

Collaborative Creativity: The Radcliffe Workshop on Technology and Archival Processing

MARY O. MURPHY

Schlesinger Library, Radcliffe Institute, Harvard University

ANNE SAUER

Digital Collections and Archives, Tufts University

Abstract: Most, if not all, archival repositories are faced with significant backlogs of unprocessed materials that are inaccessible to researchers. In attempting to address the barrier to access these backlogs present, archivists have attempted to streamline processing using a variety of approaches, including “More Product, Less Process.” While these techniques have had an impact, there is still much to be done. Utilizing an applied research approach, the Radcliffe Institute for Advanced Study hosted a Workshop on Technology and Archival Processing in May 2011. The event brought together a carefully chosen, cross-disciplinary group of archivists and technology leaders to examine new ways to leverage technology to improve access to information in archives. Attendees were asked to consider a set of use cases that illustrated archivists’ current workflows and the pressures of significant backlogs in all formats, including non-digital material. The technologists were challenged to think about how recent advances in automation and visualization could assist in the descriptive process, while archivists were challenged to re-envision our own practices. Each use case was grounded in real world scenarios of access and use. The Workshop yielded a number of novel and innovative approaches to the very real problems facing archives, if archives can find ways to test and implement the ideas. Most importantly, the Workshop demonstrated the utility of an applied research approach to bringing together diverse expert communities and fostering creative problem-solving.

Introduction

Most, if not all, archival repositories are faced with significant backlogs of unprocessed materials that are inaccessible to researchers. In attempting to address the barrier to access these backlogs present, archivists have attempted to streamline processing using a variety of approaches, including “More Product, Less Process.” While these techniques have had an impact, there is still much to be done. On May 16-17, 2011, the Radcliffe Institute for Advanced Study hosted a Workshop on Technology and Archival Processing. The event brought together a carefully chosen, cross-disciplinary group of archivists and technology leaders to examine, in an applied research approach, new ways to leverage technology to improve access to information in archives. Attendees were asked to consider a set of use cases that illustrated archivists’ current workflows and the pressures of significant backlogs in all formats, including non-digital material. The technologists were challenged to think about how recent advances in automation and visualization could assist in the descriptive process, while archivists were challenged to re-envision our own practices. Each use case was grounded in real world scenarios of access and use. The paper will present a description of the workshop with an explanation of the implications and potential impact of opportunities that emerged from the event.

Problem Statement

Technology-driven user needs and economic conditions complicate the challenges presented by backlogs.¹ There remains a pervasive problem of unprocessed collections that is exacerbated by inadequate levels of funding for staff to process collections, and inadequate tools to make collections digitally accessible to researchers.² And though in recent years, a number of technological approaches have been implemented across various repositories to decrease backlogs and ease barriers to access, sustainable processing methods have remained elusive.³ It was against this backdrop in 2010 that Marilyn Dunn, the Executive Director of the Schlesinger Library and Radcliffe Librarian, began to engage in talks with computer scientist and former Radcliffe Dean Barbara Grosz. They, along with other members of the Harvard archival community, discussed possible technological solutions to the on-going challenges presented by backlogs in archives.

Dunn conceived of an idea for a multidisciplinary workshop that would push the thinking about backlogs, processing, and access beyond all previous approaches to new areas of thought. Dunn wanted to host a workshop that would be both collaborative and creative, from the planning stages to the event itself, where she hoped to bring together archivists, computer scientists, and other technologists to brainstorm collectively about new ways to use technology to provide maximum access to archival collections and to create sustainable archival methods.

Methods

Dunn caucused with fellow professionals in the Boston/Cambridge area to form a group of twelve planning committee members from across the technology and archival sectors. They represented various institutions including the Massachusetts Historical Society, The John F. Kennedy Presidential Library, Massachusetts Institute of Technology, Tufts University, and Harvard University.

The planning committee agreed to meet once each month to strategize about all facets of the workshop including who might participate, what the agenda might look like, and what the content of workshop should include. They also appointed a project manager to coordinate their work and to execute planning decisions.

