

## **Building a Digital Library: With Comments on Cooperative Grant Projects and the Goals of a Digital Library**

*Elaine Peterson*  
*Special Collections Librarian*  
*Associate Professor*  
*Montana State University Libraries*  
*P.O. Box 173320*  
*Bozeman, MT 59717-3320*

### **Introduction**

From 1998-2000 Montana State University (MSU) received from the Institute of Museum and Library Services (IMLS), an agency of the U.S. government, a \$138,000 National Leadership grant to build an image database of Native American peoples that would be searchable on the web. Also included in the funding were monies for user education through the annual meeting of tribal college librarians held every year at MSU. An initial partnership was developed between three institutions—the three campuses of MSU, the Museum of the Rockies, and Little Big Horn College. Currently the database has over 1,500 images and can be found at [libmuse.msu.montana.edu:4000/nad/nad.home](http://libmuse.msu.montana.edu:4000/nad/nad.home).

IMLS annually funds digital projects in several categories to various types and sizes of libraries and museums. Awards and guidelines are listed at [www.ims.gov](http://www.ims.gov). IMLS is currently the largest federal funding agency granting monies for digital projects.

The Montana IMLS project can be seen as a model program of cooperation. Smaller campuses and museums usually do not have available the resources to create a functional digital library. Perhaps the biggest hurdle is the financing. The bulk of the project's grant monies purchased hardware and software. Without sharing, the individual institutions could not purchase a high-end scanner, expensive software such as that produced by the Oracle Corporation, or a Sun server.

### **What Is a Digital Library?**

The question of what really constitutes a digital library is beginning to be addressed by some authors. In his research on document imaging and digital libraries,<sup>1</sup> Levy writes that digital documents will be characterized by their materiality, boundaries, permanence, and variability. Furthermore, he asserts that these properties will be socially or politically determined by the interaction of the documents and people. What will make up a digital library will be the combination of a collection of documents and individuals' work. He notes that there is a complex set of relationships between documents, individual people, and the

technology itself. How this complex relationship is developed and maintained will be a key factor in any of the distinct document's use.

It is useful to reflect on the components of a digital library. It is not merely a collection of text documents or images or video clips, any more than a physical collection of books and photographs could be called a true "library." A library has a focus and within its focus it should be rather extensive. Furthermore, the individual items should be searchable, so they could be retrieved by an outside person. Thus, for example, a large group of photographs of my European vacation that I have posted on the Internet could not be called a true digital library. Those photographs have no focus, their meaning for an outside researcher is unclear, and they are not indexed or searchable individually.

What would make it a digital library? Levy's key items of documents, people, and technology provide an excellent framework. The first item, the documents themselves, is perhaps the least interesting part of a digital library! It is important to have the documents, of course, and any type can be included in a digital library. My hypothetical digital library might lack interest if it had text documents only, with no photographs of my trip. On the other hand, a digital library with only textual documents, and none of the bells and whistles of video clips and photographs, could be a very important one. After all, we all still search databases for the information contained in text files.

But it is the second item, people, which really pulls together a digital library. It must have a focus, and who can give documents focus other than an individual familiar with the items? Selection of documents for a digital library is as key to a project as a collection development policy for creating a "paper" physical library. Hazen, et al.<sup>2</sup>, from the Council on Library and Information Resources succinctly point out guidelines for selection of items. We must consider an array of issues for a digital library including their copyright, the source of the materials, current and potential users, the anticipated use of the materials, the relationship to other digital efforts, and their maintenance. All of these components are crucial.

The final element, technology, is merely the tool that pulls together the product that has been created by the people and the items contained in the digital library. While the technology is more than the hardware that runs the digital library, it is also more than the software indexing it or the search engine. Again, it is the people involved in creating the digital library, those who make those selections, and who also make the choices of terms for indexing. Although there is much written about creating a virtual digital library which can be indexed by machine, this is in many ways an illusion, since the person writing the program must at some point determine the indexing choices for the computer to execute.

### **Practicalities: Organization of the Workflow and the Database**

How were the components of documents, people, and technology applied to our endeavor? The focus of our IMLS-funded digital library is the "Images of the Indian Peoples of the Northern Great Plains." This topical focus was agreed to by the grant participants. One of the first steps was to write guidelines for inclusion of images in the database. Broad

guidelines were written relying heavily on the Museum of the Rockies' Photo Archivist/Curator, Steve Jackson, and a small committee.<sup>3</sup> Guidelines included such things as uniqueness, age (pre-1940 preferred), and quality of the image. The actual selection of the pieces would be done by each institution. The guidelines were left general so that there was flexibility for each participant. For example, when one participant found an image that was a 1890s photograph portraying a unique subject, but was of poor quality, it was still included in the database. Another example would be the discovery of an original, handwritten treaty with participants' signatures. That document was included as an image document, although that type of "image" had not been envisioned in the original planning.

