Surviving Change: The First Step toward Sustaining Your Digital Library*

Marie E. Gallagher and Christie Moffatt, National Library of Medicine, USA

http://profiles.nlm.nih.gov/

Abstract

Profiles in Science®, a Web-accessible digital library showcasing the personal manuscript collections of several Nobel Prize winning scientists and other prominent biomedical researchers and leaders in public health, has been publicly available since September 1998 through the National Library of Medicine® (NLM) at the National Institutes of Health, U.S. Department of Health and Human Services. Profiles in Science features digital reproductions of historical items selected from the archival collections of the NLM as well as collaborating institutions. In addition to the expected challenges posed by technological evolution, our experience with Profiles in Science over the years has taught us that change affects virtually every area of our project, including established procedures, project mission, legal considerations, and stakeholder expectations. Sustaining a digital library requires surviving all types of change. This paper will include a behind-the-scenes look at situations we have faced and will continue to encounter, as well as a "survival guide" for coping with change. By sharing our experiences and strategies, we hope to provide those planning new digital projects a glimpse at what they may face over time. Institutions with established projects may discover challenges and strategies they have not previously encountered, and may inspire improvements to these coping methods.

Keywords: surviving change, adaptability, sustainability, organizational strategies, digitized manuscripts, National Library of Medicine

Introduction

Profiles in Science®, an information resource of the National Library of Medicine® (NLM) on-line since 1998, provides Web users access to digitized manuscripts from the collections of prominent biomedical researchers and leaders in public health. Selections from the personal papers of Oswald T. Avery, Joshua Lederberg, Marshall Nirenberg, C. Everett Koop, Florence R. Sabin, Linus Pauling, Barbara McClintock, and Francis Crick, among others, are available on Profiles in Science. The collections contain various types of materials including published and unpublished manuscripts, diaries, laboratory notebooks, correspondence, photographs, speeches, oral history interviews, poems, drawings, and audiovisuals. Nineteen collections are currently featured on the Web site. Digital collections are updated with new materials periodically, and new collections are added throughout each year. Most of the physical papers have been donated to the National Library of Medicine, where they are processed, housed, and used by researchers who visit the Library. Some Profiles in Science collections were created in collaboration with other institutions that hold collections of interest. The digital collections may be viewed at *http://profiles.nlm.nih.gov/*.

* In: Gallagher, M. and C. Moffatt, Surviving Change: The First Step toward Sustaining Your Digital Library, in J. Trant and D. Bearman (eds.). *Museums and the Web 2006: Proceedings,* Toronto: Archives & Museum Informatics, published March 1, 2006 at http://www.archimuse.com/mw2006/papers/gallagher/gallagher.html

Fig. 1: Home page of Profiles in Science Web site. *http://profiles.nlm.nih.gov/*

Profiles in Science, nicknamed "Profiles," continues to present interesting challenges and provide learning opportunities about making special collections available over the Web. Some of these challenges have been predictable, such as changes in technology. Others challenges have been unexpected, requiring project staff to adapt and find creative solutions. Our responses have been guided by previous experience and general principles and best practices for digital library development (McCray & Gallagher, 2001.)

