

***BUILDING THE PENNSYLVANIA DIGITAL LIBRARY:  
Guidelines for Topics, Materials, Priorities, Best Practices, Next Steps***

## **1. Introduction**

In 2007, the Collection Development Working Group of the Pennsylvania Advisory Committee for Collaborative Digitization (PACCD), a statewide coalition of leaders in digitization of and access to important resources across the Commonwealth of Pennsylvania, began the development of guidelines for the subject matter and content of a comprehensive, distributed statewide digital collection of Pennsylvania's history, culture, and society.

Funded in part by a grant from the Pennsylvania Library Services and Technology Act, the group convened a committee of experts in the history and resources of the Commonwealth to begin to develop a working list of important subjects and types of material that should be considered for digitization.

These guidelines will help libraries, museums, and cultural heritage organizations begin the process of digitizing their collections in order to share them with a wider audience, both in Pennsylvania and around the world.

This guide will explain how and why the material will be used by a variety of communities and audiences and will list desiderata for digitization, both of which, in turn, can be utilized to set priorities that will help funders from Pennsylvania and across the country determine the most urgent funding needs for digitization. Additionally, these guidelines provide a starting point for gathering information on the technical processes and recommended best practices of digitization. It is hoped that following these guidelines, practices, and specifications will result in high-quality digital resources for, by, and about Pennsylvania's history, culture, and society that will be of benefit to Pennsylvanians (and scholars) at home and abroad.

We hope to hear during a public review period from all interested parties about the guidelines, recognizing that not everyone will agree with the priorities our group has suggested. In order to keep these guidelines current and flexible and ensure that they meet the changing needs of the range of collecting institutions, we propose that the recommendations made as part of these guidelines remain in place for three years and then be reviewed and revised.

## **2. Audiences for digitized resources**

The Pennsylvania Digital Library content will serve the following audiences:

**Middle and high school students and teachers / National History Day.** Students will be able to use an online environment that is enticing to them -- and teachers will be able to create lesson plans and class assignments -- using sources that are unique and rare and in most cases do not leave their home repository. In so doing, they will meet Pennsylvania Standards for history calling for students to develop the skills of explaining, analyzing, interpreting, synthesizing, and evaluating the many kinds of historical sources. This is especially true in the case of activities relating to National History Day, a program that reaches nearly 1 million middle and high school students nationwide and engages them in historical research using primary sources.

**Local and family historians.** Nationwide, about 35 million people use the Internet for family history research, and Pennsylvania libraries and archives report significant use of their collections for this purpose. Increased digitization of these materials will aid a growing audience interested in connecting with personal and community histories, and will encourage a significant number of these researchers to travel to Pennsylvania to work in its libraries and archives.

**College students.** Pennsylvania ranks fifth in the nation in higher education enrollment and third in the number of higher education institutions. The same motives that animate the Pennsylvania Standards also prompt instructors of history courses in higher education to ask their students to use archival and other primary sources more innovatively. Having access to a large body of documents and other historical sources online will encourage college students to be more ambitious in undertaking their class projects and will provide them with a wider array of resources than may be found on their own campus.

**Scholars and researchers.** This audience, numbering in thousands, both uses Pennsylvania's historical collections and extends awareness of them by teaching, publishing, and giving presentations. A large array of digitized Pennsylvania materials available on the Web will call attention to the key role Pennsylvania has played in many histories, the many histories which, in turn, will allow Pennsylvanians to gain a richer sense of the Commonwealth's heritage. Like other segments, this group is also likely to want to travel to Pennsylvania's libraries and archives to view the materials in their original.

**Heritage tourism development.** Heritage tourism has proven itself as a significant economic driver in Pennsylvania. Digitized Pennsylvania materials serve the heritage tourism sector in two ways: by providing the raw material from which tourism professionals develop public programming; and by providing history travelers, who prefer to research destinations and topics as part of their trip planning, with additional reasons to visit or to extend their stays.

