RESEARCH ESSAY:
DIGITIZATION FOR THE CONTEMPORARY ARCHIVIST

Stan Prager
15 Crescent Hill
East Longmeadow, MA 01028
413/219-8897
stan@gogeeks.com
American Public University
Graduate Degree Program: Public History
Student # 1058335
HIST522 C001 Spr 13: Archives and Manuscript Management
September 11, 2013

Research Essay: Digitization for the Contemporary Archivist

Introduction

My interest in Public History in general, and archiving more specifically, is focused upon digitization. Why? Actually, there are a number of reasons. First of all, it is critical to archival science in the twentieth-first century. Second, it is cutting-edge: who when entering a new field would not be attracted to the most innovative elements of that field? Third, there now exists the technology – both the equipment and the processing power -- to accomplish digitization on a wide scale in at least a somewhat economically feasible manner, something that was hardly true even a decade ago. Fourth, there is a compelling need for it: our paper records are deteriorating every hour through the entropy enshrined in the Second Law of Thermodynamics. Finally, it interests me: I own a computer business and am pursuing a Master's Degree in Public History – digitization of archival records is to my mind the perfect marriage of technology and history. I will address all of these issues to some degree in the course of this rather ambitious essay, although not in any particular order. For the sake of focus and a degree of brevity in what is, after all, a very large topic, I will confine the discussion of digitization to paper materials; digitizing analog audio and video resources is a huge sub-topic unto itself.

Digitization as a distinct component of archival science combines unique questions and challenges, some of which are often more subjective than objective. On the objective and scientific side, the obvious questions are whether there is the technology in place to achieve digitization: does the archive possess the scanning equipment and computers and media and web storage space to accommodate the needs of the proposed project? Also, is it economically feasible? Most digitization, especially when older, more fragile materials are involved, requires

extremely labor intensive hand-scanning in order to proceed. The subjective side poses its own challenges. It is not practical or feasible to digitize everything, so priorities must be set. What part of the archive should receive priority attention? Should it be the most cherished part of the collection? Should it be the part of the collection most popular with the benefactors of the archive (i.e. genealogical documents of prominent families) that dictates the level of their contributions? Should it be materials that require the least labor-intensive handling? Or should the priority be dictated by the condition of the materials: should the documents most in danger of deterioration always be first in line? These are complex questions that frequently result in very different answers for diverse facilities. Then there is the background noise of traditional archivists who are not eager to grasp the new technology and an ever-shrinking pool of scarce resources that must be allocated to the full constellation of competing needs within each institution, of which digitization is but one component.

In the course of this essay, I plan to explore the current science of archival digitization, on both the technical and non-technical levels. Since I became involved in Public History primarily to enter the profession in this specialty, the essay will then focus on these two of the requirements: *Select a general field or aspect of archival work that interest you and explain why* and *Identify the steps you would take to pursue a career as an archivist*. It should be noted that my discussion here is drawn not only from topics related to digitization covered in this course, but those encountered in other Public History courses I have taken at APUS, as well as other research I have conducted. I have made certain to cite both my own unpublished papers, as well as all of my other sources in the footnotes and bibliography. I do this both in the interest of full disclosure and academic integrity.

The Dynamics of Entropy

The Second Law of Thermodynamics defines entropy, which in layman's terms simply means that everything will eventually decay and deteriorate.[1] Absent forces such as earthquakes, volcanos and bulldozers, rocks deteriorate very slowly in terms of human lifetimes. Not so with other materials. Only a tiny fraction of all the wisdom of the ancient world that was once recorded on papyrus still remains extant. As one writer notes: "Nearly all papyri which survive today come from Egypt, although papyrus was in use throughout the ancient world. Papyrus is a perishable, organic material, and its survival through the ages has depended heavily on the climatic conditions found in a few regions of Egypt."[2]

If stored properly, the paper products utilized in modern times could last hundreds of years, perhaps longer, but the conditions in which the paper products are stored are critical to their long term survival.[3] I once wrote an unpublished research paper on book preservation, which examined all of the environmental impacts on printed matter that contributed to its deterioration.[4] Similar forces are at work upon manuscripts and all paper materials in general. These forces include temperature, relative humidity, exposure to light, mold, insect infestation, as well as the nature of the containers and shelving that the paper is stored in or upon. Moreover, the writing on such paper could become unreadable long before the paper itself completely deteriorates.