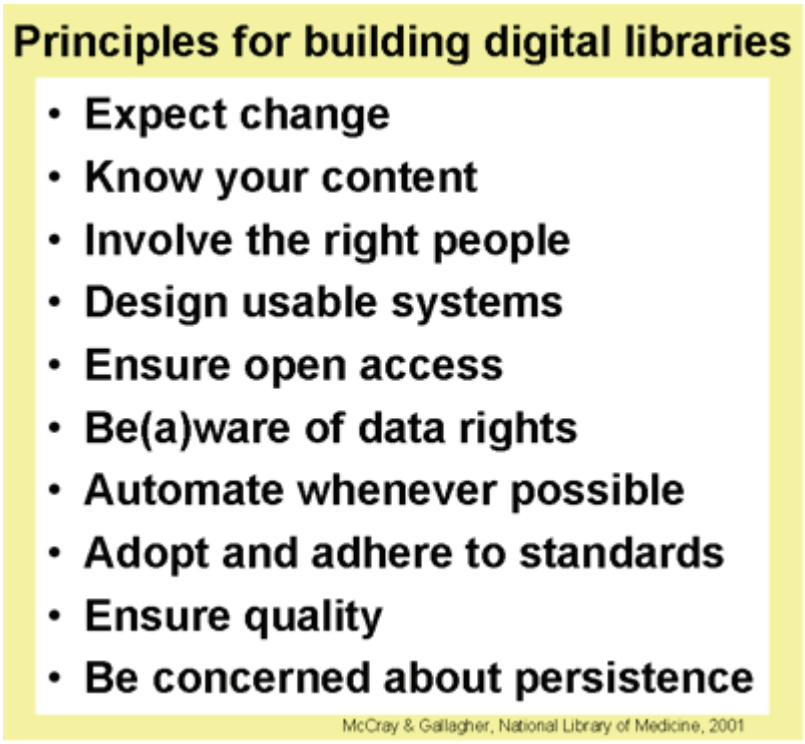

Fig. 2: Principles for building digital libraries.

The team behind Profiles in Science is drawn from the Digital Manuscripts Program in the NLM History of Medicine Division (HMD) and from the Digital Library Research Program in the NLM Lister Hill National Center for Biomedical Communications. The Digital Manuscripts Program team includes archivists and historians who acquire, process, research, describe, digitize, and preserve the physical collections. Computer scientists and software developers make up the Digital Library Research team, which develops software programs and databases, investigates and puts to use hardware and software technology, and protects the digital collections. Below we report on some of the challenges and changes observed over the years, and how we have responded.

Everything Changes

Technology

Arguably the most obvious challenge to sustaining a digital library, for local use or on the Web, is the threat of technological obsolescence due to rapidly changing technology. All of the hardware and software upon which any digital library depends are susceptible. In the best case, failing to upgrade hardware or software eventually leads to slow system responses, leaving users dissatisfied. Not keeping hardware and software current usually has more serious consequences. The risk of incompatibilities increases with the age of hardware and software. You will eventually find yourself unable to add new components such as scanners or external storage, or unable to install a new operating system or applications software on old hardware. Even when operating systems and applications software are kept current, the risk of software incompatibilities is ever present. Incompatibilities can result in limited, corrupt, or completely non-functional computer systems. Newer hardware and software provides faster performance,

increased storage capacity, and opportunities to use improved technologies such as, DVDs, USB ports, IEEE 1394 connections, and wireless networking. Improvements to operating systems and applications software also include improved performance and new capabilities. Newer hardware and software often include fixes to flaws in previous generation technology. The most serious risk of failing to upgrade hardware is complete hardware failure: hardware components wear out, and it may be impossible to find replacement parts. The most serious risk of failing to upgrade software may result in dangerous problems such as security holes that provide unauthorized access, corruption and destruction of a digital library. The prevalence of computer hacking and the potential costs associated with data loss makes patching security holes critical. However, applying these patches is not without risk. Newly patched software can conflict with your existing software or hardware. The resolution of these conflicts may not be obvious or even possible, and may cause your digital library to become unavailable.

In our experience, upgrading hardware has been necessary every three or four years. Around this time we start encountering slow responses, have trouble adding new hardware components, or become unable to run some software or queries. This time period has also approximately coincided with major releases of new Solaris and Windows operating systems that run on our servers and desktop PCs (Sun Microsystems, Inc., 2005; Microsoft Corporation, 2003.) We have experienced hardware failures before the close of the hardware replacement cycle, and fortunately we were always able to restore our data from backups to similar or identical computers quickly. To minimize the impact of hardware failures, our next system design includes redundant hardware and software at two locations.

The need for software upgrades has been constant. Changes in our project's evolving policies and procedures mean modifications must be made to all computer code related to an affected policy or procedure. We have used a combination of open source software, commercial off-the-shelf software, and software developed in-house for Profiles in Science. For all three types of software, we have observed that one of two things eventually happens: 1) the current version of the software is replaced by a new version that we are forced to use or 2) we must identify or develop replacement software because the current software either stops working or is discontinued. Continuous monitoring, testing and maintenance are needed to keep computer systems stable because commercial and open source software are constantly subject to upgrades or discontinuation. Obviously there are costs associated with upgrading and maintaining a digital library's hardware and software. Consider the cost of NOT maintaining or upgrading: this could result in permanent loss of access or loss of the information and software comprising the digital library. Foregoing hardware and software upgrades is not an option if a digital library is to be available for the long-term. Without hardware and software upgrades, it is not a matter of if, but when, the digital library will stop functioning.