**Funders.** Through the work of the PACCD, subject areas and sources are identified, ranked by importance, and linked to the needs of Pennsylvania's diverse user communities. With this information at their disposal, both funding agencies and libraries will be better able to plan and set digitization priorities. Having these guidelines will also enhance the possibilities for inter-institutional collaboration as cultural collections repositories work together to assemble critical masses of materials. Without precluding the development or funding of digitization projects which do not fit the guidelines, the guidelines will help to focus the attention of funders and libraries on those materials that would provide the most utility for historical and personal study by the groups listed above.

### **3. Topics and Material Types**

With today's technology, a wide variety of materials can be digitized for wider access. The panel of experts and others interviewed about primary historical sources highlighted many types of materials that they, their colleagues, researchers, family historians, and students use in their study of Pennsylvania's history and current affairs.

We propose a tiered approach to digitization of materials types and topics/periods. In the first tier are four "priority" topics. We advise that collecting institutions and funders give top priority to them, and we prioritize them in order to help focus the community on creating a critical mass of materials. Within material types, we recommend that priority be given to materials with high research value that are also at risk for significant degradation. Recognizing that a much wider

range of materials is desirable for digitization, we also suggest "general" topics and materials for second priority consideration.

We encourage institutions and funders to develop partnerships and cooperative projects that will devote resources to these priority materials and topics for the near future. Our listing these topics by no means should be understood as discouraging institutions from digitizing other materials based on the strength of their collection in a particular area, a local need, or an upcoming event, such as an anniversary which may focus widespread public interest on a topic. Ideally, these materials would be related to the four themes to which we have given priority, but when an institution finds that it does not have significant holdings related to these four themes, it should meet institutional and constituency needs as seems best.

These listings by no means should be understood as discouraging institutions from digitizing other materials based on the strength of their collection in a particular area, a local need, or an upcoming event, such as an anniversary which may focus widespread public interest on a topic. Ideally, these materials would be related to the four themes to which we have given priority, but when an institution finds that it does not have significant holdings related to these four themes, it should meet institutional and constituency needs as seems best.

### *Priority Topics*

The panel assumes that interest in digitizing material will consider its historical significance. All history, like politics, is local, and we recommend that collecting institutions think first about family, neighborhood, municipal, and county-level materials that illustrate themes, periods, or events of national and world interest. Collecting institutions should look for local documentation that will contribute to the study of the broader issues and events we have listed below. This group also encourages the inclusion of state birth and death records under access guidelines similar to those followed by states that have chosen to make these records available online.

Pennsylvania was and is remarkably rich in agricultural, mineral, and energy resources, which have drawn successive immigrant populations to pursue occupations in these fields, as well as to a range of manufacturing industries in the nineteenth and twentieth centuries. The people, industries, resources, and related transportation and communication systems have both shaped and been shaped by the great conflicts of those centuries, notably the American Civil War, World War I and World War II. These key subjects in the history of the Commonwealth of Pennsylvania can provide users with both a chronological view of the development of the state and a glimpse into how Pennsylvanians have dealt with progress and change. The interplay of these forces is vital to the understanding of our shared history. Throughout Pennsylvania's history, there have also been communities of conscience that reacted to war and injustice in principled and organized ways, and those reactions form part of the overarching narrative of war and conflict.

We recommend that among the topics and periods listed below, collecting institutions and funders focus specifically on four priority topics in order to build a critical mass of material:

- **The environment and land use**
- **Population groups from 1850 to the present**
- **Transportation**
- **Major nineteenth- and twentieth-century conflicts--the American Civil War and World Wars I and II, for example--and the communities of conscience that have reacted to war and injustice**

### *General Topics of Interest*

Key subjects in the history of the Commonwealth of Pennsylvania can provide users with both a chronological history of the development of the state, and a glimpse into how Pennsylvanians have dealt with progress and change. Topics for special focus in digitization, and related subtopics, include:

- **The Economy:** Digitizing materials from the 1600s to the present, related to important economic and labor trends including agriculture, commerce, and manufacturing, can provide information on the growth of the Commonwealth.
- **Politics and Government:** From the first elections held in the State to present local, state, and national elections, these materials can provide insight on how Pennsylvanians govern themselves, and their relationship with the outside world.
- **Society:** Race, gender, Native American populations, religion and social justice, slavery and the underground railroad and their place in determining the course of Pennsylvania history, along with the effect of battles and wars on the people of Pennsylvania from the earliest settlers to the present.
- **Technical, Scientific, and Medical topics:** Throughout the history of the Commonwealth, and especially from 1850 on, breakthroughs made in technology, innovation, and invention in Pennsylvania have been important to the nation and the world.
- **Culture:** From built spaces to fine arts to pop culture, especially in the 20th Century, Pennsylvanians have been cultural trendsetters.

