LSTA Indiana Digital Library Digitization Grants
2006-2007
Overview and Guidelines

1. What is this grant program and its purpose?

In this project, the Indiana State Library will offer sub-grants to Indiana libraries and their partnering organizations to digitize unique historical materials of interest to people throughout the State of Indiana. Materials will be digitized using accepted national standards as identified by the Indiana Digital Library steering group for image creation and for metadata tagging, and will serve as the cornerstone for the development of an Indiana Digital Library.

A variety of materials can be digitized, focusing on materials of use and interest to a statewide audience, with preference given to those materials of use by the educational community. These materials will be made available over the Internet to a broad audience, without restrictions on use or availability. Preference will also be given to collaborative projects, including those in which a library partners with a museum, local cultural organization, or school.

Using accepted national standards for the creation of digital images, Indiana libraries will identify and digitize unique historical resources from their collections and those of their partnering organizations. Recipients of sub-grants will be encouraged to work with libraries with expanded digitization facilities or with appropriate vendors when outsourcing is appropriate. Recipients will also be required to post information about the standards and methods used in creating the information.

For those interested in applying for a LSTA Digitization Mini-grant for the first time, the Indiana State Library will host an introductory LSTA Digitization Mini-grant Workshop in the Indiana Authors Room on Tuesday, February 7, from 9 a.m. to noon. A second workshop building on the basic morning session will be held that afternoon from 1 p.m. to 4 p.m. The afternoon session is geared toward those who have already been through the LSTA application process and would like a more advanced overview of this year's requirements. Attendees from the morning session are welcome to stay for the afternoon workshop. Both workshops will be led by Kris Brancolini and Jenn Riley from Indiana University, who will provide attendees with information about building good digital collections, new requirements for LSTA Digitization Mini-grants, and resources to help with project definition, planning, implementation, and evaluation. Please contact the Library Development Office by email at ldo@statelib.lib.in.us to sign up for either or both of these free workshops.

2. What is the application process?

Libraries must submit two copies of the application to the Library Development Office by the deadline of March 31, 2006, 4 p.m. The application may not be faxed but should be mailed. Please note --The application must be in the Library Development Office by the deadline. Postmark is not acceptable.

3. Who may apply?

The following types of libraries are eligible to apply for this grant program:

- Public libraries
- Academic libraries
- School media centers

Libraries are not required to partner with another library or cultural institution on a project; however, collaborative projects will be scored higher than single-institution applications. Collaborative projects should draw upon the participating institutions' strengths (complementary collections, skills, and other resources) and should not be viewed solely as a way to share digitization equipment for use at two institutions.

The lead project agent must be a library, and is responsible for managing the project, acting as fiscal agent, maintaining required paperwork and records, successfully implementing the project, and completing all reports.

> **These LSTA grants can only be awarded to libraries, and a library must serve as lead project agent in a collaborative endeavor. Libraries are encouraged to collaborate with non-library cultural institutions in their proposed grant projects. Special libraries, museums, historical societies or educational institutions should apply in collaboration with a public, academic, or school library.**

Other cultural institutions include, but are not limited to, Indiana archives, historical societies, and museums.

4. What amounts may be requested?

There is no minimum amount that may be requested. The maximum amount that can be requested by a lead project agent is $40,000.

5. How may the project funds be used?

The Indiana Digital Library digitization grants are intended to expand access to the unique cultural heritage of Indiana using the Internet and digital technologies. This includes digital imaging of primary research materials (manuscripts, photographs, broadsides, pamphlets, rare books, and similar materials) for unlimited access via the Internet.

<u>Allowable expenses</u>
- Equipment, including computers, scanners, printers, and computer peripherals (See CIPA information below)
- Software for imaging
- Supplies for project (CDs, photo paper, floppy disks)
- Training for participating staff
- Project staff (wages and benefits of temporary employees or temporary expansion of part-time staff)
- Metadata creation

<u>Outsourcing</u>

Funds from this program may be used to support outsourcing of digitization and encoding projects. The grant program seeks to build capacity and skills within Indiana; applicants are encouraged to work with other Indiana cultural institutions with the capacity to handle outsourced projects.

<u>CIPA Statement</u>

Use of LSTA funds for certain allowable purchases may require public libraries or public elementary and secondary school libraries to comply with the federal Children's Internet Protection Act (CIPA). For additional information, see: http://www.statelib.lib.in.us/www/isl/ldo/cipamenu.html

<u>Ineligible costs</u>

- Ongoing operating costs
- Costs for traditional arrangement, description, and preservation
- Wages and benefits for existing full-time employees
- Entertainment costs
- Overhead or indirect costs

Funds may be used for the continuation of an existing project, so long as it meets the selection criteria. Projects must conform to the identified standards for digital object creation and metadata creation, and have a long-range plan for ongoing access to the information and migration of the data.

