

Revamping the “Difficult (Potentially)” but “Mostly Good” and “Pretty Smooth” Removable Media Workflow at the Bentley Historical Library

Max Eckard
Bentley Historical Library
University of Michigan
Ann Arbor, Michigan 48109-2113 U.S.A.
1-734-763-7518
eckardm@umich.edu

Alexa Hagen
Bentley Historical Library
University of Michigan
Ann Arbor, Michigan 48109-2113 U.S.A.
1-734-647-3864
anhagen@umich.edu

Abstract

Archivists at the University of Michigan Bentley Historical Library recently concluded a project to review and initiate improvements to our five-year-old workflow for transferring born-digital archival materials from removable media to a more suitable preservation environment. This “revamp” proceeded in three phases: 1) a review of our current capture and transfer to backlog procedures, which included the process of gathering feedback on the workflow and associated documentation from Bentley archivists; 2) researching contemporary archives and digital preservation theory and best practices and conducting information-gathering interviews with colleagues at peer institutions; and 3) making a number of recommendations for improvement now being implemented by the project team. After briefly outlining our institutional context and providing an overview of our work with removable media, this short paper will give an overview of the “Removable Media Workflow Revamp” project as well as the ways our “agile” approach to the process permitted us to evolve in response to early feedback and ultimately adapt our recommendations to better align with archivists’ needs.

General Terms

Preservation strategies and workflows; Innovative practice.

Keywords

Digital curation, Workflow development, Archives.

1. INTRODUCTION

Archivists at the University of Michigan Bentley Historical Library (the Bentley) recently concluded a project to review and initiate improvements to our five-year-old workflow for transferring born-digital archival materials from removable media to a more suitable preservation environment. After briefly outlining our institutional context and providing an overview of our work with removable media, this short paper will give an overview of the “Removable Media Workflow Revamp” project. Our process was characterized by a short feedback loop, an iterative methodology and, above all, an emphasis on direct communication. This permitted us to evolve in response to early feedback and adapt our recommendations to better align with archivists’ needs. This opportunity to think

critically about our work was invaluable and we believe this approach can be broadly applied by others seeking to examine their own workflows. It has also allowed us to adapt our workflow to manage the increasing scale of digital archives.

2. INSTITUTIONAL CONTEXT

The Bentley collects and preserves the materials for and promotes the study of the histories of the State of Michigan and the University of Michigan. Founded in 1935, the Bentley’s 11,000 research collections range from the papers of Michigan governors, to the records of student and faculty life at the University, to the entire historical record of intercollegiate athletics at Michigan and much more. These holdings include more than 100 TB of digital content with extensive born-digital archives as well as web archives and digitized collections of print, photographic, and audiovisual materials.

The Bentley has actively managed large collections of born-digital content since the 1997 accession of former University of Michigan President James Duderstadt’s personal computer. As part of its mission, the Bentley is committed to ensuring the preservation and accessibility of this content over the long-term by implementing professional best practices and standards in its workflows and infrastructure. In 2011, staff began to develop a “forensics-light” framework for accessioning and transferring born-digital archives from removable media (i.e., making use of digital forensics techniques and strategies but only creating disk images when significant properties of the source media warrant it). In 2013, influenced by An Inter-Institutional Model for Stewardship (AIMS) project [1] and concurrent work at Stanford University to forensically image removable media using Forensic Toolkit (FTK) Imager software,¹ staff implemented hardware, software, and procedures for a Removable Media Workstation (RMW).² The RMW, a 64-bit system running Windows 7, with a 3.4 GHz processor and 8 GB of RAM, makes use of standard hardware and largely free and open source software to transfer content from various physical storage media to a more suitable preservation environment and digital processing backlog.³ To accommodate growing demand, the Bentley acquired a nearly identical RMW in 2016 and purchased a RipStation, an automated optical media ripping solution, in 2017.

¹ <https://deepblue.lib.umich.edu/handle/2027.42/96441>

² <https://sites.google.com/a/umich.edu/bhl-archival-curation/processing-archival-collections/07-removable-media>

³ <https://sites.google.com/a/umich.edu/bhl-archival-curation/processing-archival-collections/07-removable-media/hardware-and-software>

3. REMOVABLE MEDIA WORKFLOW REVAMP

In 2017, in line with a core operating principle to “review procedures on a regular basis [and] monitor new and emerging technologies” and as part of a larger Curation team goal to “enhance our archival management and preservation infrastructure, increase efficiencies and improve stewardship of collections,” archivists at the Bentley began a project to comprehensively review and initiate improvements to our now five-year-old workflow for transferring born-digital archival materials from removable media to a more suitable preservation environment.