---

[1] Second Law of Thermodynamics, *All About Science*, http://www.allaboutscience.org/second-law-of-thermodynamics.htm, accessed September 11, 2013.

[2] "Where do the Papyri Come From?" Papyrology Collection, University of Michigan Library, http://www.lib.umich.edu/papyrology-collection/where-do-papyri-come, accessed September 11, 2013.

[3] "The Deterioration and Preservation of Paper: Some Essential Facts," *Preservation*, Library of Congress, http://www.loc.gov/preservation/resources/care/deterioratebrochure.html, accessed September 11, 2013.

[4] Some of the material in this section of the essay relies upon previous research conducted and previously presented in an unpublished paper for APUS: Stan Prager, *Turning the Pages of Time: A Handbook for Book Preservation*, unpublished paper for HIST521 C001 Win 12: Public History, American Public University, October 12, 2012.

Let's just take climate and its impact upon books as one example. High temperatures and high relative humidity present the greatest threats to the integrity of books and paper. Temperatures above 72˚F (22˚C) causes degradation of paper, adhesives and other materials. Theoretically, the rate of deterioration doubles for every 10˚F increase in temperature! High humidity (above 60%RH) distorts covers and paper and can, as noted earlier, promote mold growth that causes permanent damage. Inactive mold spores are ever present, but only spring to life with high humidity, growing and digesting ". . . its substrate, causing permanent weakening and staining . . ." in books. Acidification is the chief concern, and high temperatures and relative humidity will cause the chemicals in paper to acidify, often in proportion to these climatic elements. While not as frequently discussed in the literature, low temperatures (below freezing) can cause adhesives to fail and produce condensation, while extremely low humidity (below 30%RH) can cause covers to warp and book materials to become brittle. Low humidity is sometimes characteristic in winter months in buildings with central heating.[5]

Still, it turns out that the most important overall climate factor is consistency that avoids extreme fluctuations in temperature. Dramatic changes in temperature and relative humidity can have a devastating effect upon a collection. Here's the boring science part: books are ". . . hydroscopic, readily absorbing and releasing moisture. They respond to diurnal and seasonal changes in temperature and relative humidity by expanding and contracting . . . [such] Dimensional changes accelerate deterioration." So the key is limiting extreme swings in temperature while attempting to keep to lower temperatures and relative humidity ratios. Of

[5] "Preserving Books in Your Home Library," Cornell University Library, http://www.library.cornell.edu/preservation/publications/PreservingBooks.pdf, accessed September 4, 2012, 10; "Preserving Personal Books and Papers," American Museum of Natural History. http://library.amnh.org/library-services/conservation/preserving-personal-books-and-papers, accessed September 4, 2012, 1; "Book Care for Bibliophiles," Tappin Book Mine, http://www.tappinbookmine.com/bookcar2.htm, accessed September 5, 20122; Alison Walker, "Basic Preservation," British Library Preservation Advisory Center, 2010. http://www.bl.uk/blpac/pdf/basic.pdf, accessed September 5, 2012, 7.

course, it should not be overlooked that human beings must abide comfortably in this same environment.[6] Add to this the dangers of light exposure, mold, insects, mice and other threats and the acceleration of entropy is clear. Modern archives might store all books and other materials properly in state of the art facilities, but since these were not always stored optimally in such facilities, the conditions of archival materials typically vary significantly.