The group acted foremost with a mission to coordinate a workshop that would leverage technology to streamline processing and improve access to archival information. In addition, planning members worked with the goal of building bridges: first to technologists, to enlist their help identifying technological solutions to defeat backlogs; then to archivists, who would need to

¹ See, for instance, Elizabeth Druga, "Raising Money Raises Questions: The Ethics of Generating Revenue from Archival Materials," *Journal of Information Ethics* Vol. 19 No. 1 (Spring 2010): 141–156; Barbara M. Jones, comp., "Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in North America's Research Libraries," White paper for the Association of Research Libraries Task Force on Special Collections (2003), <http://www.arl.org/bm~doc/hiddencollswitepaperjun6.pdf>, accessed 26 October 2011; and Kate Theimer, "What is the Meaning of Archives 2.0?" *American Archivist* 74 (Spring/Summer 2011): 58–68.

² See, for instance, Mark A. Greene, "MPLP: It's Not Just for Processing Anymore," *American Archivist* 73 (Spring/Summer 2010): 208–263; and Christopher J. Prom, "Using Web Analytics to Improve Online Access to Archival Resources," *American Archivist* 74 (Spring/Summer 2011): 158–184.

³ See, for instance, J. Gordon Daines III, "Re-engineering Archives: Business Process Management (BPM) and the Quest for Archival Efficiency," *American Archivist* 74 (Spring/Summer 2011): 123–157; and Mark A. Green and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68 (Fall/Winter 2005): 208–63.

leverage new technology to move from manual to digital processing; and finally to users, to provide them with collections in ways that meet and exceed their research needs and expectations. The purpose of the workshop would be to seek creative possibilities to cross these bridges and to begin a new conversation about technology in the aid of archives.

Workshop Design and Use Cases

The planning committee determined that a small group of professionals – numbering approximately 40-50 people – would be best suited to produce creative, substantive results from a two-day, brainstorming and problem-solving event. The idea was that fewer participants would allow for brainstorming in greater detail, providing individual participants with time and space to articulate their ideas more completely.

Following that decision, the committee then worked through its most difficult challenge to identify potential participants and determine who would be invited to attend. The committee researched dozens of archival professionals and technologists who were all well-qualified candidates.

Eventually, the planning committee invited a group of 45 cross-disciplinary professionals to the event. Those chosen represented diverse professional backgrounds from academia to industry. One participant, Richard Pearce-Moses, provided his expertise on archival theory and practice, while another participant, Jon Kolko, provided advice on technology and human-computer interaction. While all of the participants were chosen primarily because of their research specialties, it was also their potential willingness to brainstorm out loud; in a group with other professionals; creatively; and with a guarantee that they would not say, “Yah, but let me tell you why this can’t be done.” At the Radcliffe Workshop, open-minded, innovative thinking was valued most of all.

During the same period, the planning committee also began to wrestle with how to merge the two worlds of archives and technology and allow for substantive brainstorming to emerge during the event. The committee grappled with many questions such as: “How will these very diverse participants speak to one another? How can we get the technologists interested in paper? How can we get the archivists to let go and be creative about processing? And most of all, how can we succinctly and clearly explain our problem so that our participants can help us find solutions?”

In response to these questions, planning committee member Wendy Gogel, Manager of Digital Content and Projects at Harvard University, proposed the idea that the committee write “use cases” that could be distributed to the participants prior to the workshop and then used as a basis for brainstorming during the event. Use cases are typically used by software engineers and can be described as documents that outline a sequence of interactions between an actor (archivist) and a manual process (processing) that is necessary to deliver a service (processed collections) that satisfies a goal (access to information).⁴

As a result of Gogel’s idea, a smaller group of planning committee members worked by subcommittee to write a *Scenarios and Use Cases* document. This document outlined, in narrative and table form, the problem that the Radcliffe Workshop hoped to address.⁵

The document contained two narrative scenarios, with two related use cases attached to each scenario. On each of the two days of the Workshop, one scenario was used to set the scene for the day’s discussion. Each scenario described actual situations and problems relating to accessing

⁴ Alistair Cockburn, *Writing Effective Use Cases*, (Addison-Wesley, 2001).

