Best Practices for LSTA-Funded Digital Library and Digitization Projects in Oregon

Let's say your library has a box of loose and unsorted materials donated by a local historian. Interest in these materials does exist – or might exist, if the materials were more accessible – but with no way to organize or protect the contents of the box, you are hesitant to let patrons have a look and have no easy way to share your collection with institutions and individuals outside your community. A digital archive is a fine solution to these and other problems of preservation and access, but there are many factors that must be considered before your library makes the decision to embark on a digitization project. The following document outlines some of the core factors for success that the Oregon State Library has identified through the research of past Library Services and Technology Act grants and related literature and through conversations with past grantees and experts on digitization and digital libraries. Terms in **bold** are further defined in Appendix B: Glossary.

A. Material and Collection Considerations

Whether the materials you wish to digitize are old, one-of-a-kind, or falling to pieces is an important consideration when planning a digital project, but the condition or rarity of materials alone is not a satisfactory justification for digitization. Once your library has identified a collection or a group of materials that could serve as the foundation for a digital library, you must consider the following questions before embarking upon a grant project:

1. *What is the condition of these materials?* In their present form, are the materials endangered by routine physical handling? Is the need for preservation urgent due to imminent decay or, if the items are **born-digital**, data loss? What will happen to the original materials after they have been digitized? Digital libraries and archives can ensure that access to endangered materials will continue in the short run, but they cannot preserve originals and can be unreliable relative to archival technologies that produce new physical copies.
2. ***Who can access*** *these materials?* Do the materials circulate, and if not, do patrons need to secure special permission or supervision to view the materials? Do other libraries and archives have copies of your items, and if so, are they available to the public?
3. ***Who wants access*** *to these materials?* As with any library collection, your digital archive must serve an established need and be useful for those who demand it. Putting materials online may improve accessibility, but it will not create demand. Furthermore, the target audience for your digital archive – whether it is academics, casual browsers, hobbyists, K-12 students, or all of those groups and more – will determine how your archive will be presented. What expectations will your target audience have with regards to file quality, the scholastic value of citations and sources, and the Web interface that facilitates access to the collection? Does your target audience include (or potentially include) patrons who are outside your library's service area? If so, would their access to digital materials be restricted?
4. *What is the scope of your collection?* Patrons aren't willing to dig through a pile of unrelated items in order to find what they need – this is as true of digital

collections as it is with paper. Decide on a strong collection focus for your digital collection(s) and stick to it, even if it means putting aside valuable and interesting materials for use in future projects (see section 6b, *Outliers*, below). Positioning your archive as a "one-stop shop" for information on a specific topic will attract users and assure them that their needs will be met.

5. *Who owns these materials?* Can your repository get the rights to present materials in a format that your target audience can use effectively? Can your library purchase rights or physical materials that it does not own? If your library cannot distribute digital copies of the materials in a best-quality format at no cost to users, how will you ensure that the files will still be accessible in a format that is useful to your target audience (for example, your library may offer watermarked or lo-res image files for download at no charge with an option to purchase high quality copies)?
6. *Is there already a place for these materials on the Web?*
    a. *Duplicates.* Duplication of digital materials across several unaffiliated archives can confuse indexing and rights issues. Pursue cooperative partnerships with repositories that have a similar focus and that may have already digitized some of the documents in your possession.
    b. *Outliers.* Your library may own a few items that are in demand and could be good candidates for digitization, but that do not fit within the scope of any of your digital collections; for example, your "box of stuff" may include sixty photographs of local birds and one pioneer diary. Unless your library has surplus storage space to spare and could keep out-of-scope materials in a **dark archive** for later use, digitizing these materials anyway is not a good use of resources and may complicate access. Cooperative networks of local partners can help ensure that materials are included in appropriate collections.

B. Budget and Workflow Considerations

1. *Can we afford to put these materials online and keep them accessible?* The Internet provides plenty of cheap and easy solutions for storing and distributing information – if you're an average end user. For major online content providers, however, the cost of establishing and maintaining a permanent presence on the Web is much steeper. Digitization requires long-term budget commitments; your library and its partners will have to shoulder the cost of hardware, software, server space, bandwidth, labor, and training. Many of these costs are recurring, and some costs may be unexpected (for example, if hardware fails or hard copies are damaged).
2. *Is our library financially secure?* Digital libraries are especially vulnerable to financial crises. Repositories that cannot maintain the funds necessary to update their hardware, keep up with file format migrations, retain necessary staff or pay hosting fees may lose digital files to hardware failure, obsolescence, etc. Born-digital files that do not exist in any hard copy format may be lost forever. All repositories must have a rainy-day plan in place that provides for the continual storage and maintenance of digital data should the financial stability of the repository falter. Partnering with large, financially

secure organizations that have ample data storage space to share can provide a good safety net, especially for smaller institutions.

