

September 16, 2005

Basic Considerations For Digitization: Providing Access To Special Collections - A Symposium

Medical Heritage Center, The Ohio State University

The Ohio Preservation Council and the Ohio Library Council sponsored a symposium, *Basic Considerations for Digitization: Providing Access to Special Collections*, at the Medical Heritage Center on The Ohio State University campus on September 16, 2005. Approximately sixty people participated in the event, which featured four highly respected individuals with a wide variety of special expertise providing presentations identifying key issues associated with digitization. The presenters form a panel in the afternoon to discuss specific questions and issues raised by symposium attendees. The panel was moderated by Mr. Eric Honneffer, Document Conservator at Bowling Green State University and Ms. Angela O'Neal of the Ohio Historical Society Archives.

The goal of the symposium was to provide a basic introduction to evaluating whether to undertake a digitization project, and if so, how to begin. The importance of preservation and access were the focus. Digitization is not an answer to preservation, but a means to enable access to information so the original can be preserved. Many issues need to be considered before embarking on a digitization project, such as how to: select materials appropriate for digitization; determine the appropriate digitization method when there are no uniform standards; maintain the integrity of the original both during and after digitization; ensure and maintain the quality of the digitized material; and provide long-term access to digitized material. The costs involved to digitize and provide access; preserve the original material; bring digitized information forward with each hardware and software technology advance are other issues requiring careful thought and planning.

Following are brief summaries of the presentations provided by Ohio Preservation Council members:

Susan Allen Digital Library Leader, Worthington Public Library

Susan Allen is the Digital Library Leader at Worthington (OH) Public Library where she coordinates the Worthington Memory Project. She has worked as a reference librarian at Worthington's Northwest Library and the Carnegie Library of Pittsburgh. She worked on the HELIOS digital imaging project at Carnegie Mellon University and the Historic Pittsburgh Online Finding Aids project at the University of Pittsburgh. She holds a BA degree from The Ohio State University, an MLIS degree from the University of Pittsburgh, and took additional training in digital imaging and preservation management at Cornell University.

The vision for the Worthington Memory project began in the late 90's. Ms. Allen listed the life-cycle of a digitization project as a five step process:

1. *Pre-Project Activities* - includes determining your goals, identifying your sources of funding, securing commitment from your institution, and project planning. Pre-project planning can take years before any scanning actually takes place. Planning will include questions such as; who will do the work? What will we put in the project? Other elements to plan for are copyright issues, funding, equipment, cataloging and metadata, benchmarking, online access and storage.
2. *Ramping Up* - the time of the project initiation and when you begin to scan.
3. *Production*- the time of your greatest productivity.
4. *Project Wind Down*- concluding the project and dealing with problems that have been set aside.
5. *Post-Project Activities* - maintenance responsibilities for digital products.

Ms. Allen discussed project funding. "Grants are great if you can get them, but learn to make do without extra money." As far as personnel goes, Worthington utilized volunteers. Other options would be to hire extra staff and/or divert existing staff. Partnerships are also very helpful. Worthington Public Library partnered with the Worthington Historical Society and the City of Worthington Bicentennial from the beginning of their project. Outsourcing might also be an option for your project; however Ms. Allen cautioned that while you might save money, the rare and fragile items are no longer in your control and you risk damage when they are sent off site.

Selecting items to scan for the project can be done in much the same way as selecting items for preservation. Look at the physical condition, the historical significance and value, and if you are duplicating someone else's effort. Take into consideration the item's potential for digitization. Will it have low contrast or colors that will present a challenge? What about the ownership (property rights and copyright)? You will need to develop a policy first for legal issues. A good source for this is www.copyright.gov. Ms. Allen emphasized that you must use standards, known as benchmarking. This assures quality throughout the scanning project and long-term access.

Equipment needs will include a scanner, PC's, monitors and printers, workstations, and software. You will also need a network, a way for storage and backups, and a web server.

Ms. Allen discussed metadata, and noted that the Worthington Public Library used the Dublin Core system. She explained that the Dublin Core had many useful features, and is easy to create and manage. There is a Dublin Core Assistant site at www.ukoln.ac.uk/metadata/dcassist/. You will use an image production record which will follow each item from beginning to end, with information such as title, author, format, date, physical condition, etc. The record will have the digitization information including date of scan, person scanning, image file name, and many other useful items.