Standards and digital file formats

Adherence to standards has been an important activity for Profiles in Science to ensure compatibility and accessibility. During the years that Profiles has been on-line, many standards used by the project have changed. For example, the World Wide Web Consortium (W3C) HTML Specification (Raggett, Le Hors, & Jacobs, 1999) is currently version 4.01; the Dublin Core Metadata Element Set (Dublin Core Metadata Initiative, 2004) is currently version 1.1; the Adobe Portable Document Format Reference (Adobe Systems Incorporated, 2004) is currently version 1.6. Each time one of these three standards was updated, potentially thousands of the

pages on the Profiles Web site needed to be updated to remain conformant. It would have been overwhelming to manually make these changes to individual pages. Fortunately, all of the Profiles Web pages are derived from data that resides in a database. We made changes to the computer programs that generate the pages, and regenerated the site according to the new standard. We have not converted all of our PDF files to version 1.6 since there is no known problem reading our current PDFs, and it is possible that the software we use to extract PDF data and search PDFs may not be able to handle PDF version 1.6 yet. We are preparing to do this conversion when necessary. Monitoring and adhering to evolving standards will remain an ongoing effort.

File formats continue to be born and evolve. For years there have been many formats (Murray & vanRyper, 1996) to choose from, each having strengths and weaknesses. We chose TIFF as the master format for our digitized images, and derived PDF and JPEG files for the Web. TIFF is a published standard (Adobe Systems Incorporated, n.d.) and remains a widely used format; there is no shortage of software that converts TIFF files to alternative formats. There are numerous variations of TIFF files, so we always pay attention to the options presented when creating TIFF files. To date, the formats we chose remain usable, but we expect to make other derivative file formats available as alternatives emerge.

The media onto which we backup and store digital files is also evolving, and new types of media become available each year. Media formats become obsolete when the hardware or software needed to read them are no longer produced. To combat media obsolescence, we move our digital files from one format to another approximately every three to four years: about the same as our hardware replacement cycle. At the start of the project, disk capacity was smaller, disk storage was more expensive, and options for copying files off-line were less attractive. High cost, concerns about computer hacking, infrequent need to access the master TIFF files, and a hard disk crash convinced us to keep multiple copies of the TIFF files off-line at different locations. The idea was the same as that underlying the Lots of Copies Keeps Stuff Safe (LOCKSS) system (Stanford University Libraries, 2006.) We also experienced first hand that storage media degrades. When we copied files from generation 1 to generation 2 media, we verified that the copies were identical. A few years later, when we copied from generation 2 to generation 3 media, we encountered read errors on some of the generation 2 media. Fortunately we had made three versions of the generation 2 media, so we were able to recover the lost generation 2 files from one of the intact generation 2 copies. We always verified a copy of one version against another version that was supposed to be identical. Having three identical versions was essential to recovering from the media corruption problem.

Perception of acceptable quality

As the pixel resolution of computer monitors increases, digitized images will appear to become smaller and less detailed when viewed on better display screens. This is already evident in digital camera technology. Compare the quality of a photo taken with a five-year-old $400 digital still camera to a photo taken with a current $400 digital still camera. The photo that seemed amazing five years ago now looks disappointing because we have become accustomed to the higher resolution and greater number of colors captured by the newer camera. Similarly, today's digitized video will appear smaller as display screen resolution improves over time. Eventually, watching older digitized video will become frustrating when compared to watching newer video digitized with improved methods and quality.

Current technology, guidelines and recommendations for digitization may produce results that are superb today, but will be judged insufficient in time. The best we can do is capture at the highest possible quality now, hope the result will be sufficient for many years to come, and be prepared to re-digitize in the future. How soon we will have to re-digitize will depend on how well we captured using today's technology and best practices, and how fast capture, display and print technologies advance.