3. *How does this project end?* While LSTA-funded digital projects must have a continuing funding plan that will carry the project past the end of the grant period, the fact remains that some digital projects are essentially finite (as with highly specific collections that can only contain so many relevant objects, for example) while others could potentially go on forever. Open-ended projects must design workflow so that post-digitization tasks, such as fact-checking, cataloguing and proofreading, are not put off or forgotten.
    a. *Tying up loose ends.* If it seems like your project may come to an unplanned end in the near future, do not rush to fill major gaps in the collection. Concentrate on making sure that the pieces of the site that are already available to patrons are complete and accurate, and that all existing digital files are secure and may be exported. Keep written records and collection policies that you or your partners can reference in the future. Always assume that your library or a partner will someday return to your project in order to improve and build upon the work you have already accomplished.

C. Partnership Considerations

1. *Large repositories*, like academic and government institutions, can provide a stable support base for smaller digital repositories and may be able to lend their technical expertise to your project. Repositories with an established Web presence may be willing to include your materials in their collection or to provide hosting and storage for your digital archive.
2. *Small repositories*, such as museums, historical societies, and regional libraries, often have small collections of their own that could benefit from digitization or, if digitization has already taken place, improved accessibility or a stronger Web presence. Networks of small repositories can collectively sort materials into appropriate collections, and may consolidate materials and collections into one digital archive, index, or portal site. Duplication often occurs between the collections of smaller repositories; cooperative partnerships can see to it that duplication does not persist into a digital format and that all partners agree on issues of ownership and rights. In geographically isolated or unpopulated areas where access to digitization equipment is difficult to come by, small repositories can set up centralized or mobile digitization centers and provide cooperative maintenance.

D. Technological Considerations

1. *How do we share?* Your library must be able to export records into a common file format and share these files with other repositories. Cooperative sharing of digital objects and their corresponding records is critical to the development of digital archives.
2. *What software should we use?* Cost and skill must be considered when determining which digital collection management software your library or repository should use. The cost of purchasing **proprietary** software, such as

CONTENTdm and **PastPerfect**, starts high and gets higher as updates are released, but the software is easier for beginners to learn and may come with technical support services. **DSpace** and other **open-source** software costs little or nothing to acquire, but the learning curve involved can be very steep.

3. *What hardware should we use?* The expense of purchasing, maintaining, and inevitably replacing hardware is huge. Fortunately, not every library needs its own suite of servers and archival-quality digitization equipment. Seek out partnerships with libraries or organizations that have already purchased digitization hardware and are able to keep up with maintenance and updates. If your library is able to secure the funds to provide for the ongoing maintenance of hardware after the grant period has ended, create a long-term plan that will provide for continual technical support and upgrades for the hardware and that will allow other repositories and organizations to share your equipment.
4. *How do we provide the best access to these materials?* In order for your digital collection to be accessible and to facilitate resource sharing with other digital repositories, it must be searchable. This is achieved by applying thorough indexing, or **metadata**, to your digital objects; the ability to create metadata is included with most major digital collection management software. A digital photograph, for example, should at minimum be tagged with a title, creator, description, subject, place, date, and information about the publisher and rights. Consistency is critical to the metadata process. You must set standards and plan workflow, especially if the staff, volunteers or work-study students compiling your metadata are inexperienced or too numerous to train all at one time. Decide on a set of metadata standards that will guarantee uniform indexing for:
    a. *Names*. Is this biographical record for Mrs. John Smith, Jane Smith, Jane Smith (nee Doe), Jane Doe Smith, or Jane Doe?
    b. *Places*. Does this postcard depict the beach at Taft, Lincoln City, Taft (now Lincoln City), or Lincoln City (formerly Taft)?
    c. *Dates*. Was this document published on 12/20/59, 20/12/59, December 20, 1959 or 20 December 1959?
    d. *Descriptive titles*. Is this object a cider press, a cider mill, or a mechanical juicer?