We, the project team, created a Project Charter⁴ to articulate the purpose and goals of this revamp and serve as a reference document for our progress and weekly check-ins. The Charter also outlined project team member roles as well as a timeline and associated checklist. The project proceeded in three phases: 1) a review of our current capture and transfer to backlog workflows, which included the process of gathering feedback on these workflows and associated documentation from Bentley archivists; 2) researching contemporary archives and digital preservation theory and best practices and conducting information-gathering interviews with colleagues at peer institutions; and 3) making a number of recommendations for improvement now being implemented by the team.

3.1 Review

The initial intent for this project was to consider replacing our current workflow and RMWs with the BitCurator Environment.⁵ During the Review phase, we inventoried and reviewed all functions of our current workflow and the tools used to carry them out, identified similar or equivalent tools in the BitCurator Environment, interviewed processing archivists about their experiences using the RMW and associated documentation, and identified pain points and areas for more thorough research.

3.1.1 Methodology

To better understand our workflow, we created an inventory spreadsheet that listed all tools used throughout the removable media transfer process and their functions. We then mapped these functions to tools available within the BitCurator Environment as we considered its adoption and how it might impact our processes. Additionally, in an attempt to be inclusive in our review, we developed a brief set of questions for interviewing processing archivists about their experiences using the RMWs. The interview covered topics such as how often they use the RMWs, whether the documentation and tools meet their needs, and pain points in the workflow, as well as providing an opportunity to offer feedback and suggestions.

3.1.2 Findings

The responses from our interviews with processing archivists about their experiences using the workflow and RMWs were generally positive (e.g., “mostly good”, “pretty smooth”, “good and concise”, etc.). Archivists expressed that their confidence using the RMWs grew over time and that the process was relatively straightforward and well documented. Issues identified in the interviews clearly aligned with known pain points and reinforced our thinking about areas for further research and improvement. On the other hand, the responses challenged some of our assumptions

about a tool-based solution. Some of the pain points included needing to continually relearn to use the RMW after a period of non-use, communication and handoffs between processing archivists and the Archivist for Digital Curation, and transferring removable media as a processing bottleneck because of its time-intensive nature. To improve day-to-day work while our project was underway, we were able to make some immediate clarifications about handoffs and communication by updating internal processing documentation. Other pain points would require additional research.

Interviewing processing archivists during this phase ensured that our final recommendations were ultimately grounded in user needs and from this process conceptual categories for further research emerged. Upon review of the responses, it became clear that the scope of our research would need to extend beyond a tool-based approach and consider more fully all the roles, functions, and processes involved. The Review phase was critical in helping us to begin pinpointing areas for more thorough examination which were refined and explored in the Research phase. Based on the success of integrating interviews into this phase, we incorporated additional interviews as well as some unstructured meetings with archivists throughout the process.

3.2 Research

The Research phase included consulting with the Bentley’s Assistant Director for Curation to further scope targeted areas for research, speaking with colleagues at the University of Michigan Library about their experiences using the BitCurator Environment, comparing our workflow with current standards and best practices by reviewing similar workflows developed by peer institutions. We also conducted interviews with archivists at three institutions.

3.2.1 Methodology

In order to get a broad picture of practices across the field, we began gathering documentation and workflows of transfer and digital processing procedures from various institutions that made their materials openly available online. We annotated the documentation to indicate when it addressed topics related to the areas of research we identified in the Review phase and if additional innovative practices were described. During a meeting with the Assistant Director for Curation, we employed a few different methods for prioritizing areas for further research. We created an impact/effort matrix to determine how much impact each area of research would have on our workflow overall as compared with the amount of effort required to research and implement. Examining one area in more depth, we employed SWOT analysis to assess the strengths, weaknesses, opportunities, and threats associated with introducing a removable media tech position--an approach we have experimented with in the past and which is widely employed by peer institutions. Both the impact/effort matrix and SWOT analysis provided useful frameworks to help us prioritize the many suggested solutions for implementation.