6. What is the basis for selecting projects for funding?

Projects must include an eligible institution as the primary partner in the project.

The project **must** include the following required criteria:
- Project conforms to identified standards for metadata and digital data creation. See Appendix A (Indiana Digital Library Metadata Best Practices for Use of Qualified Dublin Core) and Appendix B (Digital Imaging Standards and Best Practices) for required standards. Please find other useful best practice guidelines

on the Indiana Digital Library information page at
http://www.statelib.lib.in.us/www/isl/diglibin/index.html.

- Project plan includes sustainability component, describing long-range plans for ongoing access to the data and migration of the data. General tools and resources for digital project planning and implementation may be found on the Indiana Digital Library information page at
http://www.statelib.lib.in.us/www/isl/diglibin/index.html.
- Project staff certifies that they have the right under existing copyright law to create and make available to the public digital copies of the materials identified in their grant.

Priority will be given to projects exhibiting the following criteria:
- Project involves Indiana subject matter identified as high priority for Indiana Digital Library:
  - The Indiana Digital Library Content Advisory Workshop was held in the fall of 2005 with participants from the educational, academic, and museum communities. The theme of "personal narrative, life in Indiana" was selected to narrow the focus for this year's LSTA grant-funded projects and helped to create an overall starting point for the digital library. This theme can be further developed by asking "Does it identify Hoosier traits?" and "Does it inform our understanding of Indiana history?"
  - The following content areas targeted during the 2005-2006 LSTA grant program will also be given preference in 2006-2007:
    - Famous Hoosiers -- regardless of time period, particularly scientists, inventors, artists, religious leaders. Also needed information on politicians, soldiers, businessmen, women, Native Americans, and African Americans.
    - Underground Railroad
    - Native Americans of Indiana
    - Transportation -- regardless of time period, including canals, railroads, National Road
    - Civil War -- all aspects
    - Communication -- regardless of time period, including early newspapers
    - Architecture -- regardless of time period
    - Indiana in World War I
    - Indiana in World War II
- Project involves a defined collaboration between a library and another partner or partners (museum, archives, historical society, or educational institution):
  - If your library has previous digital project experience, use Vigo County Public Library's 2005-2006 LSTA grant application as an example. Their proposed project was in the scope of their previous digitization experience, so collaboration was not necessary to ensure funding. See Appendix C.
  - If your library does not have previous digital project experience, use the Jasper County Public Library's 2005-2006 LSTA grant application as an

example. They partnered with Indiana University, an organization with ample digital project experience, to make their LSTA application more viable. See Appendix D.

- Project applicants provide matching funds for personnel, hardware, software, or any other project expenses.

7. What is the review process?

All grants must be received by 4:00 p.m. March 31, 2006.

Only complete grant applications from eligible institutions will be reviewed.

Eligible applications will be scored, and then reviewed by Library Development Office staff and external reviewers according to the grant program criteria. In addition, applications will be reviewed for clarity and quality of plan, and quality and quantity of content to be provided by the proposed project.

Announcement of funded applications will be made the first week of May 2006 on the Indiana State Library's Web site and by letter to the applicants. The information will also be posted to appropriate electronic discussion lists and included in the Library's newsletter.

8. How are grant payments made?

All grants are reimbursement grants. Payments will be made monthly on receipt of a reimbursement form accompanied by copies of invoices.

9. What reports must grantees make?

Grantees will be required to make quarterly progress reports, plus a final report evaluating the project. A final fiscal report is also required.

In addition to the standard reports required by LSTA, participating libraries and their partners will report the number of items digitized and the types of collections digitized. They will also report what standards were used in developing resources, and post those standards as part of the digital resource.

Participants will report the number of "hits" on their individual Web sites; the Indiana State Library will report the number of "hits" on the portal site linking these individual projects. Participating libraries will also be encouraged to provide the number of "hits" on individual resources.

10. What else do applicants need to know?

Projects funded through this grant program will be identified as part of the Indiana Digital Library. Projects funded during this fiscal year will become the cornerstone of

the Indiana Digital Library, designed to provide electronic access to the unique cultural resources held by Indiana libraries and allied cultural institutions.