The point of all this is that manuscripts, letters, diaries, records, books and ephemera all exist on borrowed time. This fact makes the science of archival digitization very current, very compelling and very exciting. At the same time, it is also overwhelming. As noted previously, it is not only impractical but probably impossible to digitize everything currently stored in archives across the country. The cost would be exponential to digitize even a fraction of that. Thus, it is critical that priorities should be set. I raised the issue earlier of the competing forces that often scramble to elect their "nominee" to be first in line for digitization; the selected material may not always be the best candidate. But let us assume for a moment that the wisest heads prevail and the materials selected for digitization qualify not only in overall historical value but also are the most endangered due to age and/or condition. What now? It is all about time and equipment.

Scanning Equipment and Technical Considerations

When the first scanners showed up in homes as a peripheral to the personal computer in the mid-nineties, these seemed quite miraculous. A faded baby picture of a long deceased grandmother could be scanned in, and even enhanced with photo-editing software that was often bundled with the scanner itself. Likewise, Ellis Island passage tickets, deeds for homes long

[6] "Preserving Personal Books and Papers," 1; "Caring For Your Treasures: Books – How To Protect Your Books," American Institute for Conservation of Historic and Conservation Works (AIC), http://www.conservationus.org/index.cfm?fuseaction=Page.viewPage&pageId=626&parentID=497, accessed September 5, 2012, 1; "Preserving Books in Your Home Library," 10; Walker, 7.

demolished and birth certificates could all be captured and saved digitally. Facsimiles of precious pictures or documents could be printed and hung up in sunlit rooms without fear of damage because the original was safe, saved to a hard drive or floppy disk – and later to other media such as CD's, DVD's, external hard drives, or even backed up on the web. Over time, optical character recognition – OCR – software was introduced and gradually got better and better, so that printed matter, if not handwritten letters, could be scanned into a format that was searchable and editable. Today, almost every $99 printer sold at retail contains a scanner that if not top-quality is at least quite serviceable for most users. While these are not generally suitable to professional use at an archive, this story points to how rapidly scanning technology has been integrated into our everyday lives. Somewhat more expensive, better quality scanners and more powerful OCR software is available for professional use. With the exception of certain more specialized machines, the costs involved for such equipment is relatively modest.[7]

For instance, hardware recommendations set forth by Paul Royster in his highly regarded article "The Art of Scanning," highlight the value of the Microtek *ScanMaker i700* flatbed scanner. Royster notes that while this scanner is a little slow on the performance side, the accuracy is superb and at the relatively low cost of just under three hundred dollars a unit, most budgets will permit purchase of several of these models.[8] The Microtek *ScanMaker i700* is a very high quality 48-bit scanner with 9600 x 4800 dpi capabilities and a scanner bed that will accommodate legal size (8½"x14") originals. According to manufacturer specifications, "ScanMaker i700 with 128,160-element staggered CCD reaches the resolutions of up to 4800

[7] Some of the material in this section of the essay relies upon previous research conducted and previously presented in an unpublished paper for APUS: Stan Prager, *Grant Narrative: Historic East Longmeadow*, unpublished paper for HIST636 A001 Win 2013: History & Digital Preservation, American Public University, April 18, 2013.
[8] Paul Royster, "The Art of Scanning," *DigitalCommons@University of Nebraska – Lincoln*, (1-5-2011), University of Nebraska – Lincoln, http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1097&context=ir_information, accessed March 1, 2013.

pixels per inch with high precision. Through its built-in 48-bit ADC (Analog to Digital Converter), ScanMaker i700 can convert originals into digital data in high fidelity format. It is capable of delivering full 48-bit depth color for richer and smoother tonal values." The bundled software is generous, and includes Microtek *ScanWizard 5* with DIGITAL ICE Technology, *ABBYY FineReader Sprint OCR*, and *Adobe Photoshop Elements 2*. Of most significance is the *ScanWizard 5*, which according to the manufacturer integrates: ". . . DIGITAL ICE Technology and ColoRescue™ Technology [which] automatically maps, identifies, and eliminates surface defects on photos and film that dramatically decreases the time to reconstruct surface-damaged prints."[9] It is often surprising that less than a thousand dollars will bring an institution three scanners of this quality along with the software necessary to do the job. But the cost, of course, is in the labor, as a human being must carefully manipulate a fragile document or photograph and manage the scanning effort.