We also spoke with colleagues at the University of Michigan Library about their experiences with the BitCurator Environment, which they have been using in production for about two years. Finally, we contacted four institutions whose documentation was particularly relevant to our work. We received responses from three institutions and developed a unique set of questions for each based on their workflows. We also prepared a set of general

⁴<https://docs.google.com/document/d/11kP1hpNunTmi1QcOm8PM5Ul6fc3sBIBQQwXXZHTC62k/edit?usp=sharing>

⁵ <http://bitcurator.net/>

questions to learn about their current practices and how they have evolved over time as well as more specific questions aligned with our areas of inquiry. Each interview lasted approximately an hour and included a brief introduction to the project and a history of our workflow.

3.2.2 Findings

Completing the impact/effort matrix resulted in seven areas of inquiry for our continued research and final recommendations: 1) “streams” of born-digital content, 2) technical transfer guidelines, 3) removable media tech role, 4) pre-transfer event tracking, 5) tracking and disposition of physical media, 6) reporting, and 7) photographing media. These areas are described in greater detail in the Recommendations section. Some areas, such as “streams” of digital content, were high effort but also high impact. In this instance, we decided that the potential for more efficient processes and the implications for digital processing at scale outweighed the effort required to research and develop this area of the workflow. With the results of the SWOT analysis, we determined that the overall strengths of creating a removable media tech position, including reducing the burden of training and oversight required as well as the burden on processing archivists carrying out this work, outweighed the weaknesses, including potentially creating a new bottleneck in the workflow with one person responsible for all transfers. These strengths and weaknesses could be further assessed during a pilot phase employing a graduate student. The adoption of this specialized role dedicated to the technical transfer of digital archives would permit archivists to focus their energies on the traditional archival functions of appraisal, arrangement, and description for all components of a collection whether physical or digital.

Ultimately, we decided not to pursue adopting BitCurator at this time. We found that the tools currently used within the RMW environment proved sufficient for our purposes and they were not explicitly identified as pain points. Additionally, the University of Michigan IT department does not support Linux machines, so we would have had to administer the workstations ourselves. We could not justify taking on that responsibility when considering the benefit versus the cost, especially since the functionality of the tools involved were very similar to those of the RMWs.

The search for examples of digital processing documentation eventually led us to reach out to archivists from three institutions in order to get a first hand perspective on their work. The interviews helped us to understand how the documented workflows translated to everyday practice at the institutions. The archivists provided valuable insight into their work processes, the challenges they currently face, and their goals for the future as well as how they are addressing some of the areas we are hoping to improve within our workflow. The information they imparted about pre-transfer tracking and reporting, such as the types of information tracked and methods for tracking as well as tools for generating reports, was especially helpful and has significantly informed our final recommendations. They also shared their experiences using the removable media tech model, confirming many of the advantages and drawbacks previously identified during the SWOT analysis. Speaking with colleagues both at the University of Michigan Library and elsewhere and comparing our workflow with

examples across the field ensured that our recommendations aligned with current standards and best practices.

3.3 Recommendations

Our goal in making recommendations for shorter- and longer-term improvements was to create a robust and flexible digital curation environment. This environment would be informed by theory and best practices while supporting the needs of archivists with varying roles and responsibilities and providing reasonable accommodation for the growing size and “it depends” nature of complex, heterogeneous collections. We considered the findings of the Review and Research phases and suggested a number of recommendations for improvements to the workflow. The effort we put into incorporating feedback in all phases increased our confidence in suggesting some relatively sweeping changes; our effort to be inclusive and keep people informed throughout the process has meant that implementing these recommendations has presented few difficulties thus far.

Some recommendations were relatively straightforward, including developing guidelines for field archivists and others to negotiate the technical transfer of digital material from donors to the Bentley, proposing a policy that, with some exceptions, disposition of removable media after transfer should be the default procedure, and optimizing our current webcam set-up to take better photographs of media. Others were more complex and are detailed in the following subsections.

3.3.1 “Streams” of Born-Digital Content

Our standard born-digital workflow utilizes ArchivesSpace⁶ and Archivematica⁷ and its Appraisal tab to streamline the ingest and deposit of fully processed digital archives to a repository for digital preservation and access. It works well for the majority of our collections, but it is not the most efficient way to process digital archives that are relatively homogenous, well-described or well-organized (e.g., scanned material). It also does not work well for potential candidates for preservation and access platforms other than our DSpace repository⁸, such as large batches of media or runs of content, etc. We thus recommended developing batch workflows for various “streams” of born-digital materials based on a number of criteria, identifying the following:

