

The Western Waters Digital Library: Providing Solutions through Collaboration and Technology

Susan Salem and Kenning Arlitsch

Abstract

This paper describes current strategies for promoting the long-term success and sustainability of a collaborative and distributed digital library project. The Western Waters Digital Library (WWDL) is designed to develop from a framework incorporating a common digital management system, uniform application of digital and metadata standards, and extensive collaboration. The ultimate aim is to become a comprehensive information resource for a wide and varied audience concerned about water issues in the American West. Challenges faced by the WWDL in confronting these issues will be described along with the strategies now in place to address them.

Rationale

The geography of the land west of the 100th meridian is largely arid, and without its high mountains to catch snowfall and large rivers to channel snowmelt, most of the West might be incapable of supporting large populations. The twentieth century witnessed

political maneuvers that led to engineering marvels in the form of hundreds of dams, reservoirs, and canals, diverting the snowmelt and supporting the growth and economies of major cities. Dam building had unforeseen consequences, too, affecting the environment and those unfortunate enough to be at the end of the water line.

More recently, increasing population growth, pollution, conflict over water rights, and seven years of continued drought have raised alarm about the allocation and distribution of water. To date, there is no single organization or other collaborative group that provides a comprehensive information resource about water to researchers, policy makers, educators, and citizens. If westerners are to meet the challenge of managing their water resources wisely they will need access to information resources and analysis based on research and accurate historical records. In its final report in 1998, the Western Water Policy Commission suggested that water data should be collected and archived on a river-basin basis, and every effort should

Susan Salem is Project Manager, Western Waters Digital Library, University of Utah, email: susan.salem@library.utah.edu and Kenning Arlitsch is Head, Information Technology, Marriott Library, University of Utah, email: kenning.arlitsch@library.utah.edu.

be made to make the data easily available to all basin agencies and the public.¹ The WWDL was created in response to this critical need for broad access to historic and contemporary records regarding water in the western United States.

Overview

The WWDL was launched in late 2003 by the Greater Western Library Alliance (GWLA) with a two-year grant from the Institute of Museum and Library Services (IMLS). GWLA is a consortium of thirty academic research libraries in the central and western U.S. In this initial phase, twelve GWLA libraries in eight western states are producing and hosting digital collections pertaining to the Platte, Rio Grande, Colorado, and Columbia river basins. The metadata describing those materials are being aggregated to a central server at the University of Utah to enable virtual, seamless searching of all the collections from a single web site. During the IMLS funding period, the WWDL aims to accomplish three principal objectives:

1. Create a viable technical infrastructure for aggregating geographically-dispersed collections for searching from a single web site;
2. Lay the foundation for the continued development of a comprehensive digital information resource about water in the west, and;
3. Establish a model for cooperation and collaboration.

Current strategy is focused on developing a sound but flexible technical infrastructure, productive collaboration, and additional funding sources. The long-range strategic vision is to increase both geographic participation and coverage over time. At this writing, the WWDL has created a collection of historical, legal, and government documents totaling over 40,000 images, but this is just a beginning. To fulfill the intended purpose of becoming a comprehensive digital resource, the WWDL must provide easy access to the information needs of a vast and varied audience.

Technology

The success of collaborative digital programs is contingent on the contributions of the participating institutions, but also on the benefit and recognition brought to those institutions. The technology infrastructure chosen for this project established conditions that allow each partner to contribute equally, maintain direct

control of their collections, and retain their institutional identity when their objects or collections are viewed by the user.

The Digital Projects Task Force of GWLA chose CONTENTdm™ as the digital asset management software. CONTENTdm was already being used successfully at many of the participating institutions, and it was serving as the technological base for the Mountain West Digital Library, a collaborative project of Utah and Nevada.

In the WWDL model, a CONTENTdm server is installed at each participating institution. Servers run on Windows®, Linux®, or UNIX Solaris® platforms, with each institution selecting the platform most suited for its environment. The project team at each institution digitizes and uploads their own collections, creating customized local websites for the collections as they see fit.