There exist far more advanced and significantly more expensive equipment that will cut down on labor costs but are not typically suitable to fragile materials or institutions with tight budgets. For instance, if the goal is to scan in books at high speeds, there are top-shelf scanners like the "DL-mini" from *4DigitalBooks* that boasts cutting operator time by ninety percent using an air-cushion automatic page-turner that is designed to minimize stress on the books.[10] Its big brother, the DL-3000 also features the same ". . . patented air-flow page turning process and air cushioned book cradle [to] ensure a gentle handling of the books without applying pressure or tension on the spine," and claims speeds of 3000 pages an hour.[11] The DL-3000 costs about a

---

[9] ScanMaker i700, Microtek Products, http://ww1.microtek.com/Mii/Web/Product.php?Product=Detail&P_Id=105, accessed April 14, 2013.
[10] DL-mini, *4DigitalBooks*, http://www.4digitalbooks.com/_auto_dlmini.html, http://www.4digitalbooks.com/_auto_dlmini_minfo.html, accessed September 11, 2013.
[11] DL-3003, *4DigitalBooks*, http://www.4digitalbooks.com/_auto_dl3000_minfo.html, http://www.4digitalbooks.com/_auto_dl3000.html, accessed September 11, 2013.

quarter of a million dollars, so only a very large institution with deep pockets could afford one. However, as writer Andrew Liszewski argues, the long-term per item cost is a terrific value: ". . . given the fact that the DL 3000 can operate 24 hours a day for up to 6 million cycles before needing maintenance, it can bring the production costs for scanning a book down to about 2 cents per page. As a result, even though the machine has a price tag of about $250,000, that cost can be financially amortized after about 3 million scanned pages."[12] Still, it seems to me that even with the "air-flow page turning process" this is a mass production machine not at all suitable for the old and fragile. More specialized automatic scanning designed specifically for more delicate materials is the Kirtas *KABIS I* from Kirtas Technologies, which offers: ". . . the SmartCradle™ with new automatic book centering that increases post-processing productivity. The 110 degree position is optimal for low stress, gentle book positioning and is ideal for fragile and rare books. Air flows and the Enhanced Page Separator combine to separate pages, and the newest SureTurn™ robotic arm with its vacuum head will gently lift and turn the page. Within the vacuum head is our Page Edge Sensor technology to detect multi-page or no-page lift conditions and automatically take programmable corrective action."[13] There are several models for KABIS scanners, with scanners ranging in price ". . . from $69,000 to $129,000," which again make these cost prohibitive for all except larger institutions.[14]

There are many more product offerings on the high-end that we could examine, but even with these labor-saving devices, much of the brittle stuff must be utilized in manual mode,

---

[12] Andrew Liszewski, "Digitizing Line DL 3000 Book Scanner Is Big And Fast," *Oh Gizmo*, 04/24/08, http://www.ohgizmo.com/2008/04/24/digitizing-line-dl-3000-book-scanner-is-big-and-fast/, accessed September 11, 2013.

[13] Kirtas KABIS I, Kirtas Technologies, http://www.i2s-digibook.com/products/digitize/automatic-book-scanner/kirtas-kabis-i/, accessed September 11, 2013.

[14] David Rapp, "Product Watch: Library Scanners," Library *Journal*, June 14, 2011, http://lj.libraryjournal.com/2011/06/industry-news/product-watch-library-scanners/#_, accessed September 11, 2013.

whether on a $300 Microtek *ScanMaker* or a $250,000 DL-3000. Kirtas puts the labor costs in perspective in its own promotional materials: “At an image resolution of 300 to 500 dpi. Kirtas estimated in 2005 that their automated method costs "as low as $.03" per page ($36 per hour), while manual scanning, at a rate of 100 to 150 pages per hour, costs "$.35 to $1.50" per page.”[15] The actual cost for each facility that houses archives being scanned, of course, is governed by the hourly wage of the employee and their productivity. Humans get tired, as well, and must take breaks!

