

2020

Review of Digital Curation for Libraries and Archives

Barbara Austen

Connecticut State Library, barbara.austen@ct.gov

Follow this and additional works at: <https://elischolar.library.yale.edu/jcas>



Part of the [Archival Science Commons](#)

Recommended Citation

Austen, Barbara (2020) "Review of Digital Curation for Libraries and Archives," *Journal of Contemporary Archival Studies*: Vol. 7 , Article 5.

Available at: <https://elischolar.library.yale.edu/jcas/vol7/iss1/5>

This Book Review is brought to you for free and open access by EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in *Journal of Contemporary Archival Studies* by an authorized editor of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

Stacy T. Kowalczyk. *Digital Curation for Libraries and Archives*. Santa Barbara, Calif.: Libraries Unlimited, 2018.

The Connecticut State Archives developed a new position, digital records archivist, in 2018 and tapped me for the position, given my twenty-plus years in archives and my experience (although limited) with digital preservation. With a new governor who wants to go all electronic, I need to make sure I am up on my digital preservation/curation skills. I am still learning, so I was sincerely hoping by reviewing this book to expand my knowledge and cement what I have learned so far. Even after numerous workshops, webinars, and voracious reading, however, I found certain segments of the text beyond my needs and certainly beyond my comprehension.

Stacy T. Kowalczyk has written *Digital Curation for Libraries and Archives* as an introductory textbook for students in information, library, and archival science, as well as for working librarians, archivists, and information professionals. It succeeds with the first goal but would be a difficult read for working professionals, who often lack the time to explore all the resources cited in the text, all of which I am sure would be useful.

The author takes what I think is a novel approach to the field by addressing the theoretical background of each chapter's topic, followed by the technical and practical implications, a summary, and then a list of discussion questions for use in the classroom. I found the summaries and practical applications particularly useful in that the technological jargon of the chapter was removed and the thesis fully revealed. Perhaps if the author had put the summary first, readers would be able to plumb from the text the most useful points.

In the first section, which provides a "Preservation Overview," the author draws distinctions between digital curation (managing and enhancing data), digital archiving (assessment and selection, appropriate storage, and ensuring authenticity), and digital preservation (ensuring accessibility over time) (4–5). She admits that those distinctions are not universally used but applies them that way throughout the text. She sees archiving as a subset of curation and preservation as a subset of archiving. That distinction may work in large organizations with extensive staff like universities, but the reality is that many institutions, including mine, have only one or two staff working with digital collections. The "roles" necessarily overlap, making the differences outlined in the text less important.

The author uses strategies and models to explain various steps within digital curation. In general, I found these tools more distracting than helpful. The exception was chapter 2 on "Curation Strategies and Models." It explores the differences in the amount of technological expertise needed and costs to maintain three curation strategies: technology preservation, technology emulation, and data migration or transformation. I had heard of all three strategies, but the author describes them more fully than I had seen before. Technology preservation is the largest drain on resources and includes space and equipment. Few institutions can afford to purchase and maintain a myriad of systems in order to open what are now obsolete files. This is one of the few times the author addresses the costs of a preservation program. However, she does provide information on cost models, focusing on acquisition and infrastructure costs to assess the financial impact of a preservation program on an institution. That information is extremely useful to the administrators

of a program but is quite beyond the decision-making power of most digital curators. It could, though, be useful in making administrators aware of the investment they are making.

The author next addresses lifecycle models. These include research lifecycles, data lifecycles, and digital curation lifecycles. There are two illustrations of digital curation lifecycles, one of which was familiar to me from the Digital Curation Center in Great Britain. I found the research and data lifecycles a bit harder to grasp and would have appreciated illustrations of those systems to help me visualize the context and content. These two lifecycles are specifically addressed to data curation more than data preservation.

The section on “Preservation Technology Fundamentals” succinctly covers the topics of file formats and standards. File format sustainability affects preservation, so selecting the proper format is essential. She covers the Library of Congress’s factors for format sustainability and the differences between lossless and lossy compression. In lossless compression, the size of the file is reduced without sacrificing image quality, while in lossy compression, some data may be lost thereby affecting image quality. She also introduces the important fact that digital preservation includes not just master files but production files and derivatives for access. Capturing an image or document is just the beginning. The last portion of the chapter gives recommendations for formats that are considered archival. I would guess that most readers would want acceptable formats to be addressed sooner.

The chapter on metadata contains information on descriptive, structural, and administrative or preservation metadata, all of which contribute to understanding the context of the digital object. Most of my time is spent on creating metadata so images/data can be easily found. On ingest, our repository creates the Submission Information Package (SIP) and Archival Information Package (AIP). As a result, I found that the detail in this chapter was more granular than I needed as many of the metadata elements are created by the repository into which the digital object is ingested. I do use this information, though, to make sure the repository, a consortia, is delivering what it promised.