11. Grant program timeline:

- Application posted   January 6, 2006
- Half-day workshops for grant applicants February 7, 9 a.m.-12 p.m. (Beginner) and 1 p.m. – 4 p.m. (Advanced) at the State Library
- Applications due   4:00 PM Friday, March 31, 2006
- Review Period   April 2006
- Applicants notified of status   First week of May 2006
- Contracts prepared and signed   May and early June 2006
- Grant Period   June 30, 2006—June 30, 2007

12. Where do I apply?

The application will be posted on the Indiana State Library Web site at http://www.statelib.lib.in.us/www/isl/ldo/lsta/lstadmg.html.

13. For further information:

For examples of successful grant applications and other helpful information, please visit the Library Development Office Web page at http://www.statelib.lib.in.us/www/isl/ldo/lsta/lstamenu.html.

Questions may be directed to the following Indiana State Library staff:
- For more information about prospective project partners and collaboration resources, contact Jacob Speer at (317) 232-3715 / jspeer@statelib.lib.in.us.
- For questions concerning grant administration, contact Rose Marie Kelsey-Traylor at (317) 232-3694 / rkelsey@statelib.lib.in.us.
- For questions concerning project content, contact Shannon Bloomquist at sbloomquist@statelib.lib.in.us.

Digitization Planning Worksheet

Single-page objects

	Quantity (be specific)	Dimensions	Color or black & white?	Characteristics likely to slow down digitization		
				2-sided	Fragile	Items in group not homogeneous
1						
2						
3						
4						
5						
6						
7						

Multi-page objects

	Quantity (be specific)	Average number of pages per item	Dimensions	Color or black & white?	Characteristics likely to slow down digitization			
					2-sided	Fragile	Items in group not homogeneous	Pages bound together
1								
2								
3								
4								
5								
6								
7								

Metadata Planning Worksheet

	Quantity (be specific)	Published or unpublished?	Existing unique identifier for each item?	Material in a foreign language?	If there is existing description		
					Machine readable?	Divided into fields?	Complete?
1							
2							
3							
4							
5							
6							
7							

Metadata Guidelines for the Indiana Digital Library

1. What is the Indiana Digital Library?
The Indiana Digital Library (IDL) is a resource that provides online access to the cultural and historical resources of the state of Indiana. The Indiana Digital Library is a gateway to Indiana's history and culture found in digitized books, manuscripts, photographs, newspapers, audio, video, and other materials. This digital library is made possible through the collaborative efforts of Indiana's libraries, archives, museums, and other institutions of cultural heritage to develop, maintain, and preserve digital collections and online digital resources. These digital objects will be created, maintained, delivered, and preserved according to national standards.

2. Why do we need to follow standards?
The IDL offers cultural institutions in the state of Indiana an opportunity to collaborate in ways that were never before possible. Collections that are scattered across institutions can now be brought together in the digital environment; unknown or seldom used collections can be brought to light; connections between subjects covered in a variety of formats or between materials held by a variety of institutions can be discovered. These are only a few examples and yet they offer a glimpse of the exciting possibilities ahead.
The primary objective of the Indiana Digital Library is improved access to the unique resources and special collections that have been or will be converted into digital format in cultural heritage institutions throughout the state of Indiana. One way to accomplish this goal is by bringing information about all of these diverse and scattered resources together into a single portal of access. To be able to collaborate at this level we need to be able to share our information. This is why standards are so important. Standards provide the framework for sharing information among institutions and across networks. The adoption of standards is necessary for effective sharing of resources and institutional interoperability.
To improve access to these materials it is not enough to simply convert them into digital format and make them available on the Web. Access requires information about the material. We refer to this information as metadata. Describing a resource is a difficult process but an important one if we truly want our state's unique resources to be available and accessible to the world. The more we adhere to uniform practices, the more likely these resources will be found and used.

3. What is metadata?
Metadata is simply structured information about an information resource. In the broadest sense, metadata labels your information. The reason for labeling your information in this way is to then use this labeled, or structured, information for access, administration, management, and preservation of these resources. Metadata provides the necessary tools to manage, preserve, and provide access to information in the digital environment. The creation of metadata is governed by a body of standards, best practices, and schemas that, when appropriately applied, work together to facilitate the above tasks.

There are a variety of types of metadata:

Descriptive metadata: information used for indexing, discovery, identification of digital resources, and access

Technical metadata: provides information about the scanning process – resolution, hardware/software, compression, etc.

Structural metadata: information used to display and navigate complex digital resources

Administrative metadata: provides management information such as how to access and display the resource as well as rights management

Preservation metadata: includes information such as the change history of a resource and detailed technical information useful for management of resources within a digital archive

This information or metadata is necessary for discovery of relevant materials. A digital object with no associated information can only be browsed, but this same digital object with an associated metadata record now has a title, description, keywords or subjects that can be searched. Likewise, if all members of the IDL agree to follow certain metadata standards then our records can be accurately searched and shared. The IDL recommends metadata standards to choose from in the document "Choosing a Metadata Standard for Your Digital Project." These metadata standards provide elements used to describe resources, include explanations for the type of information to include in each element, and guidelines on formatting the information. One of the most widely used metadata standards is the Dublin Core.