Moreover, one size does not fit all in scanning historical materials. It may turn out that some artifacts require use of a hand-held scanner or perhaps a digital camera because of their specific shape and/or size, legibility or fragile nature; these must be set aside and addressed on an individual basis. Constraints of budget, labor and time allotted means that decisions must be based on what can best be accomplished for the majority of the artifacts with the resources at hand. It is possible that unique circumstances may present themselves that will warrant outsourcing digitization of a specific artifact to an outside vendor. For instance, if it turns out that a high quality DSLR camera or hand-held scanner will be needed for a minority of artifacts that are not suitable for the standard scanner used in the facility for most scanning, rather than purchase that equipment for a limited number of items, the institution might seek to either rent that equipment or simply bring in an outside vendor to complete the process in this regard.

The discussion of equipment has only barely touched on scan resolution, a critical component of the digitizing process with visual materials such as manuscripts, printed matter and photographs, and a consideration that governs both methods and equipment. As such, extensive time must often be invested in preparing the materials for scanning. In fact, project preparation can actually be more labor intensive and time consuming than the actual digitization stage, but

[15] Digitizing Initiatives, http://www.tfaoi.com/aa/5aa/5aa60.htm, accessed September 11, 2013

this is the most critical phase and there is no room for shortcuts or lack of diligence in this regard. This is due to the fact that only rarely is there consistency in the condition of a bulk of historic material designated for digitization.

For documents and printed matter, the preparation process consists of first identifying the quality of the original typically utilizing the standards outlined in the "Guidelines" set forth by the Still Image Working Group in the Federal Agencies Digital Guidelines Initiative (FADGI), which assigns values based upon image quality of the original using a scale (T.1, T.2, T.3, etc.).[16] These "scores" will then be incorporated into the metadata (more on this later) compiled for tags for each printed artifact. The goal for printed material when possible will be capturing these in searchable PDF/A-1, and the system of assigning values established by FADGI will provide critical information as to which printed materials are immediate candidates for such consideration. As such, documents, letters and news clippings should be scanned in uncompressed TIFF images in 1-bit or bitonal with a 600 dpi in black and white or 8- bit 300 dpi in greyscale, following NARA recommendations to defer to the standards of the Digital Library Federation (DLF) to create raster images later suitable for OCR creation of searchable PDF/A-1 format.[17]

This is an appropriate place to discuss the greater ramifications of OCR. Despite the temptation of many institutions involved in digitization of printed matter to rush into OCR creation of searchable PDF's, this is a complicated process from a time and budget perspective

---

[16] "Guidelines: Content Categories & Digitization Objectives: Reformatting Historical Printed Matter, Documents and Manuscripts, and Pictoral [sic] Materials — Content Categories and Subcategories table," Still Image Working Group, *Federal Agencies Digital Guidelines Initiative*, http://www.digitizationguidelines.gov/guidelines/ccdo-subcat.html, accessed April 12, 2013.

[17] U.S. National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images, http://www.archives.gov/preservation/technical/guidelines.pdf, accessed April 14, 2013, 54; The Digital Library Federation Benchmark Working Group (2001-2002), "Benchmark for Faithful Digital Reproductions of Monographs and Serials," (Version 1. December 2002), Digital Library Federation (DLF), http://old.diglib.org/standards/bmarkfin.pdf, 5, accessed April 14, 2013.

and should be treated as an entirely separate component of the digitization process. As Rose Holley points out in her outstanding article on OCR accuracy, "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs," most OCR software claims a ninety-nine percent accuracy rate, yet this optimistic figure applies to clean and clear modern printed matter with recognizable typefaces. Such accuracy can plummet to as low as seventy-one percent with historic newspapers, which translates into: ". . . 145 incorrect characters in an average paragraph of 500 characters . . . 29% of the paragraph would be incorrect."[18] Accuracy is potentially even more unreliable with archaic documents on yellowed and faded paper. Thus, the critical first step must involve creating a digital image of the printed material of high-quality that can later be subjected to OCR analysis and publication, optimally juxtaposed in the online database with the original image.