Contents of the collections

The collections donated to the Archives and Modern Manuscripts Program (AMMP) in HMD are mostly non-digital, in the form of papers and photographs. Occasionally collections include audio and video tapes and films. Per the normal preservation process, these films and tapes are triaged, inspected, and stored in an on-site cold vault, separate from the paper collections. Some are copied to Betacam SP format and transferred to the Historical Audiovisuals Program. More recently donated collections of papers include text files or modern word processor files on floppy disks and CDs. So far, the electronic files have either been accompanied by hard copies, or we print the files and add the hard copies to the collection. Sometimes donors' colleagues, friends and family members offer digital photos. One donor wishes to donate historical e-mail and electronic files which are on magnetic tape from the 1960s, but locating the hardware and software needed to read the tapes has been problematic. Another collection contains a computer animation that was derived from an unknown proprietary system.

The living donors of Profiles collections continue to create materials in digital formats. One donor especially in tune with the problem of digital preservation makes every effort to keep his e-mail and computer files in plain text format, but we expect this case is more the exception than the rule. Collections donated in the near future might contain DVDs, hard disks, and maybe even computer systems. We expect to receive spreadsheets, encrypted files, and other complex or proprietary file formats. We look forward to the day when digital preservation systems are available that can handle such complicated formats. As a stop gap measure, we may copy files that are in formats that seem hard to preserve, without loss whenever possible, to more sustainable formats (Florida Center for Library Automation, 2003; Arms & Fleischhauer, 2005) to maintain access to the donated data.

Space needs

The need for space, digital as well as physical, has increased significantly since starting the project. Three renovations have been undertaken to satisfy the need for space for the project, and we always seem to be in need of more. Additional space is needed to accommodate offices for the increased staff dedicated to Profiles, which has been addressed by extensive remodeling, including removal of existing walls, and installation of cubicles. In one case, stack space was rearranged to make room for cubicles. Additional space is also needed to process and store the physical collections, some of which are hundreds of boxes in size.

The computer servers housing Profiles require space, mostly in the NLM main computer room. As space became more of an issue in the computer room, space-saving rack mountable machines were encouraged then eventually required. Disk storage is needed for the Profiles on-line images and metadata, as well as for the off-line archival master images and backups. Welcome advances in technology have resulted in larger capacity, less expensive disks and better options for storing digital data on- and off-line.

Staff expertise

The expertise of the staff needed for Profiles in Science was different from that of staff hired previously. In addition to using their library science and archives expertise, staff adjusted to the special needs of the new project by learning about digitization and describing digital resources, as well as collections, on the job. Those who enjoy using technology and becoming the experts in a new area are the quickest to adjust. The project's computer scientists and software developers also needed to acquire new expertise for the project. They gained understanding of the standard archival procedures and terminology previously unfamiliar to them. All project staff have developed expertise and remained aware of evolving standards such as Dublin Core and Encoded Archival Description (EAD), vocabularies such as the NLM Medical Subject Headings Vocabulary File® (MeSH) and Getty's Art & Architecture Thesaurus© (AAT), best practices and guidelines, and issues in digital preservation. Involving staff with varying backgrounds has been essential for the project. Archivists need not spend precious time trying to become expert computer scientists and vice versa; each expert uses his or her greatest strengths and consults others about issues outside his or her area of expertise. Software developers focus their efforts on computer science related problems rather than struggling with archival and library science issues.

The newly acquired skills and expertise made staff on the project extremely attractive candidates for positions elsewhere, so there seemed to be frequent training of replacement staff for metadata entry and digitization. This made written procedures for metadata entry and digitization even more critical than anticipated. The procedures improved continuity with new staff and prevented the total loss of expertise of departing staff.

Another new area of expertise needed was copyright research, for identifying copyright holders and obtaining permissions to digitize the items for posting on Profiles. At the start of the project, copyright research was a part-time activity shared among archival staff. Because copyright research required even more time than digitization or description, the number of items that needed copyright research quickly grew into a bottleneck that prevented items from being included in Profiles. Copyright research is now one staff member's dedicated full-time activity.

Legal considerations

Because the copyright laws are complex and subject to debate, interpretation, and revision, copyright research became a much larger effort on our project than originally anticipated. Clearing copyright is sometimes the longest step in the process of adding an item to Profiles. Each item selected for inclusion on Profiles must be evaluated to determine its copyright status as outlined in Figure 3. Depending on the responsiveness of the copyright holder, this process can require days, weeks or months. In cases where the copyright holder either cannot be identified or cannot be contacted, we may post the item on Profiles along with a disclaimer asking that anyone with information about the copyright holder please contact us. For each item, the entire process is documented and all records of contacts are retained. The procedures for obtaining copyright permission and determining the copyright status of an item have evolved during the project. Because preventing the accidental release of protected items is an important function of our metadata entry software, the software developers make ongoing modifications to the system to reflect current procedures and practices.

