

“Managing the Transition from Large-Scale Oral History Research to Digital Archive: The Digital Librarian’s Perspective”

**Cory Lampert
University of Nevada, Las Vegas**

Introduction

We live in an increasingly digital world. [Slide 2] Preparing for this conference I composed my paper in word-processing software, e-mailed the draft to my co-presenters, selected digital images and screenshots for our presentation, and announced my travel plans on the library staff blog. Cultural materials are also riding the wave to go digital and oral histories provide intimate pathways into the stories of our world. Despite experience in digitizing other formats, many institutions (such as UNLV) are just beginning to explore partnerships with oral historians, IT departments, and libraries to tackle the work of mounting transcripts, images, audio, and video into online collections. This paper will highlight some of the project design and workflow decisions that were considered while planning and building the digital archive of the Nevada Test Site Oral History Project.

Digitization for Access

Every project, large or small, needs to have a driving force behind it. For digital projects, the most common objective is to increase access to content. Often our most interesting historic materials are hidden in dusty archives yet to be cataloged, or are restricted by limited hours in special collections departments. For the dedicated researcher, this is all part of the research process and many historians enjoy hunting down these rare documents. But as Roy Tennant, a prominent librarian who advocates for increased online access has said [Slide 3], “Librarians (and I’ll add historians) like to search. Users like to find.” This quote highlights a key benefit of digitization for access; once material is online it becomes available 24/7 to multiple concurrent users across the

globe. Dedicated researchers will continue to schedule appointments at the archives or library (as there will never be a replacement for experiencing original material in person), but the digital collection will also allow anyone with Internet access to locate and utilize materials straight from their laptops via Google.

Metadata and Search Functionality

Many of us have experience with digitization in our work or at home. If you've transferred a collection of family videos from VHS to DVD, posted digital images or documents online to collaborate with colleagues, or scanned a diary or scrapbook that is deteriorating; you have worked with digital content. The conversion of analog originals to digital surrogates is one of the first steps in the process, but when the amount of materials exceeds smaller manageable collections, the issue of organization becomes central to a successful project. Bibliographic control facilitates the storage and retrieval of library materials and this is an important characteristic that distinguishes between different types of functionality found in online oral history collections. There are many methods to organizing web content. **[Slide 4]** One method is the production of a static web site that might contain collection descriptions and links to transcripts, audio and video. Possibly a site search is included to provide a shortcut to users locating information. But often, users need to browse every item to identify its characteristics or are unaware of the depth of the collection if it is not digitized in whole. Web sites without database functionality can also prove unwieldy with large collections containing complex data. This where metadata comes to the rescue.

Libraries and archives have capitalized on thousands of years of experience describing and organizing information with two main goals: information should allow for both *searching* to locate a known item and allow *browsing*, **[Slide 5]** which promotes serendipitous discovery. It isn't surprising that over the years, librarians and archivists have developed many strategies, standards,

and guidelines for facilitating this process. In the digital world, metadata combines technical and descriptive standards that have grown out of traditional cataloging practices to respond to the need for information organization in the digital format. By designing project metadata to describe each individual item (each interviewee's transcript) within a collection rather than describing the collection itself (Nevada Test Site Oral History Project), metadata allows for far more powerful searching and retrieval functionality. It is true that description at the item level takes subject specialist knowledge and cataloging expertise as well as attention to detail in data entry. But, it is often a worthwhile investment because it delivers three important outcomes: searching that retrieves highly relevant results meeting exact user needs, the ability to gather like items under subject terms to encourage browsing, and it fosters collaboration with other digital collections and digital libraries by creating standards-based, re-useable metadata that can be aggregated into larger collections or specialized interfaces.

Good metadata can also build a foundation for meeting changing user expectations. As our digital world changes, proprietary systems may come and go. With standards-based metadata it can be less painful to import/export data when migrating systems. More flexible data can also form foundations for emerging Web 2.0-type interfaces and enhancements that allow users and researchers to interact with collections in more collaborative and participatory ways (such as tagging content or annotating items in customized or collaborative workspaces).

Digitization Process

The basic digitization workflow forms a framework for every project.[Slide 6] Having said this, it is important to mention that best practices are often just a starting point. Individual projects have many variables and local adaptations are to be expected depending on resources, time, and technical set-up. The NTSOHP web site hopes to encourage the sharing of methodologies and will

contain a web page detailing our choice of technical specifications, project procedures and information about the interview protocol used. The process, as a whole, can be summarized as three phases: digitization, metadata, and interface/web design.

After the project's scope and audience were identified, audio and image digitization standards were established so that quality remained consistent. The main content focus was the born-digital word-processed transcript and while these did not require digital conversion, they did need to be edited, batched, and transferred between three geographically separate departments and quality control proved to be tricky with multiple iterations of a transcript.

