University records & information management	Library Collection management	Data Curation
Authority	Records schedule;	Collection	Data steward;

(actors and directives)	Retention schedule; Records manager or records coordinator; Legal directives.	Developer/Manager; Librarians; State-level directives; Consortial policies.	Data producer; Repository collection policies; Institutional regulations; Funder directives.
--------------------------------	---	--	---

Appraisal Criteria

The fourth element examines processes enacted upon the object; in this case, the decision to remove an object from a collection or lifecycle. As shown in Table 4, this decision-making process is rather highly developed in records management, which relies on extensive records retention schedules, comprehensive guidance created to identify disposition dates, and instructions for all types and categories of organizational information. Records schedules have admirable specificity, but schedule creators generally privilege administrative need and organizational legal obligations, which might only obliquely apply to the more uncertain environment of research data retention.¹⁰

Library collections weeding schedules are often marked by a concern for resources; as physical/digital space or budgetary resources become scarce, weeding projects are initiated and driven by a variety of criteria. In libraries where space and cost may not be critical issues, weeding projects can be driven by a concern for the ‘health’ of the collection, or a desire for managing the currency of the information.¹¹ Appraisal in data curation has developed to ensure that scientific records and data are usable over time; thus metrics of cost and historical use may not be entirely relevant. Perhaps the most urgent criteria for assessment in data curation is that of compliance; data that contain sensitive information, whether due to personally identifiable

information or represent a security risk (e.g., credit card information), should be managed and disposed with a high degree of care.

Table 4: Comparison in Appraisal Criteria

	University records & information management	Library Collection management	Data Curation
Appraisal Criteria	Criteria include: liability; administrative need; superseded, obsolete, rescinded; time period after event/action.	Criteria include: space; currency; subject; coverage; usage/cost-per-use; duplication (in format or consortial location).	Criteria include: funder ROI; compliance; (repository) collection alignment; scientific/historical/continui ng value of data (in terms of re-usability); quality; integrity.

Resources (Human, Financial, and Spatial)

The fifth element addresses the cost needed to maintain the object within the collection (see Table 5). Apart from large paper records storage operations, records management can be a cost-effective force multiplier for data management: records managers are unique within organizations in that they are responsible for the disposition of information created by others. Libraries may find themselves grappling with a variety of concerns, including the cost of purchasing/licensing collections, the high value of library real estate (location in city/on campus, stacks vs. study space), or the quality of the metadata provided by vendors (where costly staff time may be needed). Like the records management or library collections areas, there is significant cost associated with data stewardship; however, the cost of data curation is still unknown. Recent studies and tools have emerged in Europe from the ‘Collaboration to Clarify

the Costs of Curation' Project (also known as 4C), which aims to emphasize the value of investing in curation infrastructure.¹²

Table 5: Comparison in Resources

	University records & information management	Library Collection management	Data Curation
Resources	Costs of staffing and centralized records management program; costs of staff time and resources for records management tasks within records-creating units	Costs: budget and subscription/purchase models, staffing resources, space resources (physical); digital space counted by numbers of titles/items than by storage size.	Storage/backup costs; Preservation costs; Cost of creating and managing preservation metadata (to ensure discoverability).

Discussion

By focusing on these five elements (terminology, scope, authority, actions/appraisal criteria, and resources), we now summarize processes in use across records management, library collections, and data curation in order to provide insight into practices of planned data retention and deletion.

University Records & Information Management

In the discipline of records management, appraisal for records retention predominantly is concerned with the primary administrative use of information by the creating organization, with a general emphasis on addressing liabilities or inefficiencies associated with ongoing maintenance of the documents by the original creators. The secondary value of information, or the measure of its enduring utility for audiences outside of the creating unit or organization, is also a focus of records retention scheduling. However primary use is often the first concern, and

appraisal approaches associated with records management are notable for relying on authorities more familiar and significant to research administrators than to academic departments or information managers outside of the records profession.¹³ A retention decision from the records realm will rely on formal legal requirements for recordkeeping (when available) and other guidance found in state records retention guidance (when applicable and present), federal statute and administrative law, or on local institutional (e.g., university) policies based on business needs.¹⁴