For digital images, a sound strategy might be to follow the digitization standards and best practices employed by the Smithsonian Institution Archives, which is based upon the Federal Digitization Guidelines Initiative.[19] This utilizes the TIFF (Tagged Image File Format) format with a 24 bit RGB setting, with resolution standards of a minimum of 660 *ppi* and 6,000 pixels on the long side of the image.[20] For the more technical aspects of achieving goals using best methods and practices, institutions often rely upon the U.S. National Archives and Records

[18] Rose Holley, "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs" *D-Lib Magazine*, March/April 2009, Volume 15 Number 3/4,ISSN 1082-9873, http://www.dlib.org/dlib/march09/holley/03holley.html, accessed February 28, 2013.
[19] "Digitization," *Smithsonian Institution Archives*, http://siarchives.si.edu/services/digitization, accessed April 11, 2013; "Digital Imaging Framework," http://www.digitizationguidelines.gov/guidelines/DIFfinal.pdf, accessed April 11, 2013.
[20] "Digitization," *Smithsonian Institution Archives*.

Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access.[21]

Once digitized, materials must be catalogued properly so they can be indexed and searched appropriately. The encoding that accompanies digitized materials is called metadata, which is basically data about data. One definition is: "Metadata describes other data. It provides information about a certain item's content. For example, an image may include metadata that describes how large the picture is, the color depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document."[22] This can be a complex subject replete with technical jargon that I will simply summarize here, but the need for proper metadata that goes hand in hand with digitization should not be overlooked.

In general, an archivist will attempt to obtain identification for each artifact that will contain, at minimum: Title, Subject, Description, Date, Owner of Artifact, Location of Artifact, How Artifact Obtained, Type, Format, Identifier, Language, Image Score Quality, and technical details of the digitization; other data may be added as appropriate. For encoding, most utilize XML and follow the standards of the Library of Congress XML MIX Schema Version 2.0 known as "NISO Metadata for Images in XML (NISO MIX)."[23] More technical information can

[21] U.S. National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images, http://www.archives.gov/preservation/technical/guidelines.pdf, accessed April 14, 2013.

[22]Metadata, *TechTerms*, http://www.techterms.com/definition/metadata, accessed September 11, 2013.

[23] "MIX NISO Metadata for Images in XML Schema," *Standards*, The Library of Congress, http://www.loc.gov/standards/mix/, accessed April 14, 2013;

be sourced in the foundational manual for this technology, *Data Dictionary - Technical Metadata for Digital Still Images*.[24] Metadata can, of course, be revised at a later date.

Once digitized, images are typically stored to hard drives, backed up to media and uploaded to the web – either an institutional intranet or the internet itself, with access often public but sometimes proprietary. Grants for digitization frequently require public access to these materials once uploaded. More critically, the digital repository where the materials are consigned must be trustworthy and reliable. As one writer put it: "Digital information is so complex and vast that no one institution—or even a hundred institutions—can be responsible for the preservation of the world's digital cultural heritage. In a world bound by a complex array of legal, ethical, cultural, and economic obligations, the imperative of long-term access to information further complicates the roles and responsibilities of digital repositories."[25] There are existing specific guidelines for digital repository certification per NARA standards.[26]

Another issue the archivist involved with digitization must address is copyright law. Does the institution possess the right to digitize the material and make it available to the public? This is a highly complex legal issue that defies simple definition, but one that the archivist must address rather than risk inadvertent improper, illegal and unethical actions in course of digitizing a collection.[27]

---

[24] *Data Dictionary - Technical Metadata for Digital Still Images*, (ANSI/NISO Z39.87-2006), December 18, 2006, National Information Standards Institute, http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-87-2006.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVaGZww13HuDlSn6cvwjex0ejiIKSaTYlErPbfamndQa6zkS6rLL3oIr, accessed April 14, 2013.

[25] "Digital Repository Certification," OCLC Research, http://www.oclc.org/research/activities/repositorycert.html, accessed September 11, 2013.