CONTENTdm aggregating software, known as the Multi-Site Server (MSS), is installed at the University of Utah. The MSS harvests metadata from the individual CONTENTdm servers in a manner similar to an OAI harvester, and harvested fields must be mapped to Dublin Core fields. Since no images are harvested the index is small and efficient; a daily, scheduled harvest takes only a few minutes and does not tax the remote servers.

A user searching the WWDL actually searches the aggregated index at the University of Utah. Upon selecting a search result the user is linked directly to the CONTENTdm server where the object or collection resides, and those objects are displayed with the owning institution's metadata, self-designed template, and identifying characteristics. Harvested metadata is never shown to the user; it is used only for search purposes.

Project Standards

At the outset of the project, WWDL members agreed to scanning standards prescribed by the Digital Library Federation (DLF), Research Libraries Group (RLG), and the Northeast Documentation Conservation Center (NEDCC). Some participants have chosen to do all scanning in-house, while others have outsourced scanning, particularly of large documents.

The project team also agreed to adhere to the "Western States Dublin Core Metadata Best Practices" (WSDC).² The WSDC is intended to allow

flexibility at the local level where necessary or desirable, while promoting effective searching of aggregated metadata. The WWDL allows local metadata fields to be named according to local needs, provided they are mapped back to appropriate Dublin Core fields in CONTENTdm.

Yet even generally accepted metadata standards could be subject to interpretation. Seemingly insignificant local application resulted in more significant problems once the metadata were aggregated. Following some discussion it was agreed that many of those fields did not matter for the aggregated index, since most users would not search on them anyway. It was agreed that only seven DC fields should be absolutely required: Title, Creator, Subject, Description, Date, Original, Identifier, and Type.

Aggregation of metadata by its very nature reduces search granularity. A user searching a local website that has been designed for a specific collection will benefit from customized search engines and metadata fields. Aggregation necessarily reduces that customization. In a 2003 report to the Digital Library Federation, Brogan admonishes that users must understand the levels of search granularity as well as the relationship of the collection to the host institution.³ To address these challenges, the Metadata Committee is working to develop specific guidelines for the WWDL. The WWDL Metadata Committee shares some membership in the Western States Dublin Core Working Group and thus serves as a conduit for refinements to WSDC and concerns that emerge in WWDL project implementation. In addition, project participants routinely review adherence to digital standards and report compliance to the GWLA Task Force.

Website Interface

Developing an effective interface for a varied audience poses serious challenges for an aggregated index, especially when expressed needs differ or conflict. Competing with well-established and pervasive search engines like Google poses another challenge.

Initially the WWDL search interface included a “Quick Search” box on the main page that searched across all metadata fields, including full text OCR, and across all types of materials. This search resulted in large numbers of hits, including individual pages of documents mixed together with photographic images, and many participants felt it was not useful. Parameters

were then altered to search only the assigned subject headings, but this was felt to be too limited.

As of this writing the new strategy is to code the “Quick Search” so that the user can decide to search “Documents” or “Images” prior to searching, much the way Google allows users to select “Web”, “Images”, “Groups”, “News”, etc. prior to conducting a search. Selecting “Documents” on the main page of the WWDL website allows the user to perform a keyword search but automatically requires the Dublin Core “Type” field value of the retrieved items to equal “text”. An “Images” search requires the DC Type field to equal “image”. Since the DC Type field is not populated for individual pages of documents this method ensures that searches retrieve only the item-level record for a document rather than all its individual pages. Once the document is viewed a user can perform a second search that will lead directly to the page whose OCR text contains the word they seek.

Results displayed for both types of searches can be varied so that the “Images” search displays thumbnails while “Documents” might display a bibliographic view. An Advanced Search option that allows a variety of other search methods, including phrase, Boolean, and multiple field searching is part of the Multi-Site Server interface.

Collaboration

To develop a truly useful collection that will continue to grow and evolve over time, the WWDL team must not only collaborate amongst its members and with GWLA, it must create productive new relationships on campus and with external stakeholders. As described by Bunker and Zick, collaboration has become a basic prerequisite in developing and sustaining digital libraries.⁴ In fact, collaboration is essential not only in developing workable technology, but in developing the collections themselves, and in getting the word out to end-users and stakeholders.