### *Material Types*

In our discussions and interviews we heard about the many kinds of library and archival materials that can be digitized to meet user needs--published works of local history; maps (especially insurance maps); print and photographic collections; postcard collections; newspapers; diaries and letters and other personal/family documents; biographical information; business records; arts magazines and architectural sourcebooks; materials in library vertical files including primary resources and press kits; university histories and course catalogs; tax records, wills, death records, court records; city directories and telephone books; records of churches and other local organizations and groups; naturalization records; lists of almshouse residents; Pennsylvania German materials; deeds; Pennsylvania Archives; and census records.

We recommend that significant attention be given to historical sound and moving image documents, most of which will be from the 20th Century. Documentarians, scholars, and students all know the value of being able to locate, use, and incorporate into their work evidence of the sort provided by radio and television programs about local issues and history, industrial films, speeches by politicians, recordings of the dedications of buildings, etc. That historical sound and moving image materials are fragile and are decaying rapidly is amply documented; more worrisome still, most such materials are not even collected at this point into repositories, let alone preserved for future access.

While the urgency attached to the collection and preservation of audiovisual materials is very real because of the threat to their survival, we recognize that digitizing them poses some formidable

barriers to entry in terms of expertise, equipment, and cost. Copyright is another issue for many 20th century materials, and, once digitized, these materials require major investments in infrastructure for storage, preservation, and ongoing access.

In cases where digitization may not be practical at this time, we urge Pennsylvania libraries to start collecting these materials from the private, commercial, and governmental sources that hold them and to survey and prioritize them as an important first step in a statewide conversation about preserving this threatened part of our heritage. Since few institutions will have the staff expertise and equipment to digitize these materials in house, we hope, moreover, that our recommendation of this priority helps to leverage the attention of all parties toward taking a collaborative approach to solving the technical problems of file creation, storage, and ongoing access.

We also suggest that prioritization take into account the funding programs already in place from public and commercial organizations for digitization of such materials as newspapers, diaries, printed works, and census records. We recommend, therefore, that, after they consider digitization of historical sound and moving image materials, collecting institutions focus on kinds of primary sources that are outside the purview of current initiatives for newspapers, diaries, printed works, and census records. Leading examples of a material type of particular interest to many of the audiences in Pennsylvania are vital records, including birth and death records and land records. Preserving the access to these important materials, which were initially published in paper form, is an important goal for digitizing the primary sources on which historical research is based.

With a plan that focuses on high-risk formats and important topics in the growth of the Commonwealth, Pennsylvania cultural heritage institutions can develop a digital presence that truly represents the history and important contributions of the state.

#### **4. Best practices and technical specifications for digitization**

Libraries, museums, and cultural heritage institutions planning to digitize their collections should follow well-regarded and widely recommended best practices and technical specifications, developed by such organizations as the Pennsylvania Office of Commonwealth Libraries, the Open Archives Initiative (OAI), and the Digital Library Federation (DLF), among others. By following these standards and practices, institutions will create high-quality images and metadata, which will make their digital resources easily searchable and accessible to the greatest possible audience.

##### **Getting started**

The Pennsylvania Advisory Committee on Collaborative Digitization (PACCD) and its Metadata Working Group have created guidelines for the digitization of collections for, by, and about Pennsylvania's history, culture, and society. These guidelines are available via the Access Pennsylvania Digital Repository site at [http://www.accesspadigital.org/pdf/access\\_pa\\_dig\\_rep\\_guidelines.pdf](http://www.accesspadigital.org/pdf/access_pa_dig_rep_guidelines.pdf).

The PACCD guidelines are designed to improve the management and preservation of the Commonwealth's digital resources, as well as to ensure preservation of the physical objects upon which the digital resources are based.