4. What is the Dublin Core?
Dublin Core is an open, global standard designed primarily to support discovery and retrieval of digital resources. The Dublin Core metadata standard is a set of fifteen elements and optional qualifiers that can be used to describe a variety of digital resources. The Dublin Core was intentionally created to be simple to use and understand. This allows a non-specialist to create descriptive records for digital resources easily and efficiently. The terminology used in Dublin Core is universally understood and generic enough to be applicable to a variety of disciplines and formats. The Dublin Core elements are defined, but usage standards are left to the individual groups implementing the standard.
Dublin Core also allows for extensibility. What this means is that these core elements can be added to and built upon to meet the needs of the creating organization; whether this means providing greater descriptive detail, domain specific information, or information to support preservation activities. Dublin Core is meant to be a general standard that will coexist with richer standards.
The IDL intends to ensure as broad a range of access to these materials as possible by participating in the Open Archives Initiative (OAI). OAI is a protocol for sharing information by making metadata open to harvesting. In order for OAI to function effectively, the harvested records must follow the same standards and employ a set of common elements. OAI requires the use of Unqualified Dublin Core to achieve this goal of having common elements to harvest, and allows more

robust metadata formats to be exposed as well. The metadata records, not the actual digital items, are then compiled from all participating organizations worldwide into a single, searchable interface. By providing crosswalks to Dublin Core, the IDL will open all collections to OAI harvesting regardless of the native metadata standard used for the collection.

The IDL includes Dublin Core as an option for a metadata standard because it is already in use in many institutions within the library, archival, and museum communities. It is a simple set of elements designed to be easy to implement. This simplicity, however, limits how robust the metadata can be. For guidance in using Dublin Core, please refer to the companion document "Indiana Digital Library Metadata Best Practices For Use of Qualified Dublin Core."

We understand that not every institution has the financial resources, staff, or technological expertise necessary to implement a full-fledged metadata program. By providing best practices for institutions who choose to implement Dublin Core, we hope to make participation in the IDL an option for every institution across the state. These best practices are intended to provide everyone with the information they need to create metadata records with confidence regardless of whether the records are created by professional catalogers, library staff, student workers, or volunteers.

Key Points About Dublin Core:
Simple and easy • • • • •
Accepted standard [ANSI/NISO Z29.85-2001]
Requires minimal training
Easily adapted for local circumstances
Enables record harvesting

5. Why might I use a standard other than Dublin Core?
Dublin Core metadata elements provide very basic information about the resources they describe, and therefore serve as the basis for interoperability. However, institutions that want to provide more robust description, such as those with collections in a special field or those serving specialized audiences, will likely require more access points and specialized information in their metadata. For more examples, refer to "Choosing a Metadata Standard for Your Digital Project."

When choosing a metadata schema, use what works for your kind of material. Keep in mind that you have the option to use more than one schema. You may choose to use the core set and augment that with fields from other standard metadata sets or you may develop local fields to capture data in a meaningful way for your institution. Whatever you end up implementing, always remember that the goal is for your metadata to be transformable into something that will interoperate with metadata from other IDL institutions. Remember too that you always have the option to change the metadata. Start with a small set of digital items. Try the metadata elements you have selected and see if they work and meet your needs. If not, make changes.

6. If I choose a standard other than Dublin Core, will my collection still be part of the IDL?
You may choose any standard that works for your collections and still contribute your metadata to the IDL. This is possible because of the use of crosswalks. Dublin Core has value as a means for cross walking, or mapping, between richer, more complex metadata standards. In essence, a crosswalk is a table that maps the relationships and equivalencies between two or more metadata standards. This in turn allows search engines to effectively search across heterogeneous databases. In order to be able to search across collections built around various metadata standards, there needs to be a way to translate the data into a shared language. Dublin Core provides this shared language. Because of this ability to crosswalk from other metadata standards to Dublin Core, participants in the IDL are able to implement or continue using other metadata standards, such as MARC, Encoded Archival Description (EAD), Text Encoding Initiative (TEI), or whichever metadata standard works best for their collection.
The IDL will create and maintain base crosswalks that translate between the various metadata standards that it recommends. Institutions using a locally developed metadata format or a standard not covered by the IDL crosswalks, have the option to create their own crosswalks. The IDL will consult with these institutions to minimize the loss of information during the transformation process and maximize interoperability. Other institutions need only ensure that the metadata standard they choose is among the standards supported by the IDL crosswalks.

