NEWSLETTER HIGHLIGHTS:

FROM THE CHAIR

Dear Colleagues:

The Boston College (BC) Libraries, where I serve as *Preservation Manager*, are actively pursuing digital projects even as a formal, system-wide digital preservation strategy remains in development. Is this wise? I suppose one could charge us with acting prematurely, but I view our circumstances – widely mirrored in cultural institutions across the country – as a practical reality. Archivists, librarians, and curators increasingly work in digital environments and, sooner or later, must engage with bits and bytes. The challenge, of course, is to do so responsibly, i.e., programmatically; and programs take time to develop and fully mature. In the meantime, BC is doing its best to follow sound, established practices regarding application of digital technologies.

Much has been written and taught about traditional preservation planning. Indeed, collegiality and the free-sharing of approaches to that noble task helped shape the core activities

comprising today's successful preservation programs. A similar dynamic has evolved now regarding development of digital preservation programs. All over again, as we share our stories – successes and failures – we collectively draw closer to discerning best practices.

In such a spirit of collegial sharing, both this and the next issues of *Infinity* will spotlight digital preservation activities in the field via a series of brief case studies. We are most grateful to: the National Archives and Records Administration (NARA); the National Library of Medicine (NLM); the Smithsonian Institution Archives, Rockefeller Archives Center, and the Collaborative Electronic Records Project (CERP); the Massachusetts Board of Library Commissioners (MBLC); and the Post Street Archives for their valuable contributions in this issue (see below).

Your institution is also invited to participate by contributing a brief case study. Submission guidelines are available from *Infinity* Editor Tonia Sutherland, who will be responsible for ensuring balanced subject coverage. If you are considering making a contribution, please notify Tonia of your intentions at tsutherland@library.umass.edu. And please be aware that, if the topic you are targeting already appears to be sufficiently addressed, Tonia may suggest an alternative choice.

Please note that "Digital Preservation" will also be the main topic of discussion at this year's meeting of the *Preservation Section*, which will be held on Friday, August 31st, from 12:00 to 2:00 p.m. A panel of experts will be on hand to make short topical presentations and to respond to your digital preservation questions. It should be fascinating!

I look forward to seeing everyone in Chicago.

Steve Dalton
Chair, Preservation Section

FROM THE EDITOR

Dear Preservation Colleagues:

Each issue of *Infinty* offers new areas of exploration and brings to light new challenges in preserving our cultural and historical resources. I'd like to take a moment to encourage you to consider contributing a case study, a new initiative, examples of a collaborative effort, or even a program announcement for the Summer/Fall issue.

My colleagues at the University of Massachusetts Amherst will soon be joining forces with Massachusetts's other land-grant university, the Massachusetts Institute of Technology (MIT), to digitize a significant portion of materials relating to the history of agriculture in Massachusetts. Projects such as these are noteworthy not only because they represent large scale digital preservation efforts, but also because they inspire a collaborative spirit.

As we work toward establishing standards and best practices in digital preservation, we evolve and flourish together both as a profession and as a community. Many of the articles in this issue are the product of exactly this type of innovative collaboration. For this I applaud you, and I look forward to continuing to learn from all of you.

Think Spring!

Tonia Sutherland
Editor, *Infinity*

NARA Transitions to a Digital Workflow

While best known for the Charters of Freedom housed on the Mall in Washington, DC, National Archives and Records Administration (NARA) also includes a nationwide network of record centers, regional archives, and eleven Presidential Libraries. In combination, these facilities serve to preserve and provide access to the records that document our nation's history.

At the NARA, the transition from a traditional, analog reformatting program to one consisting of an entirely digital workflow is a complex, ongoing process. NARA's requirements to digitally reformat its holdings include levels of security, capacity, availability, accessibility, sustainability and authority not often encountered in the cultural heritage community. Experts from across the organization are involved in this transition to ensure that the resulting program properly reflects NARA's archival and preservation principles

The replacement of still and microform cameras, motion picture duplication equipment, and analog audio equipment with digital cameras, secure storage, and digital asset management tools demand IT input. We are working to integrate the high-tech aspects of the project, with the application of preservation techniques and considerations to ensure that the needs of the original records are met. The benefits of scanning are obvious; there are also risks. Mishandled records can be damaged during the reformatting process. Improper alignment of originals, poor focus or poor exposure will result in images that are useless in either the analog or digital world. Lost, misidentified, or poorly described digital files are undiscoverable. NARA intends to minimize these and other problems through the involvement of preservation staff in every aspect of this program and by infusing it with the core principles that have helped to define reformatting as a preservation tool.