The PACCD guidelines provide several sources for information on optimum scanning and digitization practices for a variety of material formats—audio, video, text, et al.—recommended by such organizations as the Digital Library Federation (DLF) (<http://www.diglib.org/>) and the Collaborative Digitization Program of the Bibliographic Center for Research (BCR) (<http://www.cdpheritage.org/>).

### **Audiovisual resources**

In this guide, historical sound and moving image resources are one of the material types designated as a priority for collection, preservation, and digitization. While all digitization projects pose challenges, as noted previously, the digitization of audiovisual resources may offer formidable barriers to success to cultural heritage organizations.

Detailed and reliable information on audio digitization can be found in a number of sources. These include *Digital Audio Best Practices* (<http://www.bcr.org/cdp/best/digital-audio-bp.pdf>), prepared by the Colorado Digitization Program (CDP) Working Group, and *Recommended Minimum Standards for Preservation Sampling of Moving Image Objects* ([http://rucore.libraries.rutgers.edu/collab/ref/dos\\_avwg\\_video\\_obj\\_standard.pdf](http://rucore.libraries.rutgers.edu/collab/ref/dos_avwg_video_obj_standard.pdf)), published by RUCORE, the Rutgers (University) Community Repository.

### **Standardized metadata and access**

The PACCD guidelines recommend the use of Dublin Core metadata to catalog and describe digital resources and provide detailed specifications on the proper treatment of descriptive fields.

Through these guidelines, institutions have created the digital resources that constitute the Access Pennsylvania Digital Repository (<http://www.accesspadr.org>), a statewide planning effort to create a unified approach to digitization throughout the Commonwealth. As part of this effort, the Office of Commonwealth Libraries' Access Pennsylvania initiative has purchased OCLC's CONTENTdm software on a statewide license to provide to any library that would like to do so the ability to store their digital content on the Access Pennsylvania server. This is a physical way to make digital resources available, allowing for both storage of resources and their retrieval by scholars.

Collections included in the Pennsylvania Office of Commonwealth Libraries' Access Pennsylvania's Digital Repository also meet the standards for interoperability established by the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), details on which are available at <http://www.openarchives.org>. Thus, as part of their adherence to best practices and standards, libraries, museums, and cultural heritage organizations in Pennsylvania should follow these protocols as well in their digitization efforts.

## The Pennsylvania Digital Library

As a result of following these protocols, digital collections for, by, and about Pennsylvania can then be harvested by and searched through the Pennsylvania Digital Library (<http://padl.pitt.edu>). Developed by the University of Pittsburgh's Library System, under the auspices of the PACCD, the Pennsylvania Digital Library (PADL) is a state-wide metadata repository for non-subscription, non-commercial digital resources created by Pennsylvania libraries, museums, and cultural heritage institutions.

Rather than serving as a repository for digital collections themselves, the PADL harvests, indexes, and provides a common search interface for metadata created for Pennsylvania digital collections. Thus, collections reside at their home institutions (or the Access Pennsylvania Digital Repository) while metadata adhering to OAI-PMH standards is gathered regularly and then made searchable through the PADL.

Institutions desirous of having their collections searchable through the PADL and meeting OAI-PMH standards should register with the PADL (<http://padl.pitt.edu/index.php/register>) so that their collections' metadata will be harvested and made available through this common search interface. Institutions that are part of the Access Pennsylvania Digital Repository do not need to register with the PADL as their collections' metadata is automatically harvested.

## Copyright

Copyright expert Peter B. Hirtle writes in *Digital Preservation and Copyright* ([http://fairuse.stanford.edu/commentary\\_and\\_analysis/2003\\_11\\_hirtle.html](http://fairuse.stanford.edu/commentary_and_analysis/2003_11_hirtle.html)) that

*Digital preservation and access is all about copying. In copyright law, copying is known as "reproduction," and it's one of the exclusive rights of the copyright owner. The right to publicly display a work is also an exclusive right of the copyright owner, as is the right to make an adaptation, known as a "derivative work." Our desire to keep digital information around for the future runs smack into the exclusive rights of the copyright owner.*

Thus, the process of digitization—transforming a work or object from one form, chiefly analog, into an electronic or digital format—is a form of reproduction. Therefore, national and international copyright laws and institutional copyright policies must be considered when digitizing collections and resources with the aim of sharing them with a larger world.