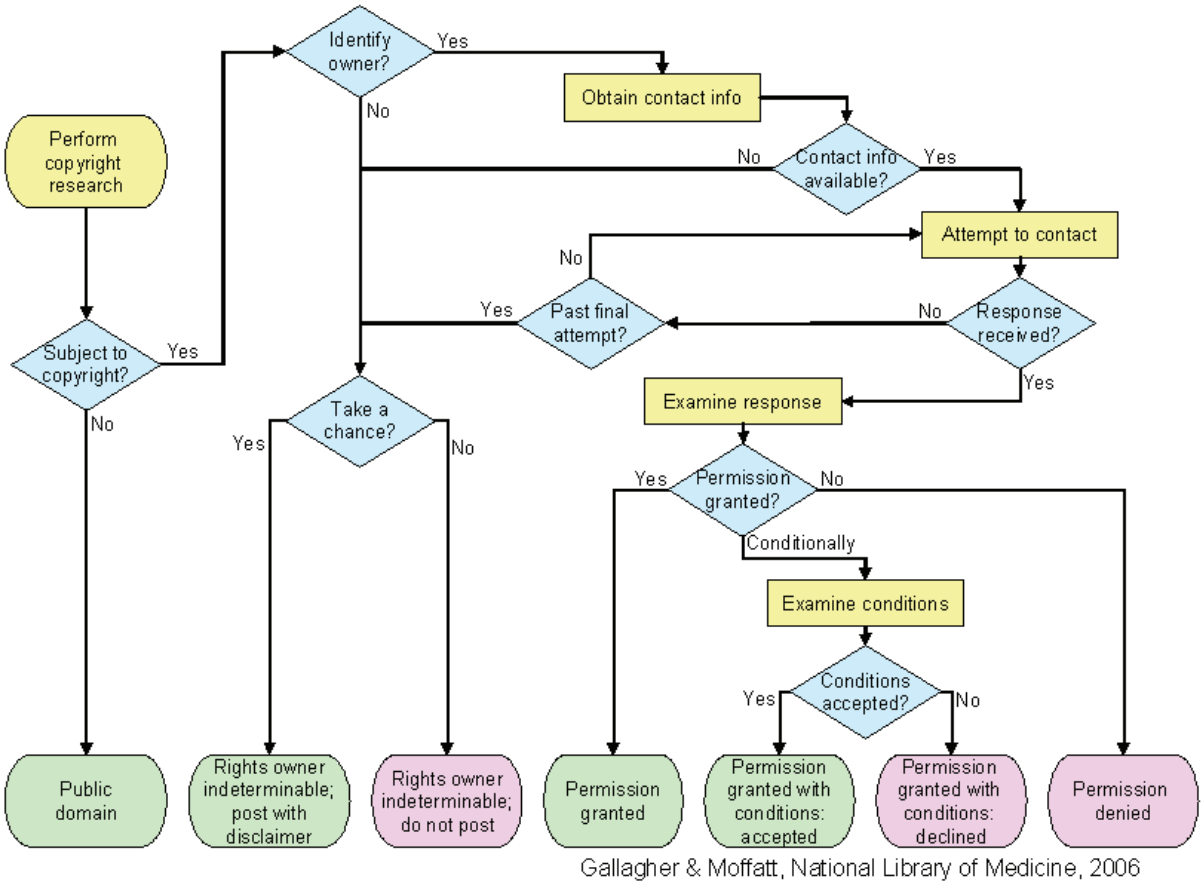

Fig. 3: Determining copyright status. Items with a "green" status can be posted on Profiles in Science.

NLM adheres to professional ethics and legal statutes regarding privacy. Medical records, personnel records, and grant applications are samples of items subject to privacy concerns. Legal statutes such as the Privacy Act of 1974 apply to NLM. However, the NLM is not a covered entity under the Health Insurance Portability and Accountability Act's Privacy Rule [45 CFR 160; 140.] The NLM HMD has adopted guidelines (Rees & Gehrlich, 2004; National Library of Medicine, 2004) to allow maximum access to its archival holdings while respecting the privacy rights of individuals whose records may be in the NLM archives. Profiles in Science adheres to these guidelines, which may prevent posting of some items. Because information is so widely available through the site, Profiles takes additional precautions to ensure privacy concerns are addressed. For example, photographs of doctors with their patients would not be posted without consent. If a photograph of an individual is taken in a private home, permission would likely be sought from the copyright holder (usually the photographer) as well as the subject of the photo as a courtesy.