Chapter 5, “Data Assurance,” encompasses data quality. Digital curators and digital repositories need to ensure the accuracy, completeness, integrity, consistency, and accessibility of the digital object, and these qualities must be monitored regularly over the life of the object. Quality control assurance includes the use of checksums for integrity to ensure the file has not changed and verification to determine “the level of conformance to the encoding specification of a certain format,” which employs a tool like JHOVE, the JSTOR/Harvard Object Validation Environment, a software framework that identifies, validates, and characterizes digital objects (80). The author provides several examples of workflows that incorporate quality assurance into the preservation process. However, it is this, and the succeeding portion of the chapter on persistent identifiers (PIDs), that was hard to follow with the author cramming acronym after acronym into two pages. Digital curation is filled with “jargon” like OAIS (Open Archival Information System), SIP, AIP, and DIP (Dissemination Information Package), most of which are familiar to those even starting out in digital preservation and are essential building blocks. However, I found myself having to flip through preceding pages or ahead to the resources list to find the meaning of other acronyms the author uses.

One concept in this section I found difficult to grasp was maturity models, which the author states, “help an organization assess its methods, structure, and/or processes against a set of criteria” (92). Kowalczyk discusses data, metadata, and digital collection maturity. More concrete examples of these models would have helped, as the discussion was more theoretical than practical.

The section “Planning for Preservation” addresses topics that did not seem to fit with preservation as I think of it and might have been brought up in the text sooner. If digital curation means managing and enhancing data, then emphasis on managing data, organizing files, and file naming needs to be more central to the text and not relegated to later in the book under “Planning for Preservation.” More pertinent to the preservation section are disaster planning and a chapter on “Digital Curation Assessment and Planning.” Considering the latter, the author assures archivists that tasks such as working with donors, collection appraisal, and determining levels of processing and metadata (description) for digital collections are the same as for analog collections. The central concept in this section is the importance of documenting decisions and the processes used in preparing digital materials. This documentation is perhaps more critical for digital collections than paper-based ones, considering the need for accuracy and integrity in managing the files. It is also something I often fail to follow as much as I should, so it was a good reminder (and remonstrance).

The final section, “Preservation in Practice,” addresses issues concerning personally identifiable information, the challenges of archiving social media, and how to curate research data. The author incorporates recent events such as policing in Ferguson, Missouri, and the Black Lives Matter movement in the chapters on personal data and social media, which makes the issues of privacy more immediate and helps to prove her points. Privacy and ethical issues are paramount in documenting these current and highly emotional events. It is in the chapter on personal data that the author addresses forensic discovery, which again seems to be an odd placement. A more detailed discussion of the pros and cons of forensic discovery and suggestions for when it should be used would have been extremely helpful.

The volume ends with a brief glossary and a handbook for teachers with suggestions for a course schedule and class activities or assignments. The schedule essentially follows the outline of the book chapters. Class activities, often done in groups, include investigating national curation models, exploring file formats, and developing a disaster recovery program, among many others. These activities as part of a class on digital preservation would be especially important to help students grasp complex concepts and unfamiliar terms.

Instructors will find the handbook for teachers and the discussion points at the end of each chapter useful. The discussion questions include such things as:

- What concepts would you include in a mental model of digital curation?
- How could lifecycle models help organizations when thinking about a preservation strategy?
- What is preservation metadata?
- What functionality would be the most important when choosing a repository system?
- How would you define data management for a repository? For a workshop for researchers? For a workshop for the general public.

Each chapter is followed by an extensive list of sources, both national and international, most of them available on the Internet.

As mentioned above, the organization of the chapters into theory, technical aspects, and practical aspects is unique from any textbook on digital curation that I have read. It was helpful for me to get additional background on the theories behind what I do in the area of digital curation. The overall presentation, however, is uneven. There are illustrations for concepts that do not seem to need them, and concepts that need illustrations but are lacking. I was also very disappointed with the editing and/or proofreading. Misspelled words abound, parts of sentences are missing, and some words are missing, all of which make it harder to understand an already complex topic.

The text will be useful to students learning about digital curation when used along with other resources and when they have the time to explore the resources sections. The instructor will need to explain some of the technological information. For someone already working in the field, the theory could be useful but the amount of technical detail may be overwhelming without a lot of hand-holding. I found the Society of American Archivists' series on digital preservation, particularly *Digital Preservation Essentials* (Chicago: Society of American Archivists, 2016), edited by Chris Prom, more helpful for someone like me who needs to learn quickly while actively practicing in the field.