Many historical materials, particularly those created before the 20<sup>th</sup> century, may be in the public domain, outside of copyright, and suitable for digitization. However, copyright may still be held for more recent materials, especially audiovisual resources.

As organizations plan and carry out digitization efforts, resources such as the OCLC Digitization & Preservation Online Resource Center's *Copyright Online Resource Kit* and the U.S. government-sponsored Section 108 Study Group's "Background Papers" website (<http://www.section108.gov/papers.html>) can assist with making informed decisions about copyright law and common practices.

## **Accessible design**

Providing rich scholarly content to a wide audience is one of the main goals of digitization. However, that content may be made unusable to many scholars unless content creators pay attention to technical compatibility, ease of use, and issues in accessible design.

For those organizations that design their own websites and portals to provide gateways to digital collections, it is important to follow good practices in website design. Designs that are clean, uncluttered, easy to search, and compatible with different web browsers are preferred as they make collections accessible to the greatest number of scholars working at all levels of interest. Another important consideration for portal design is accessibility to persons with different abilities and disabilities, including visual, auditory, intellectual/cognitive, and mobility impairment.

The Worldwide Web Consortium (W3C) (<http://www.w3.org/>) is the main international standards organization for the worldwide web. One project of the W3C is the Web Accessibility Initiative (WAI) (<http://www.w3.org/WAI/>), which creates Web Content Accessibility Guidelines (WCAG) in order to make webpages and web technologies accessible to all. Additionally, Section 508 of the U.S. Rehabilitation Act (29 U.S.C. ' 794d) (<http://www.section508.gov>) requires U.S. government agencies to give disabled employees and members of the public access to information that is comparable to the access available to others. While there are currently no requirements for accessible website or portal design under the Americans with Disabilities Act (ADA), design that is accessible to as many scholars as possible should guide website development.

## **Equipment, training, and funding**

Although the rewards to scholarship and cultural heritage can be immense, both locally and globally, digitizing collections is no small undertaking. In addition to standards and guidelines for digitizing and cataloging collections, often equipment must be purchased, staff must be trained, and funding must be secured to pay for the work to be done.

Through its digitization and preservation program ([http://www.palinet.org/digitization\\_preservation\\_services.aspx](http://www.palinet.org/digitization_preservation_services.aspx)), PALINET, the regional library network for Pennsylvania, provides opportunities for training and continuing education, information on vendors who sell digitization equipment or offer digitization services, and possible sources for funding. Other organizations, such as the Pennsylvania Historical and Museum Commission (PHMC), offer similar direction and information for museums, historical societies, and cultural heritage organizations.

National, regional, and state professional associations for librarians, archivists, and museum specialists provide opportunities for networking and learning from colleagues who have developed digital collections or operate digitization programs. Nationally, these include

- American Association for State and Local History (AASLH) (<http://www.aaslh.org>)
- American Library Association (ALA) (<http://www.ala.org>)
- Association of American Museums (AAM) (<http://www.aam-us.org/>)
- Association of College and Research Libraries (ACRL) (<http://www.acrl.org>)
- Digital Library Federation (DLF) (<http://www.diglib.org/>)
- Society of American Archivists (SAA) (<http://www.archivists.org/>)

Regionally and locally, these include

- Pennsylvania Academic Library Consortium, Inc. (PALCI) (<http://www.palci.org>)
- Pennsylvania Library Association (PaLA) (<http://www.palibraries.org>)
- Pennsylvania Federation of Museums and Historical Organizations (PFMHO) (<http://pamuseums.org>)
- Philadelphia Area Consortium of Special Collections Libraries (PACSCL) (<http://www.pacscl.org/>)
- Mid-Atlantic Regional Archives Conference (MARAC) (<http://www.lib.umd.edu/MARAC/>)

## 5. The future

In thinking about the many aspects of collection access and especially about the environment for digitization, the Pennsylvania Advisory Committee on Collaborative Digitization (PACCD) sees several related steps that might be taken to promote cooperation in the development of a digital library about Pennsylvania. Many workflows, organizations, funding streams, and kinds of expertise come into play in collection digitization, and the PACCD represents a strong initial effort at interconsortial cooperation to build the technical, collection development, and administrative network a successful, coherent digital collection for Pennsylvania will require. We therefore want to note here the many dimensions of the digitization picture in Pennsylvania and suggest that they constitute one picture rather than a number of separate sketches.