Another legal consideration to which all federal government information technology is subject is Section 508 of the U.S. Rehabilitation Act. Section 508's "primary purpose is to provide access to and use of Federal executive agencies' electronic and information technology (EIT) by individuals with disabilities." (Section 508, 2002.) Before enforcement of the Section 508 legislation went into effect June 21, 2001, we had already been trying to adhere to the 1999 W3C Accessibility guidelines (Chisholm, Vanderheiden & Jacobs, 1999.) As a result, Profiles in

Science did not need a major overhaul to be compliant. Minor changes were made to the programs that generate the Web pages, and the pages were refreshed.

Standard procedures

The procedures for processing archival collections remain the same as before the existence of Profiles: materials are organized and arranged into series, sub-series, et cetera, to the folder level. Items are filed in acid-free folders, which are stored in acid-free boxes. Individual items are selected from collections by historians for digitization for Profiles. In the beginning we briefly considered treating all of the contents of a folder as one, digitizing and describing them as one single item. The items in a folder sometimes have different copyright requirements, and electronic searching for a particular item would return the contents of the entire folder instead of only the item of interest, so we immediately realized that considering a folder to be a single item was unworkable. When working with the physical collections, we wanted a visual cue to indicate which of the items had been digitized; this is particularly helpful when two items are similar in appearance, and only one was digitized. We first experimented with making item level notes in one collection's EAD finding aid, indicating which items had been digitized. We did not make item level notes in subsequent EAD finding aids, as this was labor-intensive and the result was not user-friendly. We used to indicate that an item had been digitized by placing the item in a Mylar® sleeve and marking the sleeve with the Profiles unique identifier. Placing whole Mylar sleeves around each digitized item eventually added so much bulk, as well as cost, that doing so became impractical. Now we use a two-inch by eleven-inch slip of acid-free paper, folded at the bottom to hook around the paper document. On the slip, we note in pencil the unique identifier, box, folder, and scan date.

In the beginning we experimented with selecting items for digitization from unprocessed manuscript collections. We wanted to see if we could find ways to add newly acquired collections to Profiles quickly. Selection of items for digitization necessitated labor-intensive repetitious passes through the mysterious contents of boxes of unprocessed materials. Re-filing an item temporarily removed for digitization into an unordered collection was difficult at best, and threatened our ability to preserve the original order of the collection. We experienced duplication of effort such as selection of multiple versions of the same document and multiple copyright requests to the same copyright holder rather than consolidating requests for a set of items. The digital collections created from unprocessed manuscript collections also required more retrospective cleanup of metadata describing relations between items as well as addition or cleanup of finding aid information that would become available after the completion of processing. Dealing with the unknowns associated with an unprocessed collection increased the amount of time and effort spent on almost every step taken to add a collection to Profiles in Science. Working with processed collections has been far more efficient and resulted in smoother creation of digital collections.

Expectations

Before Profiles in Science, the NLM Modern Manuscripts Collection was available only to those able to come to Bethesda, Maryland and visit the Library in person. The Web has changed the expectations of donors and users of Profiles collections. Nowadays, portions of the papers are available instantly to anyone with a Web browser and a connection to the Internet anywhere in the world. Access to the papers on the Web gives users who might never have an opportunity to

make the trip to NLM a glimpse of these one-of-a-kind collections. Users can search multiple collections whose physical counterparts reside at institutions all over the world, a capability nearly unimaginable without the Web. Digital items from various collections can be viewed side-by-side on one's computer screen.

This has obvious advantages as well as some unintended side effects. We review each item selected for digitization for privacy issues, and we do not post sensitive items on the Web. Even so, a potential donor of a collection who is a particularly private individual might have second thoughts about personal correspondence, diaries and photographs being accessible instantly around the globe. One way we could alleviate these concerns is by delaying the posting of some items on the Web until some number of years into the future.