At its heart, records management is centered upon the idea of the ‘record’ which may be deemed ‘official’ as the product of state business or governance, ‘active’ in that it is considered current, or ‘inactive.’ These categories can affect the retention schedule and disposal method along with the content or coverage of the record. Thinking about research data as an official record can introduce novel approaches to determining how to retain and dispose of data, potentially offering new perspectives from which to address some problematic situations in data curation. For example, considering your local records management approaches to sensitive or confidential datasets may be informative to those developing data management plans or data retention policies; though the majority of research data may not be governed by an externally-mandated retention/disposal schedule, the data could fall under the mandate of other local policies intended to govern information access and security, such as those maintained by Institutional Review Boards or related administrative units. Consulting local or state-level records management policies regarding issues of liability and security could help in answering

questions about data retention and inform the decision to remove a dataset from the curatorial process.

Library Collections

Creating policies and identifying the criteria for removing items from a collection has long been a part of maintaining a healthy library collection.¹⁵ Library collections have grown beyond just the physical; however, in many libraries, physical volume counts still function as a metric for library status,¹⁶ and if appraisal criteria for digital collections exist, they are often based on the same criteria for print-based collections. Library collection processes have identified a variety of criteria for what physical materials to withdraw (or weed) such as: appearance, duplication in other collections, outdated content, and low usage.¹⁷ In most libraries, where space is often at a premium, physical dimensions and shelf space are also a concern; the ever-expanding suite of library services necessitates careful consideration of physical collections.

Like records management, library collections concern both physical and electronic records; however, a library's digital collections, such as e-books, e-journals, and other e-resources, present slightly different concerns. While currency remains an issue for electronic materials, as Mike Waugh et al. note in their discussion of an e-book weeding project conducted at LSU,¹⁸ concerns over physical space and appearance do not apply to digital collections; the e-book weeding project at LSU was based on criteria of currency rather than space. However, physical concerns could easily translate to criteria of financial resources or cost: digital collections are usually hosted by the publisher and provided to libraries on a subscription model. While they don't require physical space within the library, the monetary cost of these resources

could be a critical factor for retention or deselection. Metrics of cost-per-use are emerging as vitally important criteria for assessing digital resources and are figuring into deselection policies and activities, though they are not without significant drawbacks.¹⁹

In addition to concerns over space and cost, criteria for deselection may also be set by membership in consortia or agreements with multi-institutional digital libraries. HathiTrust, for example, uses member institutions' print holdings to determine legal use of in-copyright digital materials; in order for a user at a member institution to gain access to the digital copy of an in-copyright work, their institution must have at one time owned a print copy of the work.²⁰ In this scenario, a physical volume could be removed from the collection without losing access to the digitized copy; however, it may be resource intensive to do so and special care must be taken to ensure that the correct metadata record for the digital copy remains. Similarly, membership in state- or regional-library consortia may affect these decisions. For example, the Association of Southeastern Research Libraries (ASERL) has formed a cooperative print journal retention policy,²¹ and joined the Washington Research Library Consortium in forming a print journal archive.²² In these agreements, libraries agree to retain certain print materials for a specified amount of time (in the case of ASERL, until 2035). Thus, these consortial agreements and memberships influence what can and cannot be removed from the collection.

While considering data as 'just another' library collection may gloss over some of the uniqueness that emerged from the disciplines of data curation and data management, it also presents a history of suitable criteria that could be used to assess research data. The term 'data curation' itself implies a curatorial framework of management; merely uploading data into a

digital or institutional repository is not data curation, nor is it good data management. These collection management processes and criteria, used for decades by librarians to curate and care for physical (and now digital) library collections, could serve as an initial framework for assessing whether data should stay or exit the research lifecycle.

Data Curation

While records management and library collections practices could inform deselection within the data lifecycle, often data retention is only assessed at the end of a project when the researcher, often attempting to comply with funder sharing requirements, determines which data to deposit into an archive. In these cases, the appraisal and de-selection practices within records management and library collections are applied, but only after the object is in its final form. Digital curation, which is defined as “maintaining and adding value to a trusted body of digital information for future and current use; specifically, the active management and appraisal of data over the entire lifecycle”²³ needs to operate *in situ*: before, during, and after the research process. Unfortunately, representing the research process in a lifecycle, implies that the transition is seamless and smoothly progress from stage to stage. However, Carlson argues, “the most critical gap between the stages in a life cycle model is between the stages where the data are actively managed for use by the researcher who developed the data to where the data transition into being curated,” suggesting a divide between data creators/users and curation practitioners in interpretation and understanding of how data should/could progress in the model.²⁴ Carlson’s emphasis on appraisal during the research process is, in many ways, unique to data curation. Could appraising data during the lifecycle could lead to a different outcome when compared to

appraisal at the end of the lifecycle? This is an area that the discipline of data curation should more fully explore.