The project began by scanning images at the Museum of the Rockies (MOR). Because of their extensive holdings and staff expertise, the MOR photographs were in the best condition and were the best organized and indexed. I do not believe that the MSU Libraries is alone in its less-than-optimal organization of photographs and images. Librarians are trained in the organization of textual materials, and are usually not as skilled in handling other formats. There are few small libraries with staff to handle the preservation demands of photographs. I would strongly recommend that any group of libraries undertaking a similar cooperative project should include a museum participant or at least a person trained in photo archive work.

A full-time staff member was hired with grant monies, and he worked with the MOR to pull images from a multitude of collections to include in the future digital library. Each image was first photocopied and a paper worksheet attached to include key items. The worksheet notes the title/supplied title of the photograph, artist/photographer, date, tribe, geographic location, format of the material (photograph, watercolor, etc.), the accession number/call number of the photograph within the institution's collections, and subject headings. Some images did not have all of the components needed for this listing, but as many as possible were included. For purposes of building the database we required that at least a title be supplied for each image, at least one subject heading, and the institution's internal numbering scheme for that particular image. This number would be crucial later if the image was viewed on the web and there was an inquiry to the institution. As the scanning proceeded to the other participants, many images did not have much information, but the minimum was maintained. It was also very helpful to have a single staff member handling the images. Although duplicates could be caught later, he was often able to spot them in the initial handling for the scanning process. Minimal handling of the images themselves was also a preservation concern.<sup>4</sup>

After completing work at the MOR, the Agfa flatbed scanner was moved to each institution, so that the actual materials never left their home site.<sup>5</sup> This was an important part of the project since many institutions might be willing to contribute to a central database, but are unable or unwilling to ship valuable or unique materials to another location. Two scans of each image were made, once at a resolution of 150 dpi, described as photocopy quality (high enough for research but not high enough for publication), and then saved in compressed JPEG and GIF files for web access from the server. The image was scanned again at 600 dpi and saved as TIFF files to be burned onto a CD-ROM as a preservation copy. These CD-

ROM preservation copies were later distributed to all participants for an offline storage copy.

The final step was sending the images back to the Sun server residing at MSU either by disk or ftp. Both methods were tried and used successfully. The images resided in the server and were unavailable to the public until the Oracle software was selected, the subject headings added, and the indexes built. Indexes were constructed for each of the categories listed on our worksheets: tribe, geographic location, date, etc. The Oracle software allows for searching within each field, Boolean searching, or can supply a drop-down menu of choices for the user unfamiliar with the index terms.

Subject access to the images was from the beginning viewed as crucial to the success of the database. Many images had no subject headings assigned to them. To increase access, the worksheets with an attached paper photocopy of the image was sent to an independent Native American consultant who was familiar with the northern Great Plains tribes included in the project and also has museum expertise.<sup>6</sup> He was able to examine the photographs and include additional headings. An average of five subject headings was assigned to each image. Although the indexing for the database is Dublin Core, Library of Congress Subject Headings are used for authority control.

The consultant's other role was to pull out images that might be questionable for inclusion in the database and future display on the web. These images he marked "culturally sensitive" and sent them back in a separate pile. Those images would later be sent to local tribal historians for their opinion. Most of these culled out images were not included in the database. The most common reason for marking an item culturally sensitive was because an outsider had photographed the Sun Dance, and it is still not viewed as a ritual for public examination.

Two further notes about the images should be made. First, we decided not to enhance the images, even though that is possible with the Photo Shop software and the Agfa scanner. A conscious decision was made to represent the images in the digital library as close to the original as possible. Second, at this point no "electronic watermark" has been put into place to safeguard the images from copyright infringement.<sup>7</sup> It was felt that any printed copying from the screen image would not be of a high enough resolution to jeopardize the integrity of the copyright. At the discretion of each institution, some images carry a small copyright statement that has been placed near the image itself. Each institution has been welcome to place an overall statement of copyright policy as a lead-in to their contributed images.

### **Technical Choices**

Although it sounds trite, the fact remains that keeping up the technology is impossible, or, at best very difficult. This fact is especially hard to deal with when working with grant monies. Our IMLS grant was written in 1997 and awarded in late 1998. What was listed for technical specifications in 1997 changed by the end of 1998 and certainly by early 1999 when most of the equipment was ordered and fully functional. How does one avoid

those discrepancies from listing what one wants in a grant application from what is eventually bought?

IMLS, like most funding agencies, has been very flexible. Unless one is testing the viability of a particular piece of hardware or software for a grant, it is paramount to remember that the purpose of most grants is the overall project itself. The purpose of our grant was to show the possibilities of sharing equipment and expertise, and thereby make a digital library project more viable for smaller institutions. Because of that, it ultimately did not matter exactly which equipment was purchased.

The most frequent question I am asked about the grant is our selection of the software produced by the Oracle Corporation. I have heard from many people who have tried to build an indexed database in-house and have been frustrated. I believe that unless one is a large research university with adequate resources, this is not the route to take. Oracle was a good choice for us. When asked if I would select Oracle again, or recommend it for another project, I can only say that we made the best choice for us when we picked it several years ago. It is still an outstanding product, but today there are other options in the marketplace that were not available just a few years ago. For example, our online catalog runs on the Sirsi software, and that company now has an off-the-shelf database product (Hyperion) able to do the kind of indexing that we desired.