- **Born-Digital A/V Material:** Audio-formatted CDs, video-formatted DVDs and other runs of A/V content that are potential candidates for a more specialized A/V access platform (Kaltura⁹).
- **Image Collections:** Relatively homogenous, well-described or well-organized digital image collections that are potential candidates for a more specialized preservation and access platform for images (Digital Library Extension Service [DLXS]¹⁰ ImageClass).
- **Text Collections:** Relatively homogenous, well-described or well-organized digital text collections that are potential candidates for a more specialized preservation and access platform for textual material (DLXS TextClass).
- **Other Material:** Materials transferred via the RipStation or otherwise large batches of media destined for DSpace for which it seems appropriate to create one intellectual unit per piece of media. This approach could

⁶ <http://archivesspace.org/>

⁷ <https://www.archivematica.org/en/>

⁸ <https://duraspace.org/dspace/>

⁹ <https://corp.kaltura.com/solutions/education/>

¹⁰ <http://www.dlxs.org/>

also work for other types of homogenous or well-organized “runs” of material.

For each of these streams, we recommended taking advantage of the affordances of various systems beyond their Graphical User Interfaces (GUI) to work efficiently at scale, for example, by using Archivematica’s Automation Tools to automate the generation of Archival Information Packages¹¹ (AIPs) and the DSpace API to streamline the deposit of items and bitstreams. We also suggested exploring options outside of our standard DSpace repository to provide better access to born-digital material.

3.3.2 Removable Media Tech Role

Processing archivists are currently responsible for transferring digital content off of removable media. Due to the hybrid nature of modern collections, archivists often encounter digital materials. However, most archivists we spoke with said these encounters can be months apart or include only “2 to 3 collections out of 10.” For this reason processing archivists must often relearn to use the RMWs and have difficulty troubleshooting and understanding the goals of the work. In turn, this can make an already time-consuming process longer and further exacerbate processing bottlenecks. To remediate this issue, we recommended hiring a graduate student in a removable media tech position in order to test the centralization of the transfer process as a pilot project. The student in this role would be responsible for transferring content from all digital media freeing up archivists to focus on arrangement and description. If the pilot project proves successful, we would consider hiring a permanent staff member in this role.

3.3.3 Pre-Transfer Event Tracking

Archivematica tracks preservation events it performs on packages it transfers and ingests using PREMIS¹² metadata, but our workflow has not supported recording preservation events prior to Archivematica transfer as they move through the RMW processes. While many of our peer institutions take a more digital forensics “proper” approach, the information they track and the methods they use for tracking it could be adapted to fit our environment and needs. This could aid in describing the provenance of the content in our custody and we could even use such a resource to manage internal digital curation workflows and fulfill business intelligence needs.

We, therefore, recommended complementing Archivematica’s preservation event tracking by capturing “pre-transfer” events that archivists perform on removable media, such as virus checking, exporting files, imaging, Dissemination Information Package (DIP) generation, and quality control, in a spreadsheet or database to document provenance and improve internal work processes.

3.3.4 Reporting

Reports can provide valuable information about digital materials including directory structure, filename, file type, file size, etc. During conversations with both field and processing archivists, they indicated that this information was useful for appraisal and arrangement decisions. The Archivematica Appraisal tab provides this information for materials that have already been transferred and deposited into the Archivematica backlog. However, archivists currently lack the tools and information to assess digital materials prior to transfer, and do not always use the RMWs for this purpose. As a first step, we recommended communicating to archivists—in

all areas—about our goals in working with digital materials and the tools and resources we have available. Additionally, through ongoing conversations, we expect to learn more about their practices and needs. In the future, we could further explore options for accommodating pre-transfer appraisal workflows, especially with the introduction of the removable media tech role, ensuring that this type of appraisal happens with digital preservation considerations in mind.

3.4 Next Steps

In early 2018, we prepared our written recommendations and presented them to the processing archivists to solicit feedback. We drafted a final report incorporating their feedback and distributed to the entire Curation team.¹³ The project team is now testing and documenting implementations of these recommendations.

As we enact these recommendations, we are exploring a tiered approach to digital processing at scale informed by More Product Less Process (MPLP) [2], whereby the vast majority of lower complexity or relatively homogenous digital archives get processed in batch, while higher complexity or more heterogeneous digital archives move through a workflow characterized by manual appraisal, arrangement, and description but automated ingest. We are also more explicitly articulating the importance of the initial step of appraisal prior to transfer and demonstrating techniques for reviewing digital media based on written description, information provided by the donor, and other contextual clues.