7. How do I participate?
If your institution has materials that they would like to share by creating metadata records to contribute to the IDL, one of the most important steps in working towards that goal is project planning. An important aspect of project planning is deciding which metadata elements or schema to adopt. You will want to start with the recommendations in this guide, but always keep in mind that Dublin Core is a general standard that may not address all of your needs. It is certainly acceptable to create records using only the fields recommended in this guide if that is all your materials warrant or if that is all your resources will allow. However, we want to encourage you to look beyond Dublin Core and explore other metadata standards in order to provide the richest, most detailed records that you can. For further guidance please refer to the IDL document "Choosing a Metadata Standard for Your Digital Project."
Implementing a metadata schema for a digital project involves several steps:

1. Identify your users and materials
a. Determine the type of materials that will be digitized
b. Determine your potential users and audience
c. Determine the number and complexity of digital objects involved

2. Define the purpose of the digitizing project

a. Determine whether you are digitizing for access only or whether you will also digitize for preservation.
b. If you plan to collaborate with other organizations, interoperability and standardization must be taken into account

3. Brainstorm what elements you need for access
a. Look at the actual materials and decide what information you want to capture about those materials
b. Look at other sites and established digital programs for ideas and information
c. Compose lists of metadata elements used in these other projects

4. Research metadata standards
a. Use this guide as a starting point to determine what metadata standards are available and what needs and types of information they address
b. Look at the documentation for those standards
c. If implementing the entire standard is not an option, choose elements from the standard that can augment your descriptions and metadata

5. Consider workflows and timelines
a. What information do you need to gather as you select the material?
b. What information do you need to create when you scan the source?
c. What do you need to know to effectively manage the collection?
d. Who will be in charge of quality control of the metadata?

6. Examine your available resources
a. Knowledge and expertise of staff
b. Technical infrastructure available to the institution or collaborative
c. Funding

7. Design a database, set up a content management system, and/or use cataloging utilities to create and manage records and files
a. Dublin Core records can be created directly in OCLC's Connexion
b. OCLC also offers CONTENTdm as a collection management system that provides tools for building, managing, and sharing digital collection on the Web. The metadata standard you choose should fit the goals, formats, subject coverage, and budget of your institution. Metadata creation is very time consuming, and in most cases takes more resources than digitization. Creating the best quality digital object and the best metadata possible is cost effective because it saves you from redoing work down the line. The following are a few issues to keep in mind when planning a digitization and metadata creation program:
Other types of metadata
To manage digital objects into the future, other types of metadata, described above as technical metadata, administrative metadata, and preservation metadata, are needed. Recording information about the digital objects and how they were created is important as staff, computer systems, and file formats change over time.

Pointers to metadata schemas in these areas can be found in the IDL document "Choosing a Metadata Standard for Your Digital Project."

Complex vs. simple items
Another consideration is the type of material you plan to digitize. The more complex the material, the more complex your metadata will be. For example, with multiple page digital objects such as books, structural metadata to relate the pages of the book to each other becomes vital.

Item level vs. collection level
Different types of materials require different levels of descriptive metadata. A large collection of disparate materials probably requires item-level description with separate metadata records for each item in the collection. For a cohesive archival collection with large numbers of like items, an institution may choose to only describe the items at the folder, box, or even collection level. The IDL will accept metadata at any or all of these descriptive levels.

Metadata Working Group
August 2, 2004
Lisa Cahill
 Indiana Historical Society
Jenn Riley
 Indiana University
Yu Su
 Indiana Humanities Council

Choosing a metadata standard for your digital project

DESCRIPTIVE METADATA

Dublin Core (DC)

Type of collection: Groups of materials whose organization is not a significant access point to the collection. Generally these materials will not be organized hierarchically. The collection may be in any format, or a mixture of formats. Collections for which item-level description is planned and none of the other metadata needs described in this document apply.

Appropriate metadata standard: Dublin Core Metadata Element Set 1.1

Example collection: Worthington Memory <http://www.worthingtonmemory.org/>

Resources to consult: Dublin Core Metadata Element Set, Version 1.1: Reference Description <http://www.dublincore.org/documents/dces/>; Dublin Core Usage Guide <http://www.dublincore.org/documents/usageguide/>; DCMI Metadata Terms <http://www.dublincore.org/documents/dcmi-terms/>

Usage notes: Simple Dublin Core, consisting of only the 15 core elements, can be used, although much desired robustness (e.g., distinguishing between roles of the creator of a resource, specifying the name of the controlled vocabulary from which a value is selected) is not possible. Qualified Dublin Core, using the DCMI "Other Elements and Element Refinements" as specified in <http://www.dublincore.org/documents/dcmi-terms/#H3> can be used to slightly increase robustness, although the refinements available may not be appropriate for many uses. Creating local qualifiers for Dublin Core elements is another option, however this practice reduces the interoperability of the metadata created. The emerging Dublin Core Library Application Profile <http://dublincore.org/documents/library-application-profile/>, still in Working Draft status, attempts to address these issues.