The Library Team

Digital project teams must continually adapt to new and evolving technology, shifting institutional priorities, and funding constraints. The ability to do so may depend as much on wide-ranging collaborative relationships as on technical savvy, particularly when project teams are widely dispersed. One obstacle in maximizing internal collaboration stems from the

funding crisis pervading academic institutions across the country. While public universities struggle to keep afloat in the current climate of reduced funding and increasing costs, library staff are shouldering mounting responsibilities with scant, if any, additional remuneration. In this climate, the creation of collaborative digital libraries relies mainly on reassignment of existing personnel, who often juggle multiple projects and sometimes conflicting institutional priorities. A side-effect of such added responsibility is often turnover within the project team, which, on a two-year or three-year project, can have a detrimental effect. In addition, digital library teams tend to be interdisciplinary, bringing together members of departments that have not traditionally worked closely together, and it takes time for such teams to become fully functional. A resulting challenge is to ensure buy-in for the project at the outset, especially when new endeavors necessitate the full engagement and commitment of project staff.

The WWDL is approaching internal challenges by striving to maintain clear, open, and regular communication between the extended project team, GWLA leadership, and other staff of participating libraries. The master timeline is designed to be flexible enough to accommodate local circumstances. The GWLA Digital Projects Task Force is charged with guidance and oversight, a member of the GWLA Board of Directors serves as liaison between the Board and the project team, and sub-committees are established on an ad-hoc basis. However, while open communication and flexibility encourage productivity, they are not enough to support digital libraries for the long-run. A major challenge to collaborative projects in academia is outmoded organizational structures that require transformation to achieve technological currency.⁵

Campus and External Stakeholders

The adage, "if we build it, they will come" may no longer be true in the case of digital libraries. Resources are far too limited to risk a collection that does not meet user needs or an interface that is difficult to use. The development of a meaningful and comprehensive collection depends greatly on establishing and maintaining collaborative relationships elsewhere on campus and in the community. It is incumbent on WWDL institutions to establish on-going communication with departments that have a vested interest in access to digital information about water in the West. It will be equally

important that they work closely with state and federal regulatory agencies, advocacy groups, environmental concerns, practitioners and decision-makers, as well as the general public. The WWDL is working to establish and maintain these relationships to determine, assess, and respond aptly to end-user needs.

Collection Development

The organizational and collaborative challenges described above also affect the development of a meaningful collection. It's the same old problem: not enough time and not enough money. The answer may be identifying subject specialists with personal or professional interests in developing digital resources about water in the west.

Upon completion of the current IMLS-funded pilot-project, the WWDL will be demonstrative rather than substantive, representative of a variety of content and format. External collaboration is essential in identifying, prioritizing, and locating materials outside the holdings of contributing libraries. Many important water records from the 19th and early 20th centuries are owned by private companies, and are most often either in deteriorating condition or in danger of being discarded altogether. In addition, government capacity and will to preserve historic works is diminishing, and important government documents are now being sold. The result is that materials with the highest commercial value will be purchased, limiting access to those with means. But most critical records are not profitable, and therefore will likely disappear. For these reasons, it is imperative to partner with appropriate government agencies now, so that the records in question can be preserved in the public domain for future use. To address this situation, the WWDL has invited well-known practitioners in water-related science, law, and regulation to form an advisory board that will help guide collection development policy, obtain access to endangered documents, and provide support. The response thus far has been enthusiastic, and expansion of the advisory board is planned to ensure appropriate regional and subject area coverage. In addition to rescuing and digitizing important materials, a meaningful and relevant collection must address the following questions:

1. What are the dominant water issues in the West? (e.g., climatic trends, pollution, scarcity resulting from environmental change and population growth,

economic development, legal precedents, historical vs. equitable allocation, political influence, international and interstate conflicts, conflicts by user demographics, Native American rights, recreational use, conservation).

