Minutes, Migration, and Migraines: Establishing a Digital Archives at a Small Institution

Joseph A. Williams and Elizabeth M. Berilla

ABSTRACT

This case study discusses the implementation of a sustainable born-digital institutional archives plan at the Stephen B. Luce Library, SUNY Maritime College. After conducting a review of the literature, the authors discuss a three-part plan to accession new materials, migrate records to long-term digital formats, and provide accessibility through a full-text accompanying website. Methodology and implementation of a fully digital archives are described in a practical application for small institutions. Copyright, confidentiality, and stakeholder buy-in are discussed as well as are future plans for expansion.

KEY WORDS

Migration, Digital-born, Digital Archives, Faculty Buy-in, Institutional Archives, Case Study

An institutional archives is a permanent record of an organization; a memory bank that can be used for reference and historical research. As two scholars put it, "The aims of preservation are to increase the longevity of active materials through careful storage and use."[1] While this mission has not changed, the means, in the context of the ubiquitous growth of born-digital records, has done so dramatically. As a result, the scholarly conversation is dominated by debates over the best practices of archiving digital materials including long-term preservation, appropriate access, and proper description.

The supplanting of print records by those born digital is well documented. In some cases, when the change occurred, institutional archives that centered their collections on print materials were ill prepared for the onset of the digital revolution. Aside from the technical issue of establishing a permanent digital archives, acquisitions lapsed. This was especially the case around the turn of the twenty-first century when email blasts replaced circulated paper memos, and committee correspondence was saved as ephemeral Word documents on flash drives whose existence and playback depended on the availability of a computer and the good intentions of the content creator.

This situation occurred at the State University of New York's (SUNY) Maritime College, where the Stephen B. Luce Library had been the repository for all the institution's archival documents since the 1940s. Yet, around the year 2000, the number of physical items accessioned into the print archives dropped precipitously. Various departments, which had automatically sent physical copies of memorandums to the archives, now maintained electronic copies on departmental network drives. Committee minutes, annual reports, and other ephemera often disappeared. With them, institutional memory vanished.

This article discusses how the Stephen B. Luce Library is recovering from its archival "Dark Ages" of 2000 to 2010 with a new digital paradigm for institutional records management.

About the Institution

Maritime College is one of sixty-four colleges and universities in the SUNY system. It is one of the smallest colleges in SUNY with a full-time enrollment of only nineteen hundred students. The college curriculum primarily trains students to become licensed members of the American Merchant Marine with undergraduate degrees in marine transportation, business, and engineering. It also has graduate programs in global business and maritime and naval studies.

The Stephen B. Luce Library is dedicated to supporting the academic programs of the college. It also possesses considerable institutional archives dating from the mid-nineteenth century. The library served as the institutional repository for most official documents of the college, which includes over nine

hundred linear feet of faculty meeting minutes, administrative memorandums, ship's log books, papers of the college presidents, curricular materials, catalogs, annual reports, statements, and other items. Yet, as mentioned, all of these important institutional documents ceased to be incorporated into the institutional archives around the year 2000 when institutional records became, more or less, completely digital.

Literature Review

The Value of Digital Institutional Archives

The academic literature is full of articles that discuss the digital revolution and its impact on academic and archival repositories. This encompasses more than mere format changes. One scholar wrote, "Due to low-cost, minimum space and little effort, there has been a fundamental change in the way we perceive information storage."[2]

A 2011 survey of professional writers revealed that a majority of respondents created their content primarily in digital format. Respondents also did not value their digital files or follow best practices for their preservation. The authors of the survey conceded that whether or not writers archive their work may be a moot point, since so few become eminent enough to be studied in the future, it is still valuable for digital preservation research.[3]

Yet, information professionals agree that reliable digital archives need to be established and maintained. Ross provided the best reason for this: authenticity. He emphasized that because digital records are so ephemeral, a trusted home for electronic documents, where they will be preserved as they were at the moment of creation, is needed. Ross continued that this is especially important in a hyper, self-aware, socially reflective postmodern society.[4]

While archivists and other information professionals generally agree that digital archives not only should be, but must be, established and maintained, a number of problems can arise in doing so. We reviewed these prior to making our own plans at SUNY Maritime College.