Current trends in appraisal and selection methods in data curation have been built upon archival appraisal theories and collection development methods over the last decade.²⁵ The term “appraisal” refers to the method to identify digital content’s permanent value for the purpose of long-term preservation, therefore discussions of appraisal in data curation have been closely linked to institutional repository or data archival policies on collection development.²⁶ Appraisal criteria for initial selection decisions in a repository, for instance, function to to maintain alignment with existing collections.²⁷ Early efforts to create data repositories were, largely, focused on a specific discipline or data type.²⁸ The rise of institutional data repositories and large-scale data publishing practices have expanded selection criteria and broadened existing collections beyond collection policies aimed at a specific discipline, data type, or data format. Institutional data repositories, for example, collect, preserve, and give access to the research products of an entire institution, though they often arrange materials by department, college, or institute. Open, web-social repositories like figshare.com and Dryad (datadryad.org) continue to change the landscape of data repository options, allowing a greater variety of data to be accessible openly via the web.

While similar to practices described in the library collection section, data curation focuses on digital contents rather than physical materials. Thus, a digital collection is measured by size and its value can be based on the number of files, datasets, studies, and collections available in the repository. The usage metric for digital collections, namely the number of

downloads, is still emerging as an assessment metric for the enduring value of data, and is used in retention or disposition decisions.²⁹

At the practical level, appraisal in data curation has developed to ensure that scientific records and data are usable over time. This is where, we believe, the two important issues (the value of data and the retention period) intersect and where it is important to address the question: What makes digital scientific records more or less usable? Although we lack standardized metrics to assess the value of data based on its re-usability, there is a recent effort among data stewards to document and compile cases in which their openly shared research data are being re-used by others.³⁰ This idea of reuse fuels the value assessment of data and drives the constantly evolving paradigm of federal-funder return on investment.

Another distinctive characteristic of data curation is the significant role that research communities play in appraising the value of data for long-term retention. In their data management plans researchers may say that every dataset should be preserved for the maximum period of retention (or forever, whichever comes first). We know, however, that due to resource concerns, the rapidly evolving technology environment, and changes in policy and authority, we cannot retain everything—sometimes the best we can hope for is planned obsolescence. From the researcher's perspective, appraisal criteria of scientific records and data should be biased towards relevance, significance, uniqueness, sensitivity, and the impact of their overall research output. These qualities are exactly those criteria at work in both library collections and records management. Communicating these perspectives, and the differences between them, should be a part of every retention and disposal discussion.

Conclusion

Our review of these three disciplines—university records & information management, library collections, and data curation—suggests that there are criteria for data retention and destruction that go beyond a dataset’s projected value over time; Additionally, we advise that anyone involved in deselection decisions also be aware of the local, legal, and disciplinary policies that impact data at each stage in the research lifecycle. While data curation practices may enable data discovery and retrieval, maintain quality, add value, and facilitate reuse over time, perhaps curatorial “value-add” also incorporates the assessment of liability, risk, or resource cost over potential value. In these cases, the curation decision may lead to disposal of the dataset. If the purpose of data curation is to add value at every stage of the research lifecycle, we suggest that this definition includes the consideration of when to exit the lifecycle. However, these decisions cannot be made at too high a level; like records management, the decision to dispose of a dataset must take into account a variety of factors including (but not limited to): content, risk and liability, currency, scope, cost, quality, uniqueness, and external mandate. Not all of these factors will apply to every dataset, but, we believe that these criteria, combined with local practices, will provide a thorough basis for any decisions on when to exit the research lifecycle.

