A digital library project on a shoestring

Mary S. Woodley*

University Library, California State University, Northridge 18111 Nordhoff Street, Northridge, CA 91330-8328, USA

Abstract

Supported by a modest grant and collaboration with local historical societies, the California State University, Northridge, University Library assembled and created a digital archive of 2400 images of photographs, manuscripts, maps, postcards and brochures chronicling the history of the San Fernando Valley, a suburb of Los Angeles. In the process the Library learned valuable lessons about administering a digital library project and its metadata. Supported by a modest grant and collaboration with local historical societies, the California State University, Northridge, University Library assembled and created a digital archive of 2400 images of photographs, manuscripts, maps, postcards and brochures chronicling the history of the San Fernando Valley, a suburb of Los Angeles. In the process the Library learned valuable lessons about administering a digital library project and its metadata. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Digital library; Digital archive; Metadata standards and workflow

1. Introduction

When librarians read in the professional journals about major digital library projects with impressive funding measured in the millions, all too often their reactions vary from funding envy to funding despair, and many conclude that such projects are out of the reach of four-year college libraries or public libraries. Descriptions of these mega-projects, with the profiles of the staffing and equipment, may give the reader the impression that substantial funding is mandatory to produce a worthy result. However, the experience at the California State University, Northridge, University Library contradicts this picture and presents another scenario: modest grants for small projects can create a huge impact.

* Corresponding author.
E-mail address: mary.woodley@csun.edu.

1464-9055/02/$ – see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S1464-9055(02)00255-X
In 1999, Special Collections and Archives wrote a proposal with the Library Dean and the Head of Library Development to the State Library to fund a project that would create a digital archive documenting the history and growth of the San Fernando Valley, sometimes called “America’s prototypical suburb” [1]. The resources for the project were primarily derived from the University’s and Library’s materials. Resources owned by local historical societies, libraries and private collectors enriched the project’s scope and content. The project focused on images free of copyright restrictions (or ones where copyright permissions had been granted by the owning institution). Textual material was included if it illustrated the history and growth of the Valley. Money for the project came from a grant awarded by the California State Library through the federal Library Services and Technology Act (LSTA). The State Library grants are part of the IMLS funding (Institute of Museum and Library Services) [2] to the states to provide seed money for digital projects. In California, the State Library funding ranges from $5,000.00 to $500,000.00 [3]. At the time of writing the grant, the Library had no experience with digital projects. The Project was awarded $153,298.00 in funds; the Library provided another $43,000.00 “in-kind” funds, i.e., 10% release time for the Head of Special Collections, Head of Urban and University Archives, and the Library’s System Director [4].

2. What we learned

2.1. Try, try, and try again

Special Collections and Archives’ past experience with writing grant proposals has had varying success [5]. It came as a surprise, therefore, when the LSTA grant was awarded the first time through the process. The success of this grant proposal may be attributed to lessons learned from previous experiences. One of the most important lessons is asking trusted colleagues who were not part of the grant team to read through the proposal to ensure that all of the criteria were met and that the proposal was clearly written.

2.2. Collaboration and patience pay off

One of the greatest rewards of the project was the creation of and collaboration with the San Fernando Valley Heritage Network, a confederation of local historical societies and collections spearheaded by the Library’s Special Collections and Archives. We, the digital library project developers, realized that a wealth of material on the Valley existed in local historical societies and private collections in the San Fernando Valley. These collectors and members of these societies usually work as volunteers outside of the formal municipal and educational institutions. The collections may have adequate housing but are as likely to be stored in boxes in a volunteer’s garage with very limited hours of access. More importantly, researchers rarely know these collections exist, much less the scope and nature of their resources. The Network provides a mechanism for providing preservation workshops for archival resources: documents, letters, manuscripts, photographs, maps, canning labels, etc. Its formation also provided a means for the Library to work collaboratively with Network
members to create a digital resource much richer than any resource that could be created independently.

2.3. It always takes longer than you think!

The initial design of the project required the hiring of a Project Archivist and a Project Technician. The main duty of the Project Technician was to scan the material for the project. The role of the Project Archivist was to curate the project, determine the appropriate material for the project from our Special Collections, Urban Archives Center and University Archives, as well as material from the members of the Heritage Network. Both were to share the task of creating the metadata records. This turned out not to be feasible. The Archivist’s time was consumed with interfacing with the Heritage Network members and seeking out new sources of material. Network members are mostly informal organizations whose collections are not organized in a way to be conducive for researching. The Project Archivist spent hours with individual volunteers, selecting individual images and writing down the “oral history” associated with the images. Later, the information for each image needed to be verified and enhanced by the Project Archivist or metadata catalogers.