To learn more about digital reformatting at NARA and other institutions, please plan to attend NARA's 21st Annual Preservation Conference: **Managing the Intangible:** *Creating, Storing and Retrieving Digital Surrogates of Historical Materials* April 30-May 1 to be held at the UMUC Inn and Conference Center in Adelphi, Maryland. For program and registration information see: http://www.archives.gov/preservation/conferences/2007/.

- Article courtesy of Preservation Programs – NARA

Profiles in Science
NATIONAL LIBRARY OF MEDICINE

Quality Control: Ensuring Standards and Best Practices for Profiles in Science®
By Meghan Attalla, History Associates Incorporated, Rockville, MD

Project Background

Launched nearly ten years ago as a digital library research and development project, the National Library of Medicine's Profiles in Science (http://profiles.nlm.nih.gov/) promotes the use of the Web for researching and teaching the history of biomedical science and other related fields in health and medicine. The Digital Manuscripts Program (DMP) digitizes archival materials donated to the Library or made available by collaborating institutions. Selections from the personal papers of such notable figures as Oswald T. Avery, Francis Crick, Rosalind Franklin, C. Everett Koop, Joshua Lederberg, Marshall W. Nirenberg, and Florence R. Sabin are accompanied by in-depth historical narratives. As of this writing, the site has grown to include twenty-four collections and over 28,000 active records.

With a continuously growing and evolving site, DMP staff places great importance on its quality control tools and procedures. Careful quality review ensures consistency and accuracy of data and image capture, helping to guarantee the project's authority, sustainability, and adaptability. Consistency of data allows for automated changes across individual Profiles exhibits, decreasing the need for laborious manual edits to individual records. Because technologies inevitably change, thorough quality control will make migration to new systems easier, and therefore better assure the content's survival.

Quality control also facilitates effective searching and browsing of the collections, enhances site usability, and provides a framework for content review and staff training. To that end, DMP created written procedures for its various stages of review, described in more detail below.

Metadata Entry System as Quality Control Tool

The program's Metadata Entry System and Diagnostic Server provide the tools for thorough quality review, both automated and human. Developed and maintained by the Lister Hill National Center for Biomedical Communications (LHNCBC) Digital Library Research and Development team, the Metadata Entry System, based on the Dublin Core Metadata Element Set, drives the entire digital conversion process and workflow (i.e., metadata, scanning, quality control, permissions, and search and retrieval). Tied to this system, the Profiles in Science Diagnostic Server, an online staging area, allows staff to view administrative, technical, and descriptive metadata and images; generate a variety of reports; and preview the online collections prior to their public release.

The Metadata Entry System enforces rules, for example, by requiring certain fields and by generating warning messages for improperly formatted or missing data. It provides drop-down menus for particular fields, thus establishing authority control and minimizing the errors and inconsistencies inherent in the use of free-text fields. The system does not allow for accidental release of metadata and images before certain key steps are taken, such as the completion of permissions work. The database is used to track whether records and images have undergone quality control, and whether records are ready to be or have been approved for public release. The Diagnostic Server's reports and site previews aid staff in finding and correcting errors and inconsistencies both within and across collections.

Metadata Quality Control

All records for Profiles in Science undergo metadata quality control shortly after their creation. The process allows the digital archivists to discover and discuss data needs, enabling them to draw upon one another's experience and expertise to resolve any issues or inconsistencies. Quality control is performed by a different digital archivist than the one who created the record.

Records comprise a wide range of document types (correspondence, photographs, laboratory notes, articles, etc.) each with their own quality review needs. The reviewer compares the original item to its record in the Metadata Entry System, examining the fields within each of the database's fourteen Dublin Core-based tabs. Overall, he or she checks for spelling and typographical errors; for completeness, accuracy and consistency of the data; and for adherence to rules established in the program's "Metadata Definitions and Input Guidelines."

Image Quality Control

After scans are created, the digital archivist uses Adobe Photoshop to examine both the master preservation file (TIFF) and its derivatives for Web display (JPEGs or PDF). A scan sheet, automatically generated after metadata entry, accompanying each original document contains a unique identifier and barcode to match the original item to its corresponding record and minimize filename errors.