Problems of Establishing and Maintaining a Digital Institutional Archives

Archivists, librarians, and academics have attempted to rein in haphazard practices of digital archiving with efforts to standardize it. Some scholars, such as Gaur and Tripathi, conceded that the infrastructure for print archiving is superior to that for digital preservation. To be successful, a thoroughly planned

strategy is needed to maintain a sustainable electronic archives.[5] Any plan or methodology implies the need for standardization. This is deemed important since, as Joyce Ray wrote, " . . . a digital repository is not just any data storage system. To be trustworthy, it must be managed with the intention of long-term use and in accordance with the archival principles of authenticity, integrity and provenance."[6] These concepts are central to the Open Archival Information System (OAIS) Reference Model, which offers a conceptual view of accessioning, preserving, and allowing access to digital materials.[7] It, in effect, is a framework for the entire process, from digital acquisition to public access. Adopted internationally by many institutions, the OAIS is the most ubiquitous model for digital archiving.[8] Even so, lags have been reported in the implementation of digital archives, especially on the part of libraries in the developing world. Also vexing are the inconsistent practices for implementing an electronic repository, despite the availability of well–thought-out theoretical models like OAIS.[9] At SUNY Maritime, we decided to take key concepts from the OAIS model of organization and planning when we undertook the project to build a digital archives.

Despite theoretical advances in digital archiving, there are practical limitations as pointed out by Mike Kastellec. Quickly antiquated technology, data redundancy, selection criteria, access issues (virtual, physical, and temporal), a discombobulated legal structure, and funding all contribute to the difficulties of establishing a digital archives.[10]

Gaur and Tripathi specifically commented on preservation issues asserting that " . . . digital publication deteriorates much faster than paper. A digital object may be corrupted or lost and thus become irretrievable. But even before that happens, the technology used to store the publication is likely to become obsolete."[11]

Perception is another problem in establishing a robust digital archives. Administrators commonly believe that a digital archives is cheaper and less work intensive than a physical archives. This is a gross misunderstanding. Digital curation can be as expensive as print curation. Becker noted that constant migration of data into more stable electronic formats, the maintenance of appropriate software and hardware to sustain the archives, and ensuring digital backups are just a part of an ongoing commitment by the repository to maintain the quality and accessibility of its records.[12] Lee and Tibbo supported this assertion: "In contrast to caring for analogue materials, digital curation brings a wide array of opportunities and challenges. Opportunities include both wider and integrated access, representation of an increased range of human experience, persistence through redundant copying, economies of scale, and enrollment of collective expertise. Challenges include bit rot, obsolescence, social inertia, technology monitoring, intellectual control, access environments, and the ability to convey meaning over time."[13]

We are left with, therefore, numerous difficulties and no agreed-upon best practices to follow in establishing a digital institutional archives, especially at a small institution with limited funds and staff. In this regard, practicality often trumps theory, and a middle ground of digital content management must be contemplated. Because of their own idiosyncrasies, institutions must cherry pick among best practices for what works for them. In essence, every institution must develop a unique plan.

The Role of Planning

The literature on digital archives emphasizes the importance of planning. The Colorado Digitization Project, an effort by the State of Colorado to digitize its significant historical resources, required significant planning among archives, schools, and libraries.[14]

The German National Library is another excellent example of institutional planning. Starting in the 1990s, the institution formed a staff of knowledgeable professionals to preserve digital objects. This was propitious since a 2006 law charged German libraries to collect and preserve digital publications. Despite this, they faced legal, financial, and physical problems such as those described by Kastellec.[15] Another international example is the public library of Koprivnica, Croatia, which has digitized thousands of local records including newspapers and born-digital records relevant to the local population.[16]

In another demonstration of planning, Emory University Library presented a case study on how it preserved the born-digital materials of the Salman Rushdie Collection. The library developed a working group to assess the collection and formulate a preservation plan. Taken into consideration was the creation of so-called dark and gray archives where access to certain materials is barred or limited due to the confidential nature of the items or donor preference.[17]

Yet, for all of these case studies and examples of creating digital archives, discussion about the practical application and workflows associated with digital preservation in small institutions is limited. In addition, most case studies deal with cultural or historic archival materials, not with the business materials of the institution itself.

At SUNY Maritime College, we customized a multifaceted approach from the OAIS model and best practices from the literature. We found that, much like other case studies, collaborative planning was a necessary component of our approach, albeit our collaboration was local and interdepartmental.

The Problem

Around the year 2000, the Stephen B. Luce Library's institutional archives began to suffer. Born-digital records were not being added to the collection. This problem was compounded by the retirement and discontinuation of several librarians who were responsible for processing institutional documentation. No replacement archivists were hired, and, as a result, much official correspondence for the decade between 2000 and 2010 was never acquired and is now lost to the historic record. Collection efforts were sporadic, and of the scale of the problem was little recognized until those conducting institutional research found significant record gaps.