⁵ Radcliffe Workshop Scenarios and Use Cases (insert URL when available).

archival collections and processing archives. Day 1 of the Workshop concentrated on **“Maximizing Discovery.”** The Scenario focused on “A Researcher’s Perspective on Access,” and explored two use cases entitled, “Maximizing discovery based on information from within collections and from outside sources,” and “Maximizing usability of collections that contain confidential or otherwise restricted content.”

Day 2 of the Workshop concentrated on **“Streamlining Processing.”** That scenario focused on “Creating Sustainable Methods for Archival Processing and Digitization,” and explored two more use cases entitled, “Optimizing automation and digitization to facilitate and speed processing” and “Retaining and providing access to the most complete sets of records within collections.”

Following the Workshop, planning committee members agreed that the scenarios and use cases played a key role in structuring the workshop’s discussion, prompting ideas and allowing the participants to communicate with one another.

Workshop Description

Preparation for the Radcliffe Workshop on Technology and Archival Processing required seven months of planning that culminated in a two-day event held on May 16-17, 2011.

The event opened with a multi-repository collections exhibit, organized and hosted by Bill Stoneman, the Florence Fearington Librarian of Harvard’s Houghton Library. Attendees mingled throughout an exhibit space, stopping to view specially selected collection items and talking with archivists at eleven different stations. Each station displayed items from a collection chosen to highlight specific problems outlined in one of the scenarios or use cases that would likely be touched upon during the breakout discussion sessions. The aim was to provide participants with descriptive and visual background information to help them prepare for creative thinking to follow.

Participants then moved to four round tables to begin the process of thinking out loud together. Each brainstorming session began with an introduction by a participant who suggested topics and challenges for discussion. Table leads also facilitated conversation amongst the participants and referred conversation back to the scenarios and the use cases throughout the brainstorming periods. At the close of each session, one person from each table reported on that group’s deliberations to the larger group.

The workshop closed with a summary session hosted by Clifford Lynch, the Director of the Coalition for Networked Information. This period allowed participants to offer final thoughts and other “takeaways” to the larger group. It also functioned to start a new conversation about next steps.

Results

The success of the workshop hinged on bringing both communities to the table ready to engage in a productive dialog. This required much more than intelligence and enthusiasm, which the participants had in abundance. Nothing would come of the workshop unless archivists and technologists opened themselves to radically different ways of thinking.

For the archivists, preparation meant being ready to put all practices on the table for discussion, no matter how firmly embedded in tradition or how “essential” a particular task was perceived to be. Archivists were asked to suspend their “yes, but we can’t because...” responses in favor of responses that would encourage brainstorming and creativity: “Interesting. How would that

work?” This was no small task for even the most avant-garde archivists present as the ideas started flowing. The tangible benefit of this approach was an increased awareness of the objectives of archival work and a conscious separation of those theoretical objectives from tasks. Clearly, if this was to work, the objectives would still need to be fulfilled even as the tasks might change radically. Indeed, there would be no hope of addressing the issue without recognizing that archivists’ tasks needed to change across the board.

For the technologists, there was the more fundamental problem of quickly conceiving of the vast complexity of challenges faced by archives and the nature of the work that archives do. The technologists needed to see, touch, and feel the problems that archivists deal with daily. In this way, the exhibit was a tremendous success in presenting a variety of collections in various states of processing, with format challenges, confidential data, and preservation issues. An archivist accompanying each collection was instructed to explain the particular challenges of their collection and describe how they would use current practices to deal with the collection. The technologists asked questions, took photos with their smartphones, and gained important insight into the challenges archivists face.