The scanning process was also complicated, because material was often brought in from various sources of the Network at the same time. It was necessary to create steps in the workflow to provide safeguards to ensure that various collections would remain separate and not be intermingled.

Further, the State Library sent the announcement that the Library had been awarded a grant August 31, 2000. The official start date of the grant was October 1, with the funding to be sent in two installments. Since specific individuals were not named in the grant proposal for these positions, we learned that the University required that we follow the same hiring procedures as mandated for employment at the University: a vetted job description, formal job postings, and formal interview process. Thus, instead of project staff starting in October, they did not start until the following January, three months behind schedule. However, the inability to hire anyone immediately turned out to be a silver lining, because the delay left sufficient funds in the budget to hire two part-time metadata catalogers midway through the project.

The original project plan had underestimated the time involved to create a cataloging standard and the metadata associated with individual images. LSTA and the California Digital Library both provide general metadata guidelines [6]. The LSTA grant requires the projects to follow the minimum scanning and metadata guidelines (based on Dublin Core). In November the author joined the project as an “advisor.” In this role, she wrote local standards and procedures and a basic workflow plan. However, by February we realized that metadata creation was not on schedule and we needed to redesign our workflow. The Dean provided release time to supervise the cataloging. We hired an experienced part-time cataloger as an emergency hire in April, and a second cataloger in June. The addition of two catalogers accelerated the process immensely. Temporary emergency hiring allowed us to circumvent the normal campus hiring process; however, these two individuals worked without benefits. For our new proposal, we are planning a workflow based on a nine-month schedule rather than on twelve months.
2.4. Metadata standards and workflow

Figure 1 is one of the project images, and Figure 2 shows the image’s metadata. For our project, fields are based on Dublin Core 1.1 Qualified. This is fortuitous since the metadata standards of the California State Library are also based on Dublin Core. As a test of what fields the project needed, we created a test bed of fifty records for various types of material from the Library’s collection. Only the minimum fields required by the LSTA standards were used. The staff evaluated the level of information the records provided and the appropriateness of the field names or labels. Our “best practices” [7] document the name of the field, the order of display, whether or not the fields are under authority control, whether or not the fields are part of the metadata template, and the definitions of how the fields are used. This was imperative so that each cataloger understood the use of the fields and would record content consistently. The definitions and usage of the individual elements and qualifiers follow the standards established by the California Digital Library and the California State Library Metadata Standards.

In the end, we expanded the number of fields from Dublin Core’s standard fifteen to twenty-four, all of which map to Dublin Core. The decision to expand the number of fields was driven by the need to record or display information required for identifying, retrieving, or evaluating individual images, or by the need to record administrative or preservation information. We took advantage of the fact that none of the Dublin Core fields are required but they are also all repeatable.

The names or labels we used for the elements were influenced by the fields defined by the Visual Resources Association standard, although some of the labels were locally defined. Since our audience will include the K-12 population as well as the general public, we preferred to use field names that reflect vernacular usage. For example, see the label “Neighborhood” in Figure 2, which corresponds to the Dublin Core element called Coverage. The Dublin Core Metadata Initiative’s best practice recommendation for Coverage is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names). In Figure 2, the neighborhood name “Burbank (Los Angeles County, Calif.)” conforms to the form found in the LC Name Authority File.
In fact, our policy was that names of people, corporations, places and topical headings should conform to forms established in the LC Name Authority File and Subject File, or if not established, to forms devised by following current cataloging standards. A fuller discussion of our metadata practices may be found in the paper the author presented at the Dublin Core Metadata Initiative’s 2001 annual meeting [8].