Metadata Object Description Schema (MODS)

Type of collection: Materials requiring fairly in-depth item level description, including explicit specification of creator roles, date types, or subject authorities. Materials with existing item-level MARC cataloging.

Appropriate metadata standard: MODS 3.1

Example collection: University of Chicago Chopin Early Editions Project <http://chopin.lib.uchicago.edu/>

Resources to consult: MODS official Web site <http://www.loc.gov/standards/mods/>; MODS User Guidelines <http://www.loc.gov/standards/mods/v3/mods-userguide.html>

Usage notes: MODS is intended to be applicable to any type of resource and provides a middle ground between Dublin Core and MARC. Items for which MARC cataloging already exists can be transformed into MODS records for use in digital library applications.

Encoded Archival Description (EAD)

Type of collection: Archivally-processed groups of materials whose organization and/or provenance is significant. These materials will generally be hierarchically arranged natural groupings assembled by a collector or creator (e.g., the papers or correspondence of a certain individual) rather than less-tightly related groups of materials assembled by a holding institution. A Finding Aid or inventory may or may not already exist. Materials will often be unpublished. Description at the collection level is necessary, lower levels of description may or may not be appropriate. Materials described may or may not be available in digital form.

Appropriate metadata standard: EAD 2002

Example collection: Finding Aids in the Online Archive of California <http://www.oac.cdlib.org/>

Resources to consult: Official EAD Version 2002 Web Site <http://www.loc.gov/ead/>; EAD Help Pages, especially the EAD 2002 Cookbook < http://www.archivists.org/saagroups/ead/>

Usage notes: EAD 2002 can be used to describe the collection at the item, folder-only, or collection-only level. One EAD document should be created for each collection.

Text Encoding Initiative (TEI)

Type of collection: Text collections intended for full-text searching in an online environment. Full text may or may not be intended to be used together with page images of the original document.

Appropriate metadata standard: TEI P4

Example collection: Indiana University, Wright American Fiction, 1851-1875 <http://www.letrs.indiana.edu/web/w/wright2/>

***Resources to consult:** TEI Text Encoding in Libraries Guidelines for Best Encoding Practices <http://www.indiana.edu/~letrs/tei/>; TEI Guidelines <http://www.tei-c.org/Guidelines2/index.htm>*

Usage notes: Choose an Encoding Level as described in the TEI Text Encoding in Libraries Guidelines for Best Encoding Practices. Bibliographic information for resource discovery is encoded in the TEI Header. The entire full text of the resource is marked up structurally in the bulk of the TEI document, and this markup is used for powerful full-text searching. A new version, TEI P5 is currently in development <http://www.tei-c.org/P5/>.

Visual Resources Association (VRA) Core

Type of collection: Art images whose users require in-depth indexing and retrieval using expert terms for genre, culture, style, period, etc.

Appropriate metadata standard: VRA Core Categories 3.0

Example collection: Cleveland Museum of Art Collections <http://www.clevelandart.org/Explore/> (From each item view, click "More Information" to see VRA image metadata.)

Resources to consult: VRA Core version 3 home page <http://www.vraweb.org/vracore3.htm>; CC:DA Task Force on VRA Core

Categories Summary Report <http://www.libraries.psu.edu/tas/jca/ccda/docs/tf-vra1.pdf>; Cataloging Cultural Objects <http://www.vraweb.org/CCOweb/>.

Usage notes: VRA Core is more robust than Dublin Core for describing art images and metadata in this format is consequently more powerful but more expensive to create. VRA Core contains both "work" records describing an actual art object, and "image" records describing representations of views of that object (slides, digital images, etc.) held by an institution. Best practice in creating VRA Core records is to populate fields using appropriate controlled vocabularies such as ULAN and TGM, and the rules described in Cataloging Cultural Objects <http://www.vraweb.org/CCOweb/>. A new version, VRA Core 4.0, is currently in Beta <http://www.vraweb.org/datastandards/VRA_Core4_Welcome.html>.

Categories for the Description of Works of Art Lite (CDWA Lite)

Type of collection: Art images whose users require in-depth indexing and retrieval using expert terms for genre, culture, style, period, etc.