[26] "Trustworthy Repositories Audit & Certification: Criteria and Checklist," OCLC and CRL, February 2007, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf, accessed September 11, 2013.

[27] June M. Besek, "Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment," Council on Library and Information Resources, 2003, http://www.clir.org/pubs/reports/pub112/body.html, accessed September 10, 2013.

Space constraints in this essay prohibit me from exploring many of these related, relevant issues that I have raised in greater detail, but I did not want to fail to bring these up as relevant topics that the archivist involved in digitization must address.

Stan Prager: Digital Archivist

It is my intention, once I complete my two remaining courses, to obtain a Practicum to fulfill the remaining requirements for my Master's Degree in Public History from American Public University that will directly involve me in digitization at an archive, library, museum or historical society. Since I intend to continue as owner, operator and entrepreneur at *GoGeeks Computer Rescue*, the computer repair and manufacturing business that I own, as a first step after graduation I hope to obtain a permanent part-time position with a similar institution that permits me to immerse myself in archival digitization.[28]

In the course of my studies, I have known ever since I switched my major from "Ancient & Classical Studies" to "Public History" that digitization would captivate my interest. As such, I have attempted to devote a little extra time, in each of my Public History courses, to digitization as it relates to the greater themes of history, archives, preservation, and the application of Public History to the "real world." Traditional archives had little concern for digitizing records, but the confluence of technological advances, the pressures of material deterioration and researchers' growing demands for access to archival material on a greater scale than ever before have combined to create an excitement and sense of urgency in the archival world for digitization. I find myself caught up with that same excitement, that same urgency, and I can hardly wait to complete my studies so I can get started with digitization on a professional level.

---

[28] GoGeeks Computer Rescue, http://www.gogeeks.com/, accessed September 10, 2013.

Bibliography

Besek, June M. “Copyright Issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment,” Council on Library and Information Resources, 2003, http://www.clir.org/pubs/reports/pub112/body.html, accessed September 10, 2013.

“Book Care for Bibliophiles.” *Tappin Book Mine*. http://www.tappinbookmine.com/bookcar2.htm, accessed September 5, 2012.

“Caring For Your Treasures: Books – How To Protect Your Books.” American Institute for Conservation of Historic and Conservation Works (AIC). http://www.conservation-us.org/index.cfm?fuseaction=Page.viewPage&pageId=626&parentID=497, accessed September 5, 2012.

Data Dictionary - Technical Metadata for Digital Still Images, (ANSI/NISO Z39.87-2006), December 18, 2006, National Information Standards Institute, http://www.niso.org/kst/reports/standards/kfile_download?id%3Austring%3Aiso-8859-1=Z39-87-2006.pdf&pt=RkGKiXzW643YeUaYUqZ1BFwDhIG4-24RJbcZBWg8uE4vWdpZsJDs4RjLz0t90_d5_ymGsj_IKVaGZww13HuDlSn6cvwjex0ejiIKSaTYlErPbfamndQa6zkS6rLL3oIr, accessed April 14, 2013.

“The Deterioration and Preservation of Paper: Some Essential Facts,” *Preservation*, Library of Congress, http://www.loc.gov/preservation/resources/care/deterioratebrochure.html, accessed September 11, 2013.

“Digital Imaging Framework.” http://www.digitizationguidelines.gov/guidelines/DIFfinal.pdf, accessed April 11, 2013.

The Digital Library Federation Benchmark Working Group (2001-2002). “Benchmark for Faithful Digital Reproductions of Monographs and Serials.” Version 1. December 2002. Digital Library Federation (DLF), http://old.diglib.org/standards/bmarkfin.pdf, 5, accessed April 14, 2013.

“Digital Repository Certification,” OCLC Research, http://www.oclc.org/research/activities/repositorycert.html, accessed September 11, 2013.

“Digitization.” *Smithsonian Institution Archives*. http://siarchives.si.edu/services/digitization, accessed April 11, 2013.