2. What information is necessary to help make more informed decisions about the issues, and how will we know if this is happening? What subject categories, sub-categories, and time spans would be most useful?

3. Are all the collections currently available meaningful in addressing the issues? Do they have other value?

Getting the Word Out

Digital Libraries cannot depend solely on the Internet to reach intended audiences, but marketing has not traditionally been a priority of the academic library community. Although aptly designed marketing campaigns may now be ever more necessary, resources in academic research libraries are too precious to waste on conventional mass marketing methods that are most likely to be overlooked or ignored completely. A larger and perhaps more pressing challenge is that patrons and other stakeholders do not always understand the capacity of research libraries. In addition, many who would benefit from library digitization efforts are unaware of the pervasive information loss the civilized world faces, thinking that a "Google world" will provide access to everything now on paper. The need to raise awareness of critical roles that research libraries play in a functioning society is widely recognized in the professional organizations. ALA and ACRL have responded with the "Campaign for America's Libraries", with a particular focus in 2003 on promoting the value of academic and research libraries.⁶

Publicizing value and service to the community and getting the word out about the WWDL should be complementary endeavors that strengthen both the core library and the digital library at the same time. By working with representatives of various target audiences, the WWDL intends to solicit feedback, encourage involvement, and engender on-going interest and support. These efforts constitute an as yet informal marketing effort, enhanced by presentations at professional meetings and conferences in the library, legal, and river management arenas. The result so far has been welcome publicity in a local newspaper⁷ and a national publication.⁸

When WWDL collections are harvested from all twelve participating libraries, a more formal publicity effort drawing on initial contacts will be initiated. In light of pervasive funding constraints, the most practical approach may well be using collaborative relationships as both a springboard and entrée to widening audiences of end-users and supporters.

Funding

The WWDL aim is to continue building content by adding legal, historic, scientific, statistical, and spatial data, and by increasing geographic coverage in subsequent phases of development. To that end, the WWDL is developing an orchestrated effort to seek funding on multiple tracks, including, but not limited to: content development, outreach to interested parties, collaboration in the digital environment, refinement and enhancement of digital technologies, and digital technology research. Collaborative projects are becoming more and more attractive to funding organizations, and the WWDL hopes to build on the collaborative efforts initiated thus far to seek future funding from federal agencies and private foundations and donors.

Conclusion

Because the creation of digital libraries is a fairly recent phenomenon, those now considered exemplary have only a scant lead on newly initiated projects. None have achieved the longevity by which to judge effectiveness and viability for the long-term. In addition, other central digital library issues outside the scope of this paper, such as digital preservation and evolving technologies, pose additional challenges that further the need for collaborative problem-solving.

Notes

1. United States Western Water Policy Review Commission, *Water in the West: the Challenge for the Next Century* (Washington D.C., 1998) 43.

2. See <http://www.cdpheritage.org/resource/metadata/wsdcmbp/index.html> for Western States Dublin Core Metadata Best Practices.

3. Martha L Brogan, *A Survey of Digital Library Aggregation Services*, (Washington D.C.: The Digital Library Federation Council on Library and Information Resources, 2003), 74. <http://www.diglib.org/pubs/brogan/index-old.htm>.

4. Geri Bunker and Greg Zick, "Collaboration as a Key to Digital Library Development: High Performance Image Management at the University of Washington", *D-Lib Magazine*, March 1999, 5 (3). <http://www.dlib.org/dlib/march99/bunker/03bunker.html>.
5. Ibid.
6. American Library Association, 2003. Toolkit for Academic and Research Libraries: Messages, ideas, and strategies for promoting the value of our libraries and librarians in the 21st century. <http://www.ala.org/ala/pio/campaign/academicresearch/toolkitfinaltext2>.
7. Dan Nailen, "Preserving history: Documents are drying up and disappearing", *Salt Lake Tribune*, October 7, 2004.
8. OCLC Online Computer Library Center, Inc. and Subsidiaries, "OCLC Annual Report 2003/2004", <http://www.oclc.org/news/publications/annualreports/2004.pdf>.