The question became: who would be responsible for preserving born-digital materials for future reference? At this point, the library, realizing the opportunity to present the value of its services, took the effort upon itself. We believed it our duty to restore this traditional role to the library's archives. Our goal was to provide tangible value to the college by preserving its records and making them as accessible as possible.

Objectives and Plan

The institutional archivist and the head of technical services developed a plan to reassert intellectual and physical control of the institution's digital collections. This preservation plan included best practices from the scholarly literature modified to the specific needs of the Maritime College campus.

The library took into consideration the logistical and legal aspects of publicizing records online. Mindful of this, we considered who would be responsible for the collections in the future, how the information would be distributed under copyright, and who would be permitted to access the collections. We also realized that, for long-term preservation, legal responsibility for the stewardship of electronic records requires campuswide awareness and input from additional faculty committees, departments, and offices. In the end, our plan consisted of three phases: 1) acquire; 2) process and migrate; and 3) enable full-text searchable accessibility. We wanted to acquire those electronic documents that were the clearest record of the history of the institution and had the greatest relevance for faculty governance. To this end, we targeted the electronic documents of the administration, various academic departments, and faculty committees. We knew that even if we came up with an excellent plan, it was going to be subject to change. So we decided to test the process with a single faculty committee with the intent of expanding the program later.

We had long discussions about the best methods and processes. Questions arose such as where the electronic files were going to be kept, who was going to

be responsible for ensuring that records would be continuously acquired, and in what format files should be stored. This last question, surprisingly, turned out to be the most vexing and will be discussed in further detail below.

We decided that we wanted a digital repository that would be used. Full-text searchable files would allow for that with the addition of a local search engine that could crawl through the individual documents. We believed that by offering this service to our users, we would garner further support for the project.

Methodology

Acquisition

We proceeded by approaching the various constituencies that created the born-digital documents. Our first target was the college's Curriculum Committee, which is highly active and generates, per capita, more digital documents than any other committee on campus. We also believed that this would be an ideal committee to begin the project since its documents are created for public access, so we did not need, at that point, to consider dark or gray archiving any materials.

The librarian appointed to the Curriculum Committee made our case to acquire the digital files before they became compromised, decayed, or corrupted. It is interesting to note that most of the faculty on the committee did not even consider digital preservation necessary. This is an example of how people do not view their daily activities as being of significance a century later. However, once they were educated, feedback was immediate and positive with strong buy-in for the plan. They voted approval for our proposal, and the committee chair met with the archivists to select the documents most appropriate for the archives.

The selection process was challenging. At first, we accepted all documents. This was a mistake given the multiple iterations of similar documents. For example, curricular workflow documents went through dozens of iterations as the committee met, discussed, and altered the documents into their final form. As the archivists began to process and digitize all these various documents, we came to the conclusion that it was impossible and unsustainable to keep up with the workload. We found it especially dispiriting to spend so much time working with these spurious documents of limited long-term value. Therefore, in mid-acquisition we went back to the committee and explained the problem. We agreed that it would be in the best interests of sustainability and relevance to accession only the final versions of the documents to be collected at the

end of every semester. The committee and archivists selected these documents based on their potential historic/research value to the institution. The reduction in accessioned documents was significant, but manageable. It is just as well that the librarians do not have off during the summer and winter intersessions.

Process and Migration

We decided to migrate the Curriculum Committee's digital files to standardized file formats to create long-term sustainable records. Some sentiment for keeping the files in their original form and using emulation in the future existed. Emulation would have allowed access to the documents in their original settings and software so as to re-create their original usage scenarios. Migration, however, was viewed as more practical, malleable, and economic in the long run given the capabilities of technology and staff. We approached migration as a more efficient way for the archives to evolve within our means, over time, and within the overall context of the Institutional Archives' mission.

The original formats of the electronic records were PDF, Microsoft Excel, and Word. With the assistance of staff and interns, the documents not in PDF were migrated to PDF since this format has become a popular option for long-term digital storage. The decision to convert the files into PDF was not without debate since the majority of the original files were either in Word or Excel. This meant that we would need to invest considerable time in converting the files. Also, the point was made that we would be going against the grain of best archival preservation practices by eliminating the original format entirely. In addition, the format change might dissuade some of our users. For example, a user seeking an Excel document might dislike the idea of being only able to work with a PDF version. In the end, we reasoned that long-term preservation trumped these other concerns. Also, we decided that even though the conversion of the files to PDF may be time consuming, it would only be a problem during the initial startup of the digital repository because the plan for the future is to take the files on a rolling basis. Finally, the migration of information into the more stable PDF allowed for a broader range of digital access. We chose not to use PDF/A because we thought it would compromise the printable qualities of documents—an important feature for our specific users.