During discussions of the project as a whole, it became clear that unlike most library digitization projects (where we often receive un-cataloged, un-identified originals to work with); the NTSOHP's extensive documentation provided detailed descriptive information to build our metadata template. When the transcript was finalized to go to the bindery, the digitization unit received a digital "bundle" of files that included a biographical data sheet, a table of contents, the transcript, key vocabulary lists, and an index. We were able to easily transcribe much of this data into the metadata fields, which freed up time to focus on maintaining standards and uniformity. For subject indexing, the project employed free text, controlled vocabulary fields, and local subjects for authority control of various fields **[Slide 7]**. In addition to free text for titles and dates, controlled vocabularies/thesauri were used to ensure consistency in several fields. Often concepts have many possible terms and the process of cataloging is ultimately subjective. If researchers want to locate records on a particular subject they should not have to think of and repeat searches on every possible spelling or synonym for a term, so standard library thesauri were used to guide term assignment (i.e. the approved term: "employment" was used for variant concepts such as: work, labor, or jobs). In addition to the authority files created and maintained by the Library of Congress,

the project also designated other authorities (like the U.S. Nuclear Tests list) to govern the specialized terms not found in common cataloging references. The third type of subject field in the metadata is referred to as a local subject field. This field is special in that it is a wild card field used for concepts, sites, and events that are unique to the Nevada Test Site project. In this field any term is eligible for inclusion, as long as an appropriate controlled term is not available, it is formatted to conform to cataloging rules, and its relevancy can be verified. After all the decisions for the project metadata were tested in practice, procedures were recorded in a data dictionary [**Slide 8**] to serve as a reference guide for project staff to consult when questions arose during the metadata phase of the project.

Finished metadata is now being loaded into UNLV's digital asset management system to display and present complete searchable records to users online. We found that a working database was helpful during the editing and quality control phase and offers the option of automating import of data into the software/system where it will ultimately reside. The state of Nevada has chosen to use CONTENTdm software statewide for digital collections, and the Libraries load all locally-created digitized material into this digital asset management system for long-term storage and presentation.

The design of a web-based user interface to serve as a portal or gateway into the content occurs in the final phase [**Slide 9**]. We plan to use this web site to provide background narrative about the project, highlight links to additional resources, serve as a contact page for project staff, and publicly acknowledge grant funding. Searching and browsing menus and "CONTENTdm guided searches" will use metadata fields to narrow the large collection into easier to navigate themes. This interface and the database operating behind it will be tested for accurate retrieval and usability as well as compliance to web development standards before the project is launched and

announced to the public. Once the project is published to the web, marketing and publicity efforts will be planned to increase traffic to the site. Also in this phase, data back-up strategies and data migration plans for the future will be documented. Digital preservation and changing technology dictate that digitization never really “ends” in the traditional sense. But by refreshing and migrating data, keeping the look of the web site updated, and refining metadata as needed, the project will be maintained over the long-term.

NTSOHP Unique Challenges

I'd like to highlight some of the most interesting and challenging aspects of working on the NTSOHP. Notably, this is the first time the UNLV Libraries have endeavored to create a searchable digital archive of oral histories. Many of the best practices currently in place for our image and document collections, such as metadata procedures and scanning workflow were not applicable or required revision to adapt to oral history collections. Embarking on the project without any established procedures to follow was a challenge and required the development of strong partnerships and the cultivation of an environment where experimentation was encouraged. On this project, collaboration was essential to monitoring workflow and required individuals from at least 3 distinct groups to coordinate responsibilities. [Slide 10] Oral historians, librarians, and IT professionals all hail from diverse professional cultures and possess different departmental communication styles. We overcame these divides by delegating tasks. Campus and library IT set up a shared network directory to transfer files and upload data, while the Libraries took the lead managing metadata creation, CONTENTdm administration, and maintenance of the digital archive. The NTSOHP staff oversaw the batching and transfer of digital content “bundles” containing content-relevant supporting materials like the bio sheets, TOC, and vocabulary lists. Tasks really began to overlap and mesh as we worked closely on the metadata and web site planning.

Another challenge was the complex nature of the subject matter, which required open communication as questions arose. **[Slide 11]** Sorting through innumerable governmental agencies (and their acronyms), scientific jargon, and complexities like code names for tests or operations was difficult and having subject expertise a phone call away made all the difference. It became clear during these conversations that researchers and librarians have the same goals in mind: uniformity, recognition of major topical themes, and context. To meet this need we designed metadata records that include six top fields of interest for browsing (such as Key Individual, Locations, Tests/Experiments, Agencies, and Legislation) plus cross references in the indexing to capture variant forms. We also performed OCR (Optical character recognition) on the transcripts to provide full-text searching for comprehensive retrieval. Getting users into the transcript remains a key objective and major themes in the collection have been identified to help in designing a web site with thematic pathways that lead into relevant content (i.e. users will be able to click on a link for “Women and the Nevada Test Site” and the database will return transcripts representative of this subject). Ancillary material such as images, audio, and video will visually supplement the collection and enhance the user experience. **[Slide 12]**

Conclusion

Digital collections have wonderful potential to open up unique content to wider audiences and researchers around the world. The NSTOHP has taken an innovative approach to oral history digitization that values quality metadata and collaborative workflow to produce a high quality research-focused archive that will encourage a variety of users to connect to and discover the historic significance of the Nevada Test Site through oral history.