1. As noted above, our panel recommends that the guidelines document and its priorities be reviewed every three years in order to 1) keep it fresh by aligning the topics and material types with current interests and, perhaps more important, 2) create and maintain an ongoing community of interest in the digital collection about Pennsylvania. The Bureau of Library Development might look to the PACCD for this task. In the immediate future, representatives of the LSTA grant planning group or panel of experts can take this document to professional association meetings in order to create interest in and seek comment on the guidelines and, more generally, the digitization of Pennsylvania materials. Since it is frequently the case that pockets of activity will develop and remain isolated from each other, the planning group hope that the PACCD website being developed by PALINET and partners can become a community-building hub for collecting information about digitization activities in the Commonwealth.

2. In many cases, statewide digitization projects have employed a survey to assess the amount and kind of material available about the state. In 1998 Pennsylvania participated in a study funded by NHPRC that inventoried document repositories through what is now the Council of State Archivists. The report is posted on the CoSA web site at <http://www.statearchivists.org/reports/index.htm>. Since that inventory was taken, collecting institutions and consortia have created many digital collections and are planning more. Possible parameters for a survey could be questions about:

- *Digital projects already in existence*
- *Cataloged but undigitized collections*
- *Cataloged collections the holder has no plans to digitize*
- *Uncataloged collections that are worth digitization*

3. Given the wonderful incentive offered by the HSLC/AccessPA CONTENTdm license, develop the user community for it by offering workshops on digitization and by encouraging the CONTENTdm group being fostered by PALINET.
4. Promote participation by collecting institutions in and work with the results of the Connecting to Collections grant to the Conservation Center for Art and Historic Artifacts from the Institute of Museum & Library Services; the grant will provide data that will look to digitization of materials for reasons of preservation as well as access.
5. Use the opportunity offered by the Sloan Foundation mass-digitization grant to PALINET to digitize printed materials about Pennsylvania.
6. Build on the findings of the current PACSCL hidden collections grant from The Andrew W. Mellon Foundation (<http://www.pacsclsurvey.org/>) to select materials for digitization; encourage partnerships in Pennsylvania to respond to the invitation from the Council for Library and Information Resources for hidden collections grant proposals in the next five years.
7. Develop infrastructure that will help small institutions contribute materials for digitization through partnerships with larger institutions or by provision to them of expertise and digitization services by a central agency.
8. Hold a statewide invitational meeting on the desiderata for digitization based on the guidelines document and under the auspices of PACCD with representation from State Library/Archives/PHMC members, PA History Association, PACSCL, PALINET, school curriculum people and teachers, etc. Perhaps through an LSTA grant, especially try to bring small institutions into this discussion.
9. Given the well-documented challenges to collecting, digitizing, serving, storing, and preserving historic sound and moving image materials, convene a statewide meeting under the auspices of PACCD, in partnership with Commonwealth agencies and groups representing collecting institutions, to discuss how to proceed with addressing this priority of the collection guidelines.
10. Work with organizations to make records for (digital) collections harvestable into an OAI registry.
11. Encourage Pennsylvania collecting institutions to put their finding aids on the Web so that researchers can explore collections more easily and thereby suggest priorities for digitization.

In this document, the Pennsylvania Advisory Committee on Collaborative Digitization has laid out not only a set of topical guidelines but also a mechanism for ongoing updates to these digitization priorities, together with a list of considerations for, and links to information about, issues related to planning and digitization initiatives. The committee is hopeful that, once these guidelines have been through the review and comment process, they will provide a solid foundation on which Pennsylvania can build a digital library and a reputation for leadership in statewide collaboration.

*(Revised July 30, 2008)*