Appropriate metadata standard: CDWA Lite 0.09

Example collection: See example records at <http://www.getty.edu/research/conducting_research/standards/cdwa/3_cataloging_examples/index.html>

Resources to consult: CDWA Lite home page <http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite/index.html>; List of CDWA Lite Elements, Tags, Description, and Examples <http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite/cdwalite.pdf>

Usage notes: CDWA Lite comes out of the museum environment and provides a slightly different approach than the library-based VRA Core. The CDWA Lite schema was developed specifically to share CDWA metadata over the OAI protocol. Best practice in creating CDWA Lite records is to populate fields using appropriate controlled vocabularies such as ULAN and TGM, and the rules described in Cataloging Cultural Objects <http://www.vraweb.org/CCOweb/>.

Gateway to Educational Materials (GEM)

Type of collection: Learning objects that serve education communities (pre-school, K-12, higher education, vocational and technical training, and lifelong learning). These materials require classification criteria special to the education community, such as education level of the target audience, pedagogical methodology, and standards alignment.

Appropriate metadata standard: Dublin Core Metadata Element Set 1.1 as extended in GEM profile

Example collection: NASA Space Science Education Resource Directory <http://teachspacescience.org/> (From each item view, click "More Information" to see GEM metadata.)

Resources to consult: GEM home page <http://www.thegateway.org/>; GEM top-level elements
<http://raven.ischool.washington.edu/about/documentation/metadataElements/>.
Usage notes: The GEM metadata element set is based on Qualified Dublin Core, with additional elements and qualifiers added to meet educational needs. The GEM initiative has also developed application profiles and controlled vocabularies for use with the GEM metadata element set.

OTHER TYPES OF METADATA

Metadata Encoding & Transmission Standard (METS)

<http://www.loc.gov/standards/mets/>. METS is an XML schema allowing users to "wrap" existing descriptive metadata in any format with structural, technical, administrative, preservation, and meta-metadata to create a single metadata object for a resource. METS provides extension schemas with recommended technical metadata for still images, audio, and video.

NISO Metadata for Images in XML (NISO MIX)

<http://www.loc.gov/standards/mix/>. MIX is an XML schema implementation of the NISO draft standard Z39.87-2002, *Data Dictionary – Technical Metadata for Digital Still Images*.

PREservation Metadata Implementation Strategies (PREMIS)

<http://www.oclc.org/research/projects/pmwg/>. PREMIS is not a metadata element set, rather it is an initiative to develop a set of core elements for preservation metadata for the purpose of long-term preservation of digital objects. Some preliminary recommendations of this group which include some metadata element proposals may be found in the report *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects* <http://www.oclc.org/research/projects/pmwg/pm_framework.pdf>.

Suggested Technical Metadata Elements
Indiana Digital Library

(Based on Colorado Digitization Program Western States Dublin Core Metadata Best Practices
http://www.cdpheritage.org/resource/metadata/wsdcmbp/format_creation.html)

The following elements may be used to capture technical information about the hardware, software, and processes used to create a digital resource. As a general guideline, information that describes technical aspects of the digital object's creation is beneficial for long-term administration, technical support, and maintenance of digital objects.

> <u>File Size</u> - Best practice is to record the file size as bytes (i.e. 3,000,000 bytes) and not as kilobytes (Kb), megabytes (Mb), etc.
>
> <u>Compression Scheme</u> - Compression scheme used for optimized storage and delivery of digital object.
>
> <u>Object Producer</u> - Name of scanning technician, digitization vendor, or other entity responsible for the digital object's creation.
>
> <u>Creation Hardware</u> - If a hardware device was used to create, derive or generate the digital object, indicate the manufacturer, model name, and model number of this device
>
> <u>Creation Software</u> - Name and version number of the software used to create the digital object.
>
> <u>Creation Methodology</u> - If the creation process used a standard series of steps, derivations, or techniques, either state or refer to a URL describing the creation process.

- The owning institution of the digital object may create and manage each of these elements as separate database fields.
- Refer to NISO document Z39.87-2002, TECHNICAL METADATA FOR DIGITAL STILL IMAGES for additional technical metadata elements and examples: http://www.niso.org/standards/resources/Z39_87_trial_use.pdf. While this standard is specifically for digital still images, many of the fields it recommends are useful for other types of digital objects.

Examples:

> File Size: 35,000,000 bytes
> Compression Scheme: uncompressed
> Object Producer: jenlrile [Network ID at local institution of person who digitized the object]

**Proposed Digital Imaging Standards and Best Practices
for Indiana LSTA Digitization Projects**

This document provides information on the application of published standards and best practices for digital imaging to determine specifications for individual projects. LSTA-funded projects should adhere to these guidelines.