Mission

Over time, the mission of Profiles in Science has changed. The project began as a window into the NLM Modern Manuscripts Collection, allowing digitized samples of the contents of the collections to be viewed twenty-four hours a day, seven days a week. Profiles in Science also became a tool for recruiting future collections, demonstrating to potential donors and their families how much more can be done with their papers now than in the past. As the number of digital items and collections grew over time, the Profiles in Science data were sometimes used for testing and experimentation internal to NLM.

Possibly the largest expansion in the project mission has been to digitize manuscript collections of eminent figures in biomedical research and public health that reside in the archives of other institutions. Arrangements that resulted in short-term loans of processed papers to the National Library of Medicine have been smooth, efficient and mutually beneficial. We expect to continue to participate in these successful collaborations as the opportunities arise.

Organizational commitment

Organizational commitment may be the most important requirement for carrying out the activities essential to building and sustaining a digital library for the long-term. The National Library of Medicine has provided ongoing institutional/government backing (National Library of Medicine, 2000) from the project's beginning. McCray (2001) explains the importance of this organizational commitment:

> "Also important when embarking on a project within an organization is whether its senior management supports the effort. Because most digital library projects are long-term efforts, they require the commitment of long-term financial and human resources. Beginning such a project involves an implicit, if not explicit, commitment to the continuation of the work and a promise that the digital materials will continue to be available. Lacking organizational commitment, it may not make sense to even begin a project."

Organizational support has enabled Profiles project staff to focus clearly on developing the digital library, and has been essential to the growth and stability of the Profiles in Science program. Creating a digital library for the long-term requires ongoing planning, experimentation and adjustments due to rapid changes in the digital world. Institutional commitment and support makes it possible for us to perform the activities necessary to position Profiles in Science for long-term survival.

Survival Guide

Below are some short-term survival tactics we believe can help sustain a digital library for the long-term.

- Adhere to your basic principles and guidelines.
  The principles for building a digital library shown in Figure 2 have served us well and continue to do so. Don't lose sight of them, especially when a new unexpected challenge arises.
- Monitor developments in hardware and software.
  Constant vigilance is the only way to maintain the necessary awareness of new technologies that are emerging and winning over the market. You will need to decide which new technologies may be beneficial to your digital library, and to predict which new technologies will replace the technologies you currently use.
- Maintain and upgrade your hardware.
  We try to stretch our hardware investment as far as possible, but have found that upgrading hardware (such as servers and desktop workstations) every three to four years is necessary.
- Maintain and upgrade your software.
  The risk of software conflicts is ever present, but it is essential to perform upgrades. Consider having multiple identical hardware and software systems available. Being able to rehearse an upgrade allows thorough testing in a safe, non-public environment. Knowing that any software will eventually need to be replaced, choose software that allows you to conveniently extract your data so you can import it into the new software (rather than manually copying and pasting.)
- Refresh and replace your media.
  Like hardware, plan to replace your storage media every three to four years. Make multiple identical copies and store them at different locations. Verify the copies against each other before putting them in storage.
- Monitor and make use of standards, guidelines and best practices.
  No legislation, standard, guideline, or best practice is written in stone. Monitor and use them, but expect them to undergo improvements.
- Expect the contents of your collections to change.
  Expect donations of electronic items in various formats. Consider your processing and digitization procedures dynamic documents and keep them up-to-date.
- Expect to need more space.
  Your growing physical collections, digital collections, processing requirements and staff will require more space. Without compromising the safety of your collections, try creative solutions.
- Look beyond traditional expertise.
  Staff your project with people who like using technology. Having a mix of experts (historians, archivists, computer scientists) who enthusiastically learn beyond the scope of their expertise is a good combination.

- Be sensitive to possible privacy concerns.
  The Web may make personal papers too accessible for the comfort of some potential donors. Be aware that your Web collections can be used for reasons other than originally intended.
- Expect the survival tactics to change.
  Today our delivery mechanism is the Web. Remember when there was no Web? Something else may supersede it. Expect the coping strategies needed to deal with change in the digital world will require constant revision. Be adaptable. When no ideal solution is currently available, take short-term steps to tide you over until long-term solutions are found.