4. DISCUSSION OF PROCESS AND BROADER IMPLICATIONS

The initial scope of this revamp project was relatively narrow. We set out to examine the BitCurator Environment, comparing its functionality to that of our RMWs to determine whether we should adopt it. However, our “agile” approach, characterized by a short feedback loop, an iterative methodology and, above all, an emphasis on direct communication, permitted us to evolve in response to early feedback and ultimately adapt our recommendations to better align with archivists’ needs.

Introducing processing archivist interviews into the initial Review phase of the project, for example, allowed us to understand real world user needs better than we had, as opposed to our perceptions of those needs, and challenged our assumptions about a tool-based approach. As it turned out, archivists were not having trouble with the tools we were using. Instead, they emphasized the need for workflow improvements that mitigate the issues associated with processing increasingly large amounts of digital archival materials. For instance, one archivist reported that “MPLP is not possible with the RMWs”. Subsequent recommendations reflected these suggestions, for example, by introducing sustainable digital preservation approaches that scale. The identification of various “streams” of content has enabled us to more efficiently manage large-scale ingest and processing. With this approach, we have identified a middle ground between developing custom workflows for each unique collection and trying to shoehorn everything into a single workflow.

Employing an iterative methodology was another essential aspect of our process. The Project Charter, rather than being overly prescriptive, identified milestones while allowing for flexibility in

¹¹ <https://public.ccsds.org/Pubs/650x0m2.pdf>

¹² <http://www.loc.gov/standards/premis/>

¹³ https://docs.google.com/document/d/1WRDTz3E-Zb5OPEh3G-R_Jv64JLld6AD_XKOGa2jfTco/edit?usp=sharing

how we accomplished them. Additionally, the success of the processing archivist interviews led us to incorporate more interviews into the process including those with colleagues at peer institutions. The processing archivist interviews also assisted in surfacing areas of inquiry for our research. A number of these areas, particularly issues relating to digital curation processes and roles, could not have been predicted at the outset of the project. An overly prescriptive approach that, for instance, attempted to outline interview questions for our colleagues at peer institutions at an earlier point would not have yielded information relevant to our areas of inquiry which could ultimately be adapted for use at the Bentley. Recommendations for tracking pre-transfer events as well as a removable media tech role--both of which provide solutions for issues relating to workflow processes, roles, and handoffs--were the direct result of this iterative, grounded methodology. These recommendations have the added benefit of accomplishing digital curation goals, such as maintaining authenticity and tracking provenance, more effectively than our previous procedures. As with our process as a whole, the aim for recommendations is to create guidelines that accommodate various archivist needs, leaving room for archivist discretion where appropriate and that, above all, aid archivists in making informed decisions about how to choose the best of a number of ways they might process a collection.

Finally, our process emphasized frequent, direct communication among the project team and between the project team and various groups of archivist stakeholders including accessioning, processing, and field archivists, as well as our colleagues at the University of Michigan Library and peer institutions. While researching digital preservation theory and best practices was a valuable part of our process, it was these face-to-face encounters that had the most impact on our final product. For instance, as we began to identify various "streams" of born-digital content, it became apparent that, in many cases, the software and workflows we needed to accommodate the various types of digital content were already in place for digitized materials. However, they had never been used for born-digital content. Collaboration with those product owners at the Bentley was key to aligning our various workflows and sharing technology platforms. Furthermore, the idea for reporting came about through collaboration with field archivists and is an attempt to meet their needs for pre-transfer appraisal of content with the need for these processes to occur in a minimal preservation environment.

5. CONCLUSION

While this review and our subsequent recommendations certainly represent a milestone in our removable media workflow procedures, they are by no means final. The rapid pace of evolution in systems and data dictates an iterative approach to workflow development accompanied by regular opportunities to reflect on our workflows as well as those of our peers. It is important to us to continually enhance our archival management and preservation infrastructure. For us, successful work on technology upkeep, maintenance across teams, and shared services has meant that we need to listen to the people who use our tools and workflows everyday and monitor community practices in order to make informed and accountable preservation decisions.

6. ACKNOWLEDGMENTS

We thank our colleagues at the Bentley Historical Library, the University of Michigan Library, the University of Maryland, Purdue University, and the Rockefeller Archives Center.

REFERENCES

- [1] AIMS Work Group. 2012. AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship.
- [2] Mark A. Greene and Dennis Meissner. 2005. More Product, Less Process: Revamping Traditional Archival Processing.