The digital archivist tells the Metadata Entry System to look for scanned files in a set directory; if the system detects anything "wrong" with the files, such as a missing derivative, it alerts the user to a potential quality issue. Comparing the original item to the digitized files, the reviewer checks for clarity and completeness and ensures that all project scanning specifications are met.

Final Reviews and Site Release

Additional layers of review must occur before a site is built and released to the public. Following the exhibit release checklist, the digital archivist conducts an overall review of the site, examines standard reports from the Diagnostic Server, and codes the exhibit text and chronology. The digital projects team coordinator then completes an in-depth final review, approving metadata/images and all other aspects of the web exhibit for public release. The DMP program manager subsequently examines the final product for overall quality and completeness.

The Digital Library Research and Development team generates additional reports to ensure quality and prepares the site for Web display. Only after several levels of formalized quality control does a new Profiles in Science exhibit make its public debut.

Smithsonian Institution Archives & Rockefeller Archive Center Tackle Digital Preservation Issues

The Collaborative Electronic Records Project is a three-year endeavor that began in August 2005. This project seeks to develop, test, and share technology to preserve digital documents. This initiative has a strong focus on e-mail messages and attachments. Lynda Schmitz Fuhrig at the Smithsonian Institution Archives (SIA) and Nancy Adgent at the Rockefeller Archive Center (RAC) serve as the project's archivists.

CERP was developed by Dr. Edith Hedlin, then director of the Smithsonian Institution Archives and former SAA president, and by Dr. Darwin H. Stapleton, executive director of the Rockefeller Archive Center. These well-established, stable institutions have responsibilities for the management of records that go well beyond cataloging of historical manuscripts. Both are entrusted with the strategic management of a broad range of institutional records from founding documents to nearly current files.

The project archivists conducted interviews with depositors to determine how electronic records and email fit into their daily business routines. Not surprisingly, email management varies greatly from users with more 10,000 emails in the Inbox with no organization to users who have elaborate systems of folder hierarchies.

Specific files and email account holders were chosen for testing. To date, email messages with attachments, education course planning files, an office handbook, and other electronic documents have been transferred in digital form. Transfers were conducted via server and removable media, and initial processing has begun.

The email presents interesting challenges: missing email addresses and transmission data, dead Web links, large bulk of messages, missing sent email, and specific applications needed for viewing the email.

Email guidance also has been drafted for the depositing organizations. Requirements for a digital repository for the materials are currently being refined. The results of this undertaking will be shared with the depositors and the archival community through a symposium.

The CERP team has heard many times from other archivists, librarians, and technologists that email preservation is a tough issue to tackle. Working together, these two archives expect to achieve much more than they could accomplish separately and to develop a model that will benefit a broad range of nonprofit and philanthropic institutions.

The Smithsonian Institution Archives: http://siarchives.si.edu

The Use of L.S.T.A. Funds for Digital Projects in Massachusetts

In the federal fiscal year 1995, the Massachusetts Board of Library Commissioners (MBLC) began making some of its Library Service and Technology Act (LSTA) funds available to provide grants for preservation surveys to public libraries. In subsequent years grants for the Preservation of Library and Archival Materials, Manuscript Arrangement and Description, and Digitizing Historical Resources have been added to the categories of grants funded under L.S.T.A. in Massachusetts.

The Digitizing Historical Resources grant was added in FY2000, and to date nineteen grants of up to $40,000 have been awarded. From the beginning, it has been strongly emphasized that these grants were for access purposes only and not for preservation. Applicants have had to have completed a preservation survey and undertaken any conservation work necessary to stabilize the materials to be digitized prior to the grant's commencement. In addition, the grants are targeted at 'collections' not individual items from a number of sources to create an online 'collection.' That being said, when an applicant has come in with a proposal to digitize a collection and the review panel has known of other items or collections of materials that would complement this collection, the applicant has been urged to make this a cooperative grant proposal. A prime example of this was an application to digitize the Alexander Parris Papers housed at the Massachusetts State Library. Additional Parris materials were known to be housed at the Massachusetts Historical Society; Historic New England; the Boston Athenaeum; the Boston Public Library; the Boston National Historical Park, Charlestown Navy Yard; and the Massachusetts General Hospital. As a result, the review panel strongly recommended that the applicant contact these other institutions and work to incorporate their holdings into a more complete online collection. Although this meant that the project entailed much more work, all the parties were pleased in the end, and a much more comprehensive online collection was created.