To preserve provenance and original order, the files were converted and saved to the library's server in the same hierarchical file structure as their original location on the shared institutional server. The depth of the file hierarchy was considerable, often involving more than five levels of subfolders. This led to confusion about which individual files were actually converted. To ensure completion of the task, we created a list of the file structure and checked it off as sections were completed. At the end of the conversion process, we went back

into the original file structure to double check that we had captured and converted each individual file. Needless to say, our interns and staff had their fair share of migraines during the migration process.

With the PDFs available in a controlled hierarchy, we sought to make *all* of the documents searchable through an open-access database hosted by the library's server. We wanted to implement the preservation, planning, and data management aspects of the OAIS Reference Model recommended by Carroll et al., including user access.[18] Our plan for the digital archives was to maintain functionality and searching preferences reserved for finding aids, archival description, and metadata embedded in the document and its formatting.

Access

We allowed searchable, digital access through a subpage on the library's website. Through basic Web programming using Dreamweaver MX 2004, we created HTML links for individual PDF files. Pages were then added at the "group" level according to the original folder order created by the Curriculum Committee on the shared server. Item-level documents were taken from the individual files within the Curriculum Committee folders. After these pages were populated, a final page was added to host a local Google "crawler" search box and listing of results. We limited this crawler to reveal only results located on the library's digital archives pages and to conduct full-text searches of the converted PDF files. For users of traditional search methods, each digital group and item was added to the Institutional Archives finding aid with the designation of Electronic Record under Source Type. Each entry included metadata adhering to Dublin Core standards reflecting significant properties in line with archival best practices as described by Margaret Hedstrom.[19] The metadata were then placed into a spreadsheet for the archivists' use, including Dublin Core standard elements of format, description, contributor, and creator.

Conclusions and Future Plans

Our experiences establishing an institutional digital archives provide insights and practical guidance that support the academic literature in this growing area of study. First, we found the buy-in of faculty and administrators toward the library's assumption of digital curation duties to be positive. The only negative reaction we received was the concern that confidential records might be made publicly available and thus open the college to litigation. We resolved this problem by negotiating terms for dark archiving, which is discussed below.

Second, we found that establishing a simple digital archives following archival best practices could be done for a modest price. Expenses included purchasing the necessary storage drives for less than $200 and other equipment that we repurposed to host an archival server. Future equipment upgrades will be an ongoing cost absorbed by the library. The most significant project cost was staff time to process the digital records. Any institution, especially a small one, that seeks to create a digital repository must allow its librarians and archivists time away from their normal duties to devote to the project. In fact, we estimated that creating and maintaining a complete digital archives would evolve into a full-time position. Fortunately for the Maritime College Library, we were able to get the released time to devote to the initial project, although we are concerned that it will be difficult to maintain a comprehensive digital repository in the future without the commitment of the college's administration to pay for equipment and staff. However, we are optimistic that the initial digital archives will be seen as a tangible value to the institution and will garner administrative support.

Third, we learned that digital archives must be selective. Recognizing the information glut that would result from acquiring all, digital institutional records—many of them duplicates—at once, we plan to approach the acquisition of new materials with greater deliberation as proper workflows are established through individual record creators on campus. This we believe will ease the backlog on the archivists and not overwhelm them with creating a fully accessible digital archives. In effect, we made the decision to pursue only the most relevant records as agreed upon between committee chairpersons and the archivists.

Once the library successfully laid the groundwork for its digital archives, our next step has been to expand the program. Here we have encountered problems with the confidentiality of the records. We were most anxious to acquire administrative records and digital documents from other faculty committees. For each separate governing body, we established separate processing and preservation protocols.

In the case of administrative records, we were given a large set of documents with the proviso that the records be set in a dark archives for twenty years. The library's archival unit placed these records in a separate, password-protected dark archives with limited access. Original documentation of the accessioning terms was placed within the drive. The terms of the acquisition were also detailed in a password-protected spreadsheet so future archivists would know when the records can be made available to the public. While listed in the finding aid, these records are noted as restricted; they are not included in the full-text searchable database.

We made separate proposals to each chairperson of the other faculty committees. In all cases, we found acceptance of the importance of preserving the

records. Individual accessioning terms were negotiated with the chairpersons. For example, some committee chairs were open to immediate and full access, while others, due to the confidential nature of the records, recommended storage in a time-released dark archives, similar to what we had done with the administrative records. In all cases, committees then voted on the disposition of the records. While at the time of this writing some votes are still pending, those who have voted have agreed (with various conditions) to transfer their records to the digital archives.