Opening the minds of these two groups was key to enabling the creative and exciting brainstorming that followed. As important as the workshop findings, the methods that enabled these two groups of professionals to be brought together rapidly and with a high degree of awareness was extraordinary and an important result of the workshop. Realizing any of the ideas that emerged at the workshop will require ongoing, open, and constructive connections between the two communities. The techniques used to prepare the workshop participants can be replicated to ensure the collaborative creativity can continue.

Findings

First, Digitize Everything. In order to leverage technology to its maximum capacity, the clearest recommendation to come out of the workshop was to digitize collections *before* arrangement and description. Such an approach stands standard best practices for digitization in archival collections upside down, but the technologists argued persuasively that any leveraging of technology would be enabled only if collections exist in digital form. Printed text, handwriting, photographs, notes, media—all should be captured and converted to machine-readable form prior to description.

Several variations on this theme were discussed. For large collections, perhaps not everything would be digitized at the accession point, but every item touched by an archivist in surveying the collection would be captured. How would this work in practice? Some whimsical participants imagined archivists with cameras mounted on their glasses, snapping frames by blinking. Perhaps more realistically, this sort of workflow could be accommodated by specialized worktables with overhead cameras, foot-pedal controls, good lighting, and digital recording of voice notes from the archivist. Capture systems would need to preserve the context of original order and allow digital surrogates to be associated easily with physical originals.

Arrangement and description could then take place with the digital surrogates, perhaps leaving the physical originals in place, except where preservation would be required. Minimizing physical arrangement could save valuable time, and have the added benefit of preserving the original order when access may be facilitated by an imposed arrangement.

Quick and Dirty Digitization. The need for speed in digitizing at this phase of the arrangement and description process would likely necessitate digitization at lower resolutions. Again, this is

something that stands contrary to much best practice in our field, where there is a strong belief that digitization should be done at the highest resolution possible to create a high-fidelity digital surrogate. However, if the purpose of digitization is to facilitate other processes such as text mining, image recognition, and basic information access, then the minimum resolution necessary to accomplish these goals will suffice. High-quality capture can be done later, if needed.

Digitization does not (necessarily) equal online access. While digitization is generally seen by archivists as a means to promote access, it does not need to be the case. Here, digitization is performed to facilitate description. Access is secondary.

However it is conceived, digitizing as much as possible up front is clearly key to reformulating the rest of arrangement and description. It cuts the Gordian knot by rendering collection material in formats where automation can potentially revolutionize the rest of arrangement and description. Without this crucial first step it is difficult to see a way to forward progress. Once in digital form, a variety of tools can be applied.

Assembly-line Archives. As the technologists listened to accounts of archivists performing all tasks required to bring a collection from accessioning to fully cataloged, another clear recommendation emerged. Archivists are too valuable and too skilled to perform many of the tasks that have traditionally been part of archival practice. Managers need to match skills to the job to ensure that archivists are performing tasks that others cannot do. In some cases those others may be computers, but they may also be paraprofessionals. For many collections, initial digitization tasks could be performed by technicians, rather than professionally trained archivists. Managers will have to find ways to facilitate communication among processing teams to ensure that insights into the collections being worked on are communicated and captured as work unfolds. Increasingly, archivists will become team leaders – managers responsible for creating good description in a team-oriented environment.

Confidentiality Filters. Just as spam filters keep junk mail from choking email systems; it may be possible to develop tools that would analyze text to determine if potential confidentiality issues exist. Screening for confidential information is a time-consuming – and thus costly – process. Confidentiality filters could be trained to flag based on keywords or string types (xxx-xx-xxxx for Social Security Numbers, for example). While such filters would likely never negate the need for review by an archivist prior to public release online or handing over to a researcher, a filter could make a reasonable first pass on a collection. Would automated tools in this realm be infallible? No, but archivists are not perfect, either, and automated tools could save valuable time, when used judiciously and with proper oversight.

Name Recognition. Use automation to assist in identifying names in a collection. Over time, a name index of collections will aid in speedy and relevant identification of individuals represented in institutional collections. Research on automated named entity recognition is already quite advanced in the digital library field, and could be deployed in an archival setting with minimal adaptation.