¹ Although we were unable to identify any existing work that solely features the subject ‘death of data,’ we have noticed that subscribers of Research Data Management discussion list, RESEARCH-DATAMAN hosted by JISC, have actively (and in a timely manner for our book chapter) engaged in online discussions about related topics, including: ‘data retention,’ ‘identifying archival material,’ and ‘retention of physical research data.’ Threads on these topics are archived at: <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A1=ind1508&L=RESEARCH-DATAMAN#9> (threads in August 2015), <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A1=ind1509&L=RESEARCH-DATAMAN#12> (threads in September 2015), and <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A1=ind1510&L=RESEARCH-DATAMAN#33> (threads in October 2015).

² Charlotte Brunskill and Sarah Demb. *Records Management for Museums and Galleries: An Introduction*. Oxford: Chandos Publishing. 2012, Chapter 2 in this book provides the definition and practices of records management.

³ Jinfang Niu, "Appraisal and Selection for Digital Curation," *International Journal of Digital Curation*. 9, 2, (2014): 68.

⁴ As suggested by Ross Harvey (2007), ten appraisal criteria include: value, physical condition, resources available, use, social significance, legal rights, format issues, technical issues, policies, and documentation. Niu actually develops her argument on four appraisal criteria: 1) mission alignment, 2) value of digital resources, 3) cost, and, 4) feasibility, Niu, *op.cit.*, 71-2.

⁵ Brunskill and Demb, *op. cit.*, see chapter 7 for more detailed discussion on retention schedule and records management program.

⁶ OCLC has compiled a useful bibliography on weeding and deselection, <http://www.oclc.org/en-US/sustainable-collections/bibliography.html>

⁷ Even from a more established preservation policy framework, ICPSR considers new digital content (e.g., Website, audio, video, GIS) challenges, suggesting that existing policies, procedures, and practices need to be revised or re-engineered to encompass new digital content. See ICPSR Digital Preservation Policy Framework at <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/dpp-framework.html>

⁸ National Oceanic and Atmospheric Administration (NOAA). *NOAA Procedure for Scientific Records: Appraisal and Archive Approval, Guide for Data Managers*. Washington, D.C., 2008.

⁹ OMB Circular A-110 defines data as "the recorded factual material commonly accepted in the scientific community as necessary to validate researching findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues" and has been widely adopted by many federal funding agencies.

¹⁰ For an older but representative discussion of the lightly-documented tension between basic administrative needs and scholarly concerns in university records management, see Don Skemer and Geoffrey Williams, "Managing the Records of Higher Education: The State of Records Management in American Colleges and Universities," *American Archivist* 53, 4 (1990): 544-45.

¹¹ Sharon Leslie and Ida Martinez. "Assessment and Weeding of a Clinical HIV/AIDS Collection in an Academic Library: A Case Study." *Collection Management*, 40, 3 (2015): 151 and 153.

¹² For more information on curation cost, see the project website of Collaboration to Clarify the Costs of Curation (4C) at <http://4cproject.eu/>. For more information on the cost analysis of digital collection, see the project called LIFE (Life Cycle Information for E-Literature), a collaboration between University College London (UCL) and the British Library, <http://www.life.ac.uk/>

¹³ On primary and secondary value for archives and records, see Theodore Schellenberg, "The Appraisal of Modern Public Records," in *Modern Archives Reader: Basic Readings on Archival Theory and Practice*, ed. Maygene Daniels and Timothy Walch (Washington, DC: National Archives and Records Service, 1984), 68.

¹⁴ William Saffady, *Records and Information Management: Fundamentals of Professional Practice* (Overland Park, KS: ARMA International, 2011), 64-65 provides a basic discussion of retention rationales.

¹⁵ Rajia Tobia, "Comprehensive Weeding of an Academic Health Sciences Collection: The Briscoe Library Experience." *Journal of the Medical Library Association*, 90,1 (2002): 94-98.

¹⁶ Kyrrillidou, Martha, Morris, Shaneka, and Roebuck, Gary. "Rank Order Table 2: Titles Held," in *ARL Statistics 2013-2014*, Association of Research Libraries (Washington, D.C.: ARL, 2014), 52.

¹⁷ Mike Waugh, Michelle Donlin, and Stephanie Braunstein. "Next-Generation Collection Management: A Case Study of Quality Control and Weeding E-Books in an Academic Library." *Collection Management*, 40, 1 (2015): 19.