    Provide data entry staff with a glossary of metadata terms, and always check to make sure that metadata is applied consistently across all records.

Appendix A: Selected Past Grants

1. 10-09-5a: Oregon State University Libraries: Oregon Digital Library Project
    a. Building on a previously funded grant project (*LibraryFind,* 09-15-1a), Oregon State University Libraries seeks to build a centralized portal for digital collections throughout the state of Oregon, available at odl.library.oregonstate.edu/record/search. This portal will give users the ability to search across multiple archives and will expose all harvested metadata for indexing by large search engines (Google, Yahoo!, etc.), thereby improving the Web presence of Oregon's digital archives.

2. 10-07-6y: OCTE on behalf of Oregon Encyclopedia: Libraries as Community Research Centers
    a. Through this ongoing grant project, the Oregon Council of Teachers of English (OCTE) is encouraging Oregon's minority populations to add their voices to the Oregon Encyclopedia project. The Oregon Encyclopedia (available at www.oregonencyclopedia.org) is a collaborative article-based digital archive administered through a partnership between Portland State University, OCTE, and the Oregon Historical Society. Libraries. The Oregon Encyclopedia is a great model for libraries that want to showcase original research or writing in their digital projects. All encyclopedia articles must meet rigorous scholarly criteria and are critiqued by an editorial board before they are published.
3. 08-13-4a: University of Oregon Libraries: Envisioning Oregon
    a. The *Envisioning Oregon* project brought librarians, historical society and museum directors, archivists, and other individuals involved with digital libraries together for a series of town hall meetings held at multiple locations across Oregon. At the conclusion of the project, a book was published (*Envisioning Oregon*, prepared by Gabriele G. Carey, Ph.D, PDF download available at scholarsbank.uoregon.edu/xmlui/handle/1794/9792) that summarizes the results of the town hall meetings. The book lists key organizations in Oregon involved with digital libraries, outlines objectives designed to facilitate improved collaboration between Oregon's digital repositories, and identifies best practices.
4. 08-09-2a and 09-08-5a: Oregon Institute of Technology: Crater Lake National Park Research Collection
5. 04-10-2a and 05-07-2a: Oregon Institute of Technology: Klamath Waters Digital Library
    a. The Oregon Institute of Technology has done two large-scale digitization projects with an emphasis on local environmental issues. The first project (available at klamathwaterlib.oit.edu) collected maps, photographs, and documents relating to water issues in the North Klamath Basin area, while the second project (at craterlakelib.oit.edu) focused on documents and **gray literature** associated with Crater Lake National Park. These archives provide a good example of how the online interface of a digital archive is determined by its target audience. The user interface of the Klamath Waters library contains more general interest documents and is designed for easy browsing, while the Crater Lake Research Collection's highly searchable, list-based interface is geared towards the specific objectives of researchers.

Appendix B: Glossary

born-digital: Materials that originate in a digital, rather than physical, form and are never published in hard copy.

CONTENTdm: Proprietary digital collection management software developed by the Online Computer Library Center (OCLC). CONTENTdm allows for upload,

description, management, and access to digital archives. OCLC offers free software evaluation packages by request. For more information, visit www.contentdm.org.

dark archive: An archive that restricts public access so that the materials it contains can be preserved for future use.

DSpace: Open-source digital collection management software commonly used in institutional repositories and in digital preservation. Like many open-source software options, DSpace is free and is maintained through a community development model. It is written in Java and will work on any operating system. Download at www.dspace.org.

gray literature: Materials that have not been archived or indexed, or that cannot be found through conventional channels.

metadata: Information that describes the content and context of digital files.

open-source software: Software with an available source code that may be freely studied, changed, or improved by any user under the terms of its license. Most open-source software is free or donor-supported and is produced through collaborative development between users.

proprietary software: Software licensed under exclusive legal right of the copyright holder. Licensees are given the right to use the software under certain conditions and are restricted from modifying or copying the program or its source code.

PastPerfect: Proprietary digital collection management software developed by PastPerfect Software, Inc. PastPerfect costs less than other proprietary software of its kind, but because it is designed for museum cataloguing it may not meet the needs of libraries and archives. A free software demo is available for download at www.museumsoftware.com

Prepared by:

Jane F. Scheppke
Volunteer, Oregon State Library
Completed 25 May 2011
Web addresses last visited May 3 2011
Best_Practices_for_LSTA_Digitization.doc