The Worthington Memory Project ran into some challenging situations. Oversize items (larger than 12"x17") had to be scanned in parts. Works of art and objects were difficult to scan and needed to be photographed then the photographs scanned. This presented a problem with lighting and zooming in on detail. Defining the scope of "local history" was also a challenge.

Ms. Allen's last point was on long-term access. "You need to think about this from the beginning" she advised. Early planning will mean a smoother transition. She offered three web sites for help with digital projects: the Cornell Digital Preservation Management www.library.cornell.edu/iris/tutorial/dpm, the RLG-OCLC Trusted Digital Repositories at www.rig.org/longterm/repositories.pdf [link does not work], and the Digital Preservation Coalition at www.dpconline.org/graphics/index.html.

Amy McCrory Project Archivist, The Ohio State University Cartoon Research Library

Amy McCrory is a Project Archivist at the Ohio State University Cartoon Research Library and an Encoded Archival Description (EAD) Specialist for the University Libraries. At the Cartoon Research Library she processes and describes large collections of visual materials, incorporating images of cartoons and cartoon-related merchandise with the collections' electronic finding aids. McCrory is chair of the Society of American Archivists EAD Roundtable, and a member of the Visual Materials Section.

Ms. McCrory stressed the importance of planning the digitizing project beforehand and thinking it through. She used as examples two projects she recently experienced. Ohio State University acquired the Bill Blackbeard collection from the San Francisco Academy of Comic Art. The collection contained seventy-five tons of classic comic strips collected over a 30-year period. It contained a variety of materials including bound newspapers in various sizes, Victorian story papers, elaborate illustrations, and editorial cartoons. Each type required a different set of considerations.

A second collection was that of cartoonist agent, Toni Mendez. In 2003 the Mendez estate donated merchandizing materials to the library. The budgets for these two projects were \$115,000 and \$100,000 respectively. No staff worked on the projects full time. When the budget must include staff costs, it is important that staff time be well used. Workflow must be cost efficient. Hiring student assistants with skills such as photography and art is useful. Student assistants need daily instruction so providing good documentation for them is important. Constant communication at all levels is crucial. Management must be informed as to what is possible within the project.

Ms. McCrory discussed equipment. The web should be used to find other similar projects and equipment used. Equipment should not be replaced until necessary. By "shopping around" a great deal can be found. It might be possible to duplicate some situations without having to buy expensive items. As an example, light boxes can be made from materials that might already be owned. When "shopping around" do not hesitate to ask questions. Free advice may be available from stores.

It is important to remember the objective of the project and to stick to it. Consider imaging standards and best practices and decide the highest level that can be achieved. Storage space is another important consideration. The larger files require much longer editing time. An archival copy and a use copy are necessary. The archival copy needs

to be stored where it will last. The most detailed metadata that can be afforded should be used. Be certain, however, that enough metadata is used to be adequate.

Geoffrey D. Smith

Head of the Rare Books & Manuscripts Library of the Ohio State University Libraries

Dr. Geoffrey D. Smith is Head of the Rare Books and Manuscripts Library at The Ohio State University Libraries. He is a graduate of Tufts University and holds his doctorate from Indiana University where he worked with the Lilly Library and the William Dean Howells scholarly edition, which influenced his career in rare books, bibliography and textual editing. Smith was the first Curator of the William Charvat Collection of American Fiction at The Ohio State University Libraries. He has published numerous publications on library topics and is active in local, regional and national bibliographical and scholarly associations. Smith is a founding member and first president of The Aldus Society, a Columbus-based bibliophilic society.

Dr. Smith acknowledged that digitization has become a national priority. But how do you decide what to do first? Do you digitize for commercial success or self-interests? Many libraries have chosen to digitize their most unique items, such as Ancient scrolls or a Gutenberg Bible. Others have chosen to do a core collection of titles that are important to America's literature of our historical past, e.g. University of Michigan's Making of America project. These volumes are of older well-known texts to which people need full-text access.

What you need to consider is which items get used most often, their condition, the physical stress being placed on them when they are used, and storage of the original. Dr. Smith decided to digitize a volume that was in poor physical condition because it was required reading for a course. It was necessary to take the volume apart to repair it so it was a good time to scan it. The result was access for those who needed it, and preservation of the original document.