¹⁸ *Ibid.*, 19-20.

¹⁹ Tim Bucknall, Beth Bernhardt and Amanda Johnson, "Using Cost Per Use to Assess Big Deals, *Serials Review*, 40, 3 (2014): 194-196.

²⁰ <https://www.hathitrust.org/out-of-print-brittle>

²¹ <http://www.aserl.org/programs/j-retain/>

²² <http://www.scholarstrust.org/>

²³ The most widely adopted definition of digital curation is provided by the Digital Curation Center, <http://www.dcc.ac.uk/digital-curation/what-digital-curation> and ICPSR, <https://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary#D>.

²⁴ Jake Carlson. "The Use of Life Cycle Models in Developing and Supporting Data Services," in *Research Data Management: Practical Strategies for Information Professionals*, ed. Joyce M. Ray (West Lafayette, IN: Purdue University Press, 2014), 80.

²⁵ Niu, *op. cit.* p.66.

²⁶ 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>

²⁷ *Ibid.*, p.2.

²⁸ Just as early efforts to create data repositories were, largely, focused on a specific discipline, different disciplines have required different approaches to appraisal and disposition. Esanu et al. (2006) and Faundeen (2010) emphasize the importance of disciplinary-specific appraisal criteria.

²⁹ There is a great body of work on download statistics focusing on institutional repositories, including Organ (2006) and Konkiel and Scherer (2013) to name only a few.

³⁰ In February 2016, two open data advocates from Innovations for Poverty Action and Mozilla Science Lab, Stephanie Wright and Stephanie Wykstra, have joined together to document examples of research data re-use from any scientific discipline, <https://www.mozillascience.org/share-your-story>

Bibliography

Anson, Catherine, and Ruth R. Connell, *SPEC Kit 313: E-book Collections*. Washington, D.C.: Association of Research Libraries, 2009.

Association of Southeastern Research Libraries, Cooperative Journal Retention. <http://www.aserl.org/programs/j-retain/>

Blodgett, Peter, Jeremy Brett, Cathi Carmack, Anne Foster, Laura Uglean Jackson, Chela Scot Weber, Linda Whitaker, and Marcella Wiget. *Guidelines for Reappraisal and Deaccessioning*. Society of American Archivists. 2011. <http://www2.archivists.org/sites/all/files/GuidelinesForReappraisalAndDeaccessionin gDRAFT.pdf>

Brunskill, Charlotte, and Sarah Demb. *Records Management for Museums and Galleries: An Introduction*. Oxford: Chandos Publishing. 2012.

Bucknall, Tim, Beth Bernhardt and Amanda Johnson, "Using Cost Per Use to Assess Big Deals, *Serials Review*, 40, 3 (2014): 194-196. doi: 10.1080/00987913.2014.949398

- Carlson, Jake. "The Use of Life Cycle Models in Developing and Supporting Data Services." In *Research Data Management: Practical Strategies for Information Professionals*, edited by Joyce M. Ray, 63-86. West Lafayette, IN: Purdue University Press, 2014.
- Digital Curation Center. Value of Digital Curation. <http://www.dcc.ac.uk/digital-curation/>
- Duranti, Luciana. "Preface to the Special Issue on Data, Records, and Archives in the Cloud." *Canadian Journal of Information and Library Science*. 39, 2 (2015): 91-96.
- Esanu, Julie, Joy Davidson, Seamus Ross, and William Anderson. "Selection, Appraisal, and Retention of Digital Scientific Data: Highlights of an ERPANET/CODATA Workshop." *Data Science Journal*. 3, 0 (2006): 227-232. doi: 10.2481/dsj.3.227dsj.3.227
- Faundeen, John. "Appraising U.S. Geological Survey Science Records." *Archival Issues: Journal of the Midwest Archives Conference*. 32, 1 (2010): 7-22.
- HathiTrust. Access to Out-of-Print and Brittle or Missing Items. <https://www.hathitrust.org/out-of-print-brittle>
- Harvey, Ross. "Appraisal and Selection." In *DCC Digital Curation Manual*, edited by Seamus Ross and Michael Day, Edinburgh, UK: Digital Curation Centre, 2007. <http://www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/appraisal-and-selection>
- Higgins, Sarah. "The DCC Curation Lifecycle Model." *International Journal of Digital Curation*. 3, 1 (2008): 134-140. doi:10.2218/ijdc.v3i1.48
- Inter-university Consortium for Political and Social Research (ICPSR). ICPSR Digital Preservation Policy Framework. <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/dpp-framework.html>
- Kyrillidou, Martha, Shaneka Morris, and Gary Roebuck. "Rank Order Table 2: Titles Held," in *ARL Statistics 2013–2014*, Association of Research Libraries (Washington, D.C.: ARL, 2014), 52.
- Konkiel, Stacy, and Dave Scherer. "New Opportunities for Repositories in the Age of Altmetrics." *ASIS&T Bulletin*. April/May 2003.
- Leslie, Sharon, and Ida Martinez. "Assessment and Weeding of a Clinical HIV/AIDS Collection in an Academic Library: A Case Study." *Collection Management*, 40, 3 (2015): 149-162. doi:10.1080/01462679.2015.1040570