Digitizing Initiatives, http://www.tfaoi.com/aa/5aa/5aa60.htm, accessed September 11, 2013

DL-3003, *4DigitalBooks*, http://www.4digitalbooks.com/_auto_dl3000_minfo.html, http://www.4digitalbooks.com/_auto_dl3000.html, accessed September 11, 2013.

DL-mini, *4DigitalBooks*, http://www.4digitalbooks.com/_auto_dlmini.html, http://www.4digitalbooks.com/_auto_dlmini_minfo.html, accessed September 11, 2013.

GoGeeks Computer Rescue, http://www.gogeeks.com/, accessed September 10, 2013.

"Guidelines: Content Categories & Digitization Objectives: Reformatting Historical Printed Matter, Documents and Manuscripts, and Pictoral [sic] Materials — Content Categories and Subcategories table." Still Image Working Group. *Federal Agencies Digital Guidelines Initiative*. http://www.digitizationguidelines.gov/guidelines/ccdo-subcat.html, accessed April 12, 2013.

Holley, Rose. "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs." *D-Lib Magazine.* March/April 2009, Volume 15 Number 3/4, ISSN 1082-9873. http://www.dlib.org/dlib/march09/holley/03holley.html, accessed February 28, 2013.

Kirtas KABIS I, Kirtas Technologies, http://www.i2s-digibook.com/products/digitize/automatic-book-scanner/kirtas-kabis-i/, accessed September 11, 2013.

Liszewski, Andrew. "Digitizing Line DL 3000 Book Scanner Is Big And Fast," *Oh Gizmo*, 04/24/08, http://www.ohgizmo.com/2008/04/24/digitizing-line-dl-3000-book-scanner-is-big-and-fast/, accessed September 11, 2013.

Metadata, *TechTerms*, http://www.techterms.com/definition/metadata, accessed September 11, 2013.

"MIX NISO Metadata for Images in XML Schema." *Standards*. The Library of Congress. http://www.loc.gov/standards/mix/, accessed April 14, 2013.

Prager, Stan. *Grant Narrative: Historic East Longmeadow*, unpublished paper for HIST636 A001 Win 2013: History & Digital Preservation, American Public University, April 18, 2013.

Prager, Stan. *Turning the Pages of Time: A Handbook for Book Preservation.* Unpublished paper for HIST521 C001 Win 12: Public History, American Public University, October 12, 2012.

"Preserving Books in Your Home Library." Cornell University Library. http://www.library.cornell.edu/preservation/publications/PreservingBooks.pdf, accessed September 4, 2012.

"Preserving Personal Books and Papers." American Museum of Natural History. http://library.amnh.org/library-services/conservationpreserving-personal-books-and-papers, accessed September 4, 2012.

Rapp, David. "Product Watch: Library Scanners." *Library Journal*, June 14, 2011, http://lj.libraryjournal.com/2011/06/industry-news/product-watch-library-scanners/#_, accessed September 11, 2013.

Royster, Paul. "The Art of Scanning." *DigitalCommons@University of Nebraska – Lincoln.* 1-5-2011. University of Nebraska – Lincoln, http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1097&context=ir_information , accessed March 1, 2013.

ScanMaker i700. Microtek Products. http://ww1.microtek.com/Mii/Web/Product.php?Product=Detail&P_Id=105, accessed April 14, 2013.

Second Law of Thermodynamics, *All About Science*, http://www.allaboutscience.org/second-law-of-thermodynamics.htm, accessed September 11, 2013.

"Trustworthy Repositories Audit & Certification: Criteria and Checklist," OCLC and CRL, February 2007, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf, accessed September 11, 2013.

U.S. National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images. http://www.archives.gov/preservation/technical/guidelines.pdf, accessed April 14, 2013.

Walker, Alison. "Basic Preservation." British Library Preservation Advisory Center, 2010. http://www.bl.uk/blpac/pdf/basic.pdf, accessed September 5, 2012.

"Where do the Papyri Come From?" Papyrology Collection, University of Michigan Library, http://www.lib.umich.edu/papyrology-collection/where-do-papyri-come, accessed September 11, 2013.