The L.S.T.A. grants range from $5,000 to $40,000 with a 4:1 match (L.S.T.A. to local). The concept has always been that while we want to encourage applicants, we also want them to have a stake in the project. Consequently, we have required the one fifth cash match for these grants. In addition, since we do not want to be purchasing all the equipment needed for such a project, we require the applicants to go out to bid for their project. To assist them in this matter, we provide them with a list of vendors who have successfully completed projects in the past. Additionally, we have required that the digitized collections be mounted on the Web and also be accessible through the Massachusetts Special Collections Directory. There has been one exception to the requirement that the projects be outsourced. When an applicant has a fully-functional scanning lab that has been used for projects of this type, they have been allowed to

apply for a grant to support non-permanent staff to do the digitization of the institution's designated collections.

One further requirement is that the applicants commit to maintaining the digital files into the future so that they will continue to be available to researchers. While this is not digital preservation *per se* as per the focus of this newsletter, at least the recipients know that it is up to them to maintain the files as long as it is possible. Since these are specific, special collections that have been the focus of their work, each institution has a stake in maintaining them online and is not apt to let the files become inaccessible.

In addition, during the past three years there has been a project underway (with L.S.T.A. funds) to create *Digital Commonwealth*, a central portal for searching and accessing digital collections in Massachusetts. As this project has progressed, there has been a move towards standardizing the metadata requirements for these collections so that they are Open Archives Initiative (OAI) compliant, searchable, and harvestable. Part of the evolution in the grant requirements for the Digitizing Historical Resources grants has been to move towards compliance with the *Digital Commonwealth* requirements so that these collections will be made more available to researchers.

Significant strides have been made in digitizing materials in Massachusetts. By requiring that 'collections' be the focus of the grant applications, we have ensured that the applicants have thought seriously about which collections they wish to digitize and whether or not they are arranged and described. This has also meant that the projects have not spent time and money digitizing materials that, while part of the collection, really do not add much to the digitized result. The end result has been collections that have been compact and concise and had a real focus. It has also made it possible for some smaller institutions housing important collections to be able to make them available on the Web. This in itself is an important step toward accessibility.

Gregor Trinkaus-Randall
Preservation Specialist
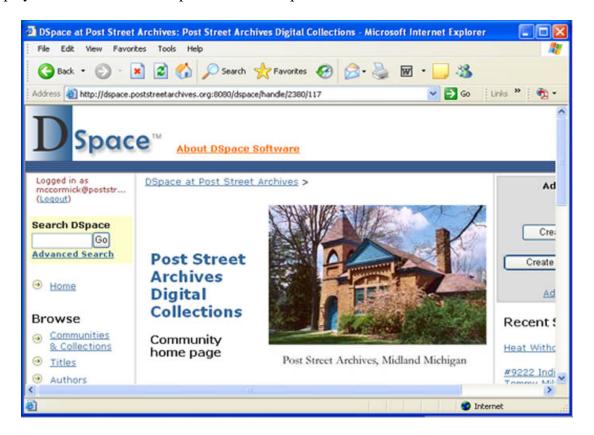Massachusetts Board of Library Commissioners

---

Digital Preservation - A Collaborative Model using DSpace

The Post Street Archives is a small privately held archives in Midland, Michigan whose mission is to preserve historical records of the Dow Chemical Company and those of its founder, Herbert H. Dow and his family. In spite of the challenges of limited staffing and resources, by 2003 it was clear that in order to continue to collect and preserve born-digital Dow Chemical Company historical information, it was time to enter the digital storage and preservation world.

The director of the Post Street Archives, Tawny Ryan Nelb, learned about DSpace, an open source software product, as a possible solution, but without in-house information technology (IT) support, she turned to MITCON, an agency that provides IT support to non-profit agencies in

Midland County for help. MITCON's director, Ed Haycock, identified the resource requirements and funding from the Herbert H. and Grace A. Dow Foundation was secured... In early 2006, the Post Street Archives hired a Digital Archives Coordinator, while on the IT side MITCON established an on-going relationship with Michigan Technological University (MTU) to provide additional IT support not available within MITCON. The required hardware was also obtained (also with Foundation funding).

Early on the decision was made to use DSpace 'out of the box' in order to minimize IT support both in development and in migration to future versions. DSpace was initially designed as an institutional repository and is most heavily used for theses and other publications. For 'out of the box' users it is important to evaluate whether the existing metadata fields, search criteria, and display elements work for the particular use in question.