A documented workflow supports the ability to work at production level, and at the same time, to keep control over materials coming from a variety of sources. Although many of the historical societies lack any formal inventories for their material, the Project Archivist listed each item borrowed on a loan form with a detailed description. Material that came together in a box or folder was kept in the same order as received. The Project Archivist collected items from multiple societies on any given day, but the Technician scanning the material worked on a single collection at a time. Each digitized image received a name based on the collection name and a number based on its original numbering system (or sequentially if not). The collections were assigned to a single cataloger who used the original while cataloging. Once the metadata records for the collection were uploaded and reviewed, the original material was returned to the appropriate historical society. These steps were important to build and maintain positive relations within the Network.
2.5. Compromises: the money never goes as far as you think [9]

2.5.1. Staffing
Small grants force project administrators to make hard decisions about how to spend the money. In our case, the cost of hiring project staff members was by far the most expensive aspect of the project. The budget for the salaries and the benefits for hiring two full-time project staff members to curate and scan the materials was originally computed as $108,951.00. This left very little money to hire a database creator or web designer, a fact that influenced our decision to buy a database from a vendor.

2.5.2. Software
One of the most difficult decisions centered on the selection of the project database. We saw three options:

1. Create and design our own database and supporting web pages from scratch;
2. Borrow already-developed databases based on filemaker pro or microsoft access and design a search interface using coldfusion;
3. Buy a product from a vendor that supplied the support and regular software updates at a price within budget guidelines [10]

We compared the costs associated with maintaining a homegrown database, which included the salary of an expert to continue to develop the database, with those costs of buying a product from a vendor. We concluded that the Library would have found it too difficult to maintain our own system without a guarantee of continuous grant money. Although the product chosen may not offer every bell-and-whistle that we would have liked, it allowed us to focus on the content of the digital library rather than on database design. Fortunately, the funds that we had budgeted for software and contracts covered the costs of the vendor selected [11].

2.5.3 Hardware
Fortunately, by the time we were ready to purchase the hardware, the costs had dropped for many items. The LSTA grant requires that any equipment that costs more than $5,000.00 be budgeted under “Equipment.” Other hardware would be calculated as part of “Operations.” We purchased a Dell server, an oversize HP Scanjet, one workstation (desktop with a 17” monitor), and a tape storage system.

2.5.4. Hidden costs
When budgeting the entire project, it is important to discover possible “hidden” costs. On our campus, the “Corporation” uses a percentage of any grant awarded to cover the costs of its administration. In this case, the ceiling that the LSTA allowed for management of the funds was lower than normally required by the University Corporation. Otherwise the Corporation may have earmarked a higher percentage of the grant for fund management. However, purchase requests through the Corporation were processed faster than through normal channels, which meant vendors received their checks faster and we received the
products and service updates more quickly. It still fell on the shoulders of the Principal Investigator, or Project Director, named in the grant to track the account expenditures.

3. Outcome

The San Fernando Valley History Digital Library opened with much fanfare and publicity. Local state representatives proclaimed October 16, 2001 as San Fernando History Day, with music and dancing troupes celebrating the traditions in the Valley. Fourth graders from a local elementary school were invited to attend and learn about the Digital Library as well as to display school projects relating to California history. Kevin Starr, State Librarian, spoke at our evening program praising the results of our efforts. Our greatest surprise was the electronic response to the Digital Library Web site. Both the Los Angeles Times and the Daily News wrote articles about the project, publishing the URL for the site [12]. The site thrilled the community and members of the San Fernando Heritage Network. We experienced 7,173 visitors to the Digital Library homepage during the first month. Visits have now leveled off to about 2,000 a month.

4. Future

After successfully completing a digital project, the Library feels confident about developing new digital projects that would benefit the University faculty, students and the community. The first year proved to be a wonderful learning experience that the Library can apply to future efforts.

References

[1] Fulton W. The plains of Id, book review of The San Fernando Valley: America’s Suburb, by Kevin Roderick. Los Angeles Times: Los Angeles, Calif.; September 9, 2001. It was fortuitous that Kevin Roderick’s new history was released a week before our opening. The Library was able to take advantage of the timing with our opening ceremonies.

[2] Check the Website for the IML for more information: http://www.imls.gov/


[4] LSTA Grant F-7, FY 2000/01, WP99, Grant Award #40–5208. The U.S. Institute of Museum and Library Services support the project under the provisions of the Library Services and Technology Act, administered in California by the State Librarian.

[5] I would like to thank Tony Gardner and Robert Marshall for sharing their experiences with me.


[9] Estimating the costs for a digital project may seem like a moving target. For help consult the RLG tools for digital imaging (http://www.rlg.org/preserv/RLGtools.html) including their Worksheet for estimating digital reformatting costs.


[11] We chose CONTENTdm (http://contentdm.com/), which provided us a package based on the size of the database and three years of software support and updates.