Our current array consists of:<sup>8</sup>

*Type of indexing:*  
Dublin Core

*Authority file:*  
Library of Congress Subject Headings

*Software:*  
Oracle 8.0.5 database software  
Oracle Application Service 4.0.8.1 Enterprise edition  
Oracle Developer  
Oracle Developer Server  
Adobe Photo Shop 5.0  
Equilibrium Debabilizer Pro 3.0  
OmniPage Pro 8.0  
Adobe Framemaker +SGML 5.5  
Adaptec Toast 3.56

*Equipment:*  
Sun Enterprise 250 Server running Solaris 2.7  
UPS and Tape Backup Drive  
Macintosh G3 with 19" Viewsonic monitor  
Dell Pentium II with 19" Viewsonic monitor  
Agfa T2000XL scanner with FotoLook 3.03

Iomega 2gig Jazz drive  
Yamaha rewritable CD-ROM drive

## Conclusion

Collaboration between libraries, museums, and archives will be critical to create meaningful and complete digital libraries. All documents are needed no matter where they reside, a variety of staff is essential, and many types of technology should be employed. While some larger institutions are creating databases of images, this project demonstrates that there is a workable model that can be used at many types of institutions. The project's main goal of creating a shared database with shared hardware and software was accomplished.

Understanding what a digital library truly is and what it can do was not part of our naïve writing of the grant application. It was exciting to discover that a new entity, the digital library, can be created which is greater than the sum of its parts. Each of our institutions own a discrete set of images of Northern Plains Indians, often with little access, and with overlapping collections with each other that we did not fully comprehend. What we have now in our digital library is the best of each of the collections that we determined ourselves, with full indexing, and no overlap. Additionally, we have created a preservation copy of our images on CD-ROM and have limited the wear and tear on the original images by our creation of a surrogate digital library collection. We have used an array of technology from the lowly photocopy machine to the high-powered Oracle software. But, most important, we have drawn on the expertise of many people to pull together a meaningful digital library.

In explaining my project to a colleague in literature and the arts, he mentioned that it had been difficult in his research to track down some original manuscripts and images in Europe because many of the items were not indexed/listed and therefore not accessible. But, more to the point, many items resided in private collections or tiny local museums that are not open to the general public. I see great possibilities now from our IMLS project for this type of research problem. Using our model, a lead institution could define a topical subject grouping, and then proceed to create a digital library. Participants would not have to give up ownership of the items and would not have to transport them to another location in order to be included in the database. Participants could limit future access as much as they deemed necessary, but at least there would be some indication of an item's existence, even if it were only a digital surrogate. Something would be better than nothing.

In the United States there have been many projects to create digital libraries. The most publicized include the University of Michigan, the University of California at Berkeley, Stanford University, and other large research universities. However, the most valuable digital libraries need not be those with the support of a research infrastructure, nor those with all the items located in one physical place. The technology is available and we all have some important items to contribute, even if the items themselves reside in disparate locations. What will make the difference are the special choices that human beings make to define and then develop a digital library. Indeed, a digital library can and should be more than an inventory of images held at one institution.

## References

1. David M. Levy, *Documents and Digital Libraries* (Palo Alto, CA: Xerox Palo Alto Research Center, 1996).
2. Dan Hazen, Jeffrey Horrell, Jan Merrill-Oldham, *Selecting Research Collections for Digitization* (Washington, DC : Council on Library and Information Resources, 1998).
3. We considered involving individuals from all the Montana Native American tribes at this point, but for efficiency decided to include them later in the project when they could respond to something more concrete.
4. For hiring purposes, our library made the wonderful discovery that today's college majors in photography are very suitable employees for a digitizing project. The person we hired for the grant had a B.A. in photography and was trained in all aspects of the discipline, including preservation and archiving images, burning CD-ROM images, and scanners, in addition to the science of resolution, color scales, and pixels.
5. Transporting the Agfa scanner was at times a problem. It occasionally broke down, but we were never able to determine if that was because of the particular model we purchased (a lemon?) or because of physically moving the scanner. Whatever the case, the Agfa Company promptly replaced the equipment and it is has not had problems in nearly two years.
6. Joe Horse Capture, a Montana native, is currently a curator at The Minneapolis Institute of Arts.
7. Gregory Heileman, Carlos Pizano, and Chaouki Abdallah, "Image watermarking for copyright protection," in *Lecture Notes in Computer Science* (Berlin, New York: Springer, 1999). The authors give a full description and technical specifications for electronic watermarks.
8. Howard Besser and Jennifer Trant, *Introduction to Imaging: Issues in Constructing an Image Database* (Santa Monica, CA: Getty Art History Information Program, c1995) provides an excellent introduction to digitizing equipment, the image source, resolution, compression, and standards, as well as providing an overview of concepts such as "What is a digital image?"