Crowdsourcing. Where confidentiality is not an issue, release materials and allow interested individuals to contribute to making archival materials more discoverable. Archives are already taking advantage of this in a variety of ways, from tagging images to transcription of handwritten texts and correction of OCR. Leveraging expertise from the crowd should be a matter of course, and archivists should find ways to manage these contributions alongside archivist-created metadata.

Conclusion

In the end, the workshop encompassed two days of concentrated brainstorming whose effects will only be felt if and when the ideas put forth are tested and evaluated.

Implementing the wholesale changes suggested above may seem daunting for a host of reasons. Are we ready to embrace technology so wholeheartedly? Can we manage the massive digital collections these workflows would create? And there are, of course, archivists who will likely be reluctant to relinquish traditional tasks in favor of automation, or to re-envision themselves as team leaders rather than curators. But the greater challenge – that should push us to follow up on the workshop findings – is the pervasive awareness that all archives have far more work to do than can ever be done. It is overwhelming to consider the time it would take to change everything that needs changing.

Daunting, indeed. However, the risks of continuing as we have been are far greater. Resources are tight and getting tighter, outside funding is diminishing, and resource allocators are closely questioning the return on investment across organizations. In this climate, it is quite simply inappropriate to think that archivists can continue to labor in ways that are clearly inefficient. We must find better methods to get on top of the materials in our care. If not, we risk our collections and our stewardship. We risk our very existence.

Where do we go from here? Archivists participating in the workshop left energized and enthusiastic about the ideas discussed. Many of them urged continued, cross-disciplinary conversation on the topic and the development of pilot projects to experiment with the ideas that emerged from the event. As a result of this, archivists at the Schlesinger Library have established the Experimental Archives Pilot Project to test some of the workshop's themes.

The Experimental Archives is based on a collaborative model. It aims to provide a virtual and physical laboratory space where experts in the field of technology can work collaboratively with archivists to test creative solutions towards the development of sustainable, digitally enhanced processing practices. The goal of the project is to provide maximum access to information in archives and to provide that information to users in the way they want it.

Three experiments are currently underway including “Direct to Digital Processing,” “Rethinking Redaction,” and “New Methods for Newspapers.” A project Wiki is being used to document the experiments and to provide a space for conversation and coaching, and where in the future, other Radcliffe Workshop participants might contribute their pilot projects happening at other archival institutions. For more information about the Schlesinger Library Experimental Archives Pilot Project, use the following link to access the project Wiki:
<https://sites.google.com/site/experimentalarchives/>.

References

- Cockburn, Alistair. *Writing Effective Use Cases*. Addison-Wesley, Boston, MA, 2001.
- Daines III, J. Gordon. "Re-engineering Archives: Business Process Management (BPM) and the Quest for Archival Efficiency," *American Archivist* 74, no.1 (2011): 123–157.
- Druga, Elizabeth. "Raising Money Raises Questions: The Ethics of Generating Revenue from Archival Materials," *Journal of Information Ethics* 19, no. 1 (2010): 141–156.
- Green, Mark A. and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68, no.2 (2005): 208–63.
- Greene, Mark A. "MPLP: It's Not Just for Processing Anymore," *American Archivist* 73, no.1 (2010): 208–263.
- Jones, Barbara M., comp. "Hidden Collections, Scholarly Barriers: Creating Access to Unprocessed Special Collections Materials in North America's Research Libraries," White paper for the Association of Research Libraries Task Force on Special Collections (2003), October 26, 2011. <http://www.arl.org/bm~doc/hiddencollswitepaperjun6.pdf>.
- Prom, Christopher J. "Using Web Analytics to Improve Online Access to Archival Resources," *American Archivist* 74, no.1 (2011): 158–184.
- Theimer, Kate. "What is the Meaning of Archives 2.0?" *American Archivist* 74, no.1 (2011): 58–68.