Library of Virginia. Records Management: Retention

Schedules. <http://www.lva.virginia.gov/agencies/records/retention.asp>

Mozilla ScienceLab. Share Your Story of Research Data Re-

use! <https://www.mozillascience.org/share-your-story>

Niu, Jinfang. "Appraisal and Selection for Digital Curation." *International Journal of Digital Curation*. 9, 2, (2014): 65-82. doi:10.2218/ijdc.v9i2.272

National Archives. Strategic Directions: Appraisal Policy. <http://www.archives.gov/records-mgmt/initiatives/appraisal.html>

National Oceanic and Atmospheric Administration (NOAA). *NOAA Procedure for Scientific Records: Appraisal and Archive Approval, Guide for Data Managers*. Washington, D.C., 2008. https://www.ngdc.noaa.gov/wiki/images/0/0b/NOAA_Procedure_document_final.pdf

OCLC. Sustainable Collection Services: Bibliography on Weeding and

Deselection, <http://www.oclc.org/en-US/sustainable-collections/bibliography.html>

Office of Management and Budget (OMB) Circular A-110. "Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations." https://www.whitehouse.gov/omb/fedreg_a110-finalnotice

Organ, Michael. "Download Statistics - What Do They Tell Us?" *D-Lib Magazine*, 12, 11 (2006). doi:10.1045/november2006-organ

Pennock, Maureen. "Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information." Digital Curation Centre and UKOLN, University of Bath. 2007. http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf.

Saffady, William. *Records and Information Management: Fundamentals of Professional Practice*, Overland Park, KS: ARMA International, 2011.

ScholarsTrust. Program Agreement: Association of Southeastern Research Libraries (ASERL) and Washington Research Library Consortium (WRLC) Shared Archives of Print Journal Collections. http://www.aserl.org/wp-content/uploads/2013/02/ASERL-WRLC_Shared_Archive_Print_Journals_AGREEMENT.pdf

Skemer, Don, and Geoffrey Williams. "Managing the Records of Higher Education: The State of Records Management in American Colleges and Universities." *American Archivist*, 53, 4 (1990): 532-47. doi:10.17723/aarc.53.4.x50632185v6j2275

Society of American Archivists (SAA), A Glossary of Archival and Records Terminology: Record. <http://www2.archivists.org/glossary/terms/r/record>

Schellenberg, Theodore. "The Appraisal of Modern Public Records." In *Modern Archives Reader: Basic Readings on Archival Theory and Practice*, edited by Daniels, Maygene and Timothy Walch, 57–70, Washington, DC: National Archives and Records Service, 1984.

Tobia, Rajia. "Comprehensive Weeding of an Academic Health Sciences Collection: The Briscoe Library Experience. *Journal of the Medical Library Association*, 90,1 (2002): 94–98.

Waugh, Mike, Michelle Donlin, and Stephanie Braunstein. "Next-Generation Collection Management: A Case Study of Quality Control and Weeding E-Books in an Academic Library." *Collection Management*, 40, 1 (2015): 17-26. doi:10.1080/01462679.2014.965864

Whyte, Angus, and Andrew Wilson. "How to Appraise and Select Research Data for Curation". *DCC How-to Guides*. Edinburgh: Digital Curation Centre. 2010.