References

Adobe Systems Incorporated. (2004). PDF Reference, Fifth Edition, Version 1.6. November 2004. last built 25-Jan-2006. consulted 29-Jan-2006. http://partners.adobe.com/public/developer/pdf/index_reference.html

Adobe Systems Incorporated. (n.d.). TIFF. last built 25-Jan-2006. consulted 29-Jan-2006. http://partners.adobe.com/public/developer/tiff/index.html

Arms, C. & C. Fleischhauer. (2005). Sustainability of Digital Formats: Planning for Library of Congress Collections. last updated 31-Aug-2005. consulted 29-Jan-2006. http://www.digitalpreservation.gov/stage/formats/index.shtml

Chisholm, W., G. Vanderheiden & I. Jacobs. (Eds.). (1999). Web Content Accessibility Guidelines 1.0. 5-May-1999. consulted 29-Jan-2006. http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505/

Dublin Core Metadata Initiative. (2004). Dublin Core Metadata Element Set, Version 1.1: Reference Description. 20-Dec-2004. consulted 29-Jan-2006. http://dublincore.org/documents/dces/

Florida Center for Library Automation. (2003). Recommended Data Formats for Preservation Purposes in the FCLA Digital Archive. published 1-Jun-2003. revised 14-Jun-2005. consulted 29-Jan-2006. http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf

McCray, A. T. & M. E. Gallagher. (2001). Principles for Digital Library Development. *Communications of the ACM* 44(5), 48-54, May. http://doi.acm.org/10.1145/374308.374339

Microsoft Corporation. (2003). Windows Products and Technologies History. published 30-Jun-2003. consulted 29-Jan-2006. http://www.microsoft.com/windows/WinHistoryIntro.mspx

Murray, J. D. & W. vanRyper. (1996). *Encyclopedia of Graphics File Formats, Second Edition.* Sebastopol: O'Reilly & Associates, Inc.

National Library of Medicine. (2000). National Library of Medicine long range plan, 2000-2005/report of the Board of Regents, National Library of Medicine. Bethesda: National Library of Medicine. consulted 29-Jan-2006. http://www.nlm.nih.gov/pubs/plan/nlm_lrp_2000.pdf

National Library of Medicine. (2004). *Access to Health Information of Individuals*. adopted 17-Feb-2004. revised 10-May-2004. consulted 29-Jan-2006. http://www.nlm.nih.gov/hmd/manuscripts/phi.pdf

Raggett, D., A. Le Hors & I. Jacobs. (Eds.). (1999). W3C HTML 4.01 Specification. 24-Dec-1999. consulted 29-Jan-2006. http://www.w3.org/TR/html401/

Rees, J. & J. Gehrlich. (2004, October). *Sensitive and Private Information: HIPAA.* Session conducted at the Mid-Atlantic Regional Archives Conference Fall 2004 Meeting, Pittsburgh, Pennsylvania.

Section 508. (2002). Acquisition of Electronic and Information Technology Under Section 508 of the Rehabilitation Act: Questions and Answers. updated January 2002. consulted 29-Jan-2006. http://www.section508.gov/index.cfm?FuseAction=Content&ID=75

Stanford University Libraries. (2006). LOCKSS Program Home. last modified 26-Jan-2006. consulted 29-Jan-2006. http://lockss.stanford.edu/

Sun Microsystems, Inc. (2006). Sun History. last modified 27-Apr-2005. consulted 29-Jan-2006. http://www.sun.com/aboutsun/coinfo/history.html

Acknowledgments

We wish to thank our dedicated and creative Profiles in Science colleagues at the National Library of Medicine: Meghan Attalla, Caris Brown, Michael Flannick, Walter Hickel, Xiaohui Ma, Gregory Pike, Vreni Schoenenberger, Susan Speaker, Sandra Brauksieck Taylor, Michele Tourney and Karl Wolf. Thanks to John Rees for consultations about archival policies and legal issues. We are grateful for the essential encouragement and guidance provided by project leaders Elizabeth Fee, Donald King, Aaron Navarro and Paul Theerman. Thanks to the collaborators and donors of the Profiles collections for their willing and active participation. Thanks to Alexa McCray for her enthusiasm and foresight in leading the Profiles in Science program from its inception through the year 2004.

This research is supported by the Intramural Research Program of the NIH, National Library of Medicine. The Library, the world's largest library of the health sciences, is a component of the National Institutes of Health, U.S. Department of Health and Human Services.