The project began in March 2006 when the MTU team installed DSpace on a test server. During this effort the team produced a 'how to install DSpace' document geared to the particular hardware/software combination used that will allow MITCON to install additional instances of DSpace for other users (see http://www.poststreetarchives.org/DSpaceInstall.pdf)   On the user side, there was limited 'how to use DSpace' documentation which meant that a lot of experimentation was needed to learn the ins and outs of DSpace. It is not an easy solution. By the end of April, a determination was made that the product was stable, operated reasonably well and most importantly had enough flexibility to allow DSpace to be used in an archives application.

A production environment was established in June, 2006. Challenges encountered included fitting an archival collection structure into the DSpace hierarchy of communities and sub-communities, understanding the details and limitations of the DSpace authorization structure, and learning the basics about different kinds of digital records and how they can be stored in DSpace.

Future challenges include moving from a storage and access emphasis to a preservation emphasis. As technology changes, file structures will need to be migrated to newer versions or otherwise preserved. While DSpace doesn't inherently offer solutions for that challenge, it does provide a platform from which to launch those efforts and a technical user community that will jointly be addressing that challenge. IT support for those preservation efforts will also need to be addressed.

In addition to born-digital records, the Post Street Archives will also take advantage of the DSpace platform for use with digital versions of other collection material. Combined with a planned web site, this will provide the research community a wealth of information online as well as onsite.

For more information contact

Kathleen C. McCormick
Digital Archives Coordinator
Post Street Archives
989-832-0870
mccormick@poststreetarchives.org

NEDCC Conference: School for Scanning: The A-Z of Creating Digital Collections

Location:
Minneapolis Marriott City Center, Minneapolis, MN

Date(s):
May 1-3, 2007

Short Description:
This popular 3-day conference takes digitization from theory into practice and is geared toward participants with a beginning or intermediate level of digital knowledge. Participants who already have experience in digitization can obtain an up-to-date briefing. This program is taught by some of the nation's most highly respected digital experts.

For complete conference information and to register online, go to:
www.nedcc.org

SAA ANNUAL MEETING 2007 PREVIEW

ARCHIVES/CHICAGO 2007: The 71st Annual Meeting of the Society of American Archivists will be held in the "Windy City" beginning August 29th and running through September 2nd, 2007. The conference will be held at The Fairmont Chicago and hotel reservations are $129/single, $149/double in advance. For more information, contact The Fairmont Chicago, 200 North Columbus Drive, Chicago, IL 60601. Please call the hotel directly at 312-565-8000 or 800-526-2008 when making your reservation, and please be sure to indicate that you are part of the SAA group. You may also make your reservation online by visiting the SAA Website at www.archivists.org and clicking on the conference graphic.

As noted above, the *Preservation Section* **Meeting** will be held on Friday, August 31st, from Noon to 2:00 p.m. "Digital Preservation" will be this year's theme; and, a panel of experts, including former SAA President Peter Hirtle (Cornell University), Anne Gilliland (UCLA), Nancy McGovern (University of Michigan), and Elisabeth Long (University of Chicago) will be on hand to discuss such issues as: copyright, metadata, certification of trusted digital repositories, and digital preservation program development. All are welcome!

Preservation Section Committee (http://www.archivists.org/saagroups/preserv/index.html)

Steve Dalton (Chair)
daltonst@bc.edu
Brenda Gunn (Chair-Elect)
bgunn@mail.utexas.edu
Gregor Trinkaus-Randall (Past Chair & Chair, Nominating Committee)
gregor.trinkaus-randall@state.ma.us
Rebecca Hatcher (Member-at-Large & Co-Chair, Education Committee)
rebecca.hatcher@yale.edu
Stephen Van Buren (Member-at-Large)
stephen.vanburen@sdstate.edu
Susan Koutsky (Co-Chair, Education Committee)
skoutsky@umd.edu
Julie Graham (Web Liaison & Nominating Committee)
jgraham@library.ucla.edu
Patricia Morris (Chair, Program Committee)
Patricia.Morris@colorado.edu
Tonia Sutherland (*Infinity* Editor)
tsutherland@library.umass.edu
Laura Finger (Co-Chair, Outreach Committee)
lfinger@rosenberg-library.org
Anne Ostendarp (Co-Chair, Outreach Committee)
aostendarp@amherst.edu