Marketing is another reason to digitize an item, to make people more aware of what is in your collection. Researchers and academics are constantly seeking knowledge and reliable information. What items in your collection would be most useful to them? Many manuscript collections are unknown and inaccessible because they have not been processed or their inventories are inaccurate or outdated. Putting them on the web will make them more available. New collections are always coming in. World War II veterans are dying and leaving diaries of their time on the battlefield. These diaries will increase our understanding of the war. Local history and genealogical resources are a popular area for digitization.

Some issues to consider: hire someone to do the project, do not neglect the rest of the collection just for the one that is being digitized, and provide proper storage containers and conditions for the original.

Judith Cobb

OCLC Digital Collection and Metadata Services Education and Planning

Judith Cobb is the Senior Product Specialist for the OCLC Digital Archive. Currently she is managing a Library of Congress National Digital Information Infrastructure Preservation Program, which was awarded to the University of Illinois and OCLC, as well as managing the operations of the OCLC Digital Archive. She specializes in archives, digital projects and digital preservation issues. She has developed workshops and presentations in several subject areas of digital preservation including metadata, records management, web archiving, preserving digital images and digital preservation strategies. Ms. Cobb provides consulting services to organizations in the areas of archives and digital preservation. Previously she worked in the State Archives of Ohio for eight years, where she was the Assistant State Archivist.

Ms. Cobb's presentation centered on digital collections and metadata services, including education and planning. She emphasized that accessibility, sustainability and preservation are linked. Years ago, microfilming was a way of preservation, but this is gradually being replaced by digitization. Maintenance of image quality is a major issue; therefore digitization should be an institutional decision. An institution should balance its projects in order to be able to access them, to sustain them and to preserve them. The size of the project determines the decisions that have to be made.

Access is a short-term need, so image quality is not a primary concern. For access copies only, master files are not created and maintained. Therefore, images can be low quality as needed. Metadata also is minimal or as needed, however, descriptive metadata is most vital and should be based on user needs. For back-up copies no special procedures are required within the existing system. Business planning can be done as needed to facilitate the project while resource allocation can be planned for a slightly longer term. Finally, the selection process can proceed based on user and the scope of the project.

Sustainability of a project is of medium term; therefore the image should be of high quality, with resolution and bit depth appropriate to the materials, and stored in master files uncompressed (TIFF format). Quality control of images requires an ongoing sampling, with good metadata standards and practices, and good data management and back-up procedures. The back-up images and metadata should be in the existing system and in master files stored on-site as well as off-site. Good descriptive metadata includes MODS, VRA and MARC. For good sustainability, funding and staffing are required over the duration of the project, even if it is "one time." Selection criteria should be based on the needs of the audience, and the media used for preservation should be inspected and refreshed regularly.

In any project, digital preservation is a long-term process. Here, too, the images should be of high quality with resolution and bit depth appropriate to the materials, and master files (TIFF format) stored uncompressed. Quality control of images is an ongoing effort and requires planning, funding and staffing. The images and the metadata should have high standards of practice and be implemented as needed in the development of the preservation system. Decisions need to be made on the selection of metadata, e.g. Dublin Core, MODS, VRA and MARC are descriptive metadata, METS is a structural metadata, NISO z39.87 is a technical metadata, and PREMIS is a preservation metadata. In preservation, back-up procedures, images and metadata should be in the existing system for access, but stored in master files on-site and off-site. More extensive back-ups may be necessary as the preservation system grows and as a response to disaster preparedness. Some of the best practices for metadata are the use of standards to document scanned images, to back them up carefully, and to monitor and implement all aspects.

About the authors - Contributors to this issue were:

Elli Bambakidis

Archivist, Dayton Metro Library and OPC Secretary 2004-05

Sue Dunlap

Preservation Manager, College of Wooster Libraries and OPC Education and Outreach Chair 2004-05

George W. S. Hays

Director/Clerk-Treasurer, Salem Public Library and OPC Chair 2004-05

Irene Martin

Local History and Genealogy Department, Toledo-Lucas County Public Library

Cynthia McLaughlin

Columbus Jewish Historical Society