With this successful start, our small academic library has re-established a traditional role in a new medium and provided a tangible benefit to the college community. Our project also has shown any library can do it with continued administrative support. Through collaboration and best practices, we believe that our experiences with digital curation will continue to educate our college on the value of digital records and the importance of their preservation.

Notes

[1] Brendan E. Asogwa and Ifeanyi J. Ezema, "The Challenges of Preservation of Archives and Records in the Electronic Age," *PNLA Quarterly* 76, no. 3 (2012): 117.

[2] Juan Voutssas, "Long-Term Digital Information Preservation: Challenges in Latin America," *Aslib Proceedings* 54, no. 1 (2012): 84.

[3] Christoph Becker, Hannes Kulovits, Mark Guttenbrunner, Stephan Strodl, Andreas Rauber, and Hans Hofman, "Systematic Planning for Digital Preservation: Evaluating Potential Strategies and Building Preservation Plans," *International Journal on Digital Libraries* 10 (2009): 133–57.

[4] Seamus Ross, "Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries," *New Review of Information Networking* 17, no. 1 (2012): 53–54.

[5] Ramesh C. Gaur and Manorama Tripath, "Digital Preservation of Electronic Resources," *DESIDOC Journal of Library and Information Technology* 32, no. 4 (2012): 293–301.

[6] Joyce Ray, "The Rise of Digital Curation and Cyberinfrastructure: From Experimentation to Implementation and Maybe Integration," *Library Hi Tech* 30, no. 4 (2012): 608.

[7] Consultative Committee for Space Data Systems, "Reference Model for an Open Archival Information System (OAIS)," *CCSDS 650.0-M-2*, June 2012, http://public.ccsds.org/publications/archive/650x0m2.pdf.

[8] Smita Chandra and Pratibha A. Gokhale, "Implementing Open Archival Information System Model for Digital Preservation at Indian Institute of Geomagnetism," *DESIDOC Journal of Library and Information Technology* 32, no. 4 (2012): 327–34.

[9] Voutssas, "Long-Term Digital Information Preservation," 83–96.

[10] Mike Kastellec, "Practical Limits to the Scope of Digital Preservation," *Information Technology and Libraries* (June 2012): 63–70.

[11] Gaur and Tripathi, "Digital Preservation," 293–301.

[12] Devin Becker and Collier Nogues, "Saving-Over, Over-Saving, and the Future Mess of Writers' Digital Archives: A Survey Report on the Personal Digital Archiving Practices of Emerging Writers," *The American Archivist* 75 (Fall/Winter 2012): 482–513.

[13] Christopher A. Lee and Helen Tibbo, "Where's the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills," *Archivaria* 72 (Fall 2011): 127.

[14] Nancy Allen, "Collaboration through the Colorado Digitization Project," *First Monday* 4, no. 6 (2000), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/755/664.

[15] Reinhard Altenhoner, "Tailoring Digital Preservation Services for Practice: Workflow Development and Cost Modeling," *DESIDOC Journal of Library and Information Technology* 32, no. 4 (2012): 302–5.

[16] Dijana Sabolovic-Krajina, "From Digitization of Local Newspapers to the Portal of Hometown Cultural Knowledge–A Step by Step Digitization Strategy Available to Small Communities" (paper, 78th IFLA General Conference and Assembly, Helsinki, 2012).

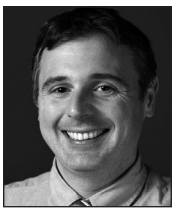
[17] Laural Carrol, Erika Farr, Peter Hornsby, and Ben Ranker, "A Comprehensive Approach to Born-Digital Archives," *Archivaria* 72 (Fall 2011): 61–92.

[18] Carrol et al., "A Comprehensive Approach to Born-Digital Archives."

[19] Margaret Hedstrom, "How Do Archivists Make Electronic Archives Usable and Accessible?," *Archives and Manuscripts* 2, no. 1 (1998): 6–22.

ABOUT THE AUTHORS

Elizabeth M. Berilla is currently the information literacy and assessment librarian at Misericordia University's Mary Kintz Bevevino Library, where she is also the library liaison to the University's College of Health Sciences. She has held the position since July 2013 and was previously the reference, instruction, and outreach librarian and institutional archivist at the State University of New York Maritime College in the Bronx. Berilla graduated with an MS (LIS) from Drexel University in 2011 with concentrations in youth services and archival studies.



Joseph A. Williams is currently the deputy director of the Greenwich Library, Greenwich, Connecticut. Formerly, he was the assistant director of SUNY Maritime College's Stephen B. Luce Library. He holds an MA in American history and an MLS, both from Queens College.