GENERAL PRINCIPLES

Capture once, use many times

Digitization is expensive, time-consuming, and requires extensive handling of original materials. Any digitization project should thus focus on creating high-quality *master* images from which many *derivative* images can be created for specific uses (e.g., web delivery). The master image should capture all "important" information from the original material, which should be explicitly defined for each digitization project. The master image should also be flexible enough to allow derivatives to be created meeting a wide variety of current and future needs. Therefore, no image processing (such as sharpening) is generally done to the master file. The best practices described in this document are designed to achieve this goal of flexibility.

Scan from earliest generation practical

As copies are made of analog materials, each generation loses some detail. From a photographic negative to a print to a copy negative, from a book to microfilm, there is generational loss of information. To capture the most information in a scanned image, always use the earliest generation of original it is practical to use. In general, scanning from negatives rather than prints and original printed material rather than microfilm or photocopies is preferable. However, there are cases where practical considerations dictate using a second- or third-generation original as the source of a scanned image. A set of cracked or broken glass-plate negatives might benefit from professional printing, then scanning the prints. A large series of bound volumes that have been microfilmed would be considerably cheaper to scan from microfilm rather than to disbind the volumes for scanning or invest expensive face-up scanning equipment. In these cases, a determination must be made if images created from later-generation originals can still meet the flexibility goals of master images for the project.

Technical issues

When setting technical specifications for digitization projects, higher is not always better. There is no advantage to scanning at a resolution higher than what is needed to capture the amount of detail on the original. In fact, there is a large disadvantage to this practice in that this excess resolution adds file size without adding detail to the digital image. The guidelines in this document are designed to help determine appropriate specifications and ensure files are as large as they need to be, but no larger.

A digitization program should employ some sort of color management solution to ensure scanners, monitors, and printers all represent image color accurately. Using "canned" International Color Consortium (ICC) profiles for each imaging device is a low-cost, somewhat effective mechanism, while using professional profiling software is a much more accurate but higher-cost solution.

Using digitization equipment appropriate to the materials being scanned is essential to an effective digitization project. Unfortunately, there are no one-size-fits-all digitization equipment

solutions. For example, A4 flatbed scanners are useful for unbound textual materials and photographic prints, while transparencies and negatives are much better imaged with dedicated film scanners, as flatbed scanners with transparency adapters have a more limited dynamic range often not sufficient for digitizing film. Materials larger than A4 require a scanner with a larger scan bed (moderately-priced scanners up to A3 are available), but these have lower top resolutions than A4 scanners. Never use a scanner at a resolution setting above its listed *optical* resolution (known as an *interpolated* resolution).

Quality control

A structured quality control program is essential to a good digitization project. An effective program might combine automated checking of objective criteria such as image resolution and bit depth for all images with manual checking of subjective criteria such as color fidelity on a subset of scanned images.

Technical metadata

Recording adequate technical metadata about scanned images is essential for long-term maintenance of master files. The NISO draft standard *Technical Metadata for Digital Still Images* in its XML Schema form from the Library of Congress at <http://www.loc.gov/standards/mix/> offers guidance on what sorts of technical metadata are appropriate to record.

VISUAL MATERIALS

In general, the specifications below conform to the guidelines put forth by the January 2003 revision of the *Western States Digital Imaging Best Practices* document at <http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf>. The Western States Best Practices also include more information on the general principles outlined above and a great deal of additional helpful information.

Photographic prints

Photographic prints are generally not scanned at a fixed resolution but instead at a fixed number of pixels across the long side, resulting in two differently-sized prints from one negative yielding similarly-sized digital files. The appropriate resolution is determined by dividing the desired number of pixels (e.g. 3000) by the number of inches of the long side of the photograph (e.g. 10" for an 8x10 photo). In this case 3000 / 10 = 300, so an 8x10" print should be scanned at 300ppi.

In general, color photographs should be scanned as 24-bit RGB color and black & white photographs in 8-bit grayscale. There are many cases, however, when black & white photographs would benefit from color scanning, for example, when they are sepia-toned or badly faded.

Master Files:
        Pixel dimensions: long side of 3000 pixels. Rarely, photographic prints will contain enough detail for scanning at 4000 pixels on the long side
        Resolution: sufficient to achieve desired pixel dimensions
        File format: uncompressed TIFF, Intel byte order
        Bit depth: 24-bit RGB color or 8-bit grayscale, as appropriate

Full-screen Files:
        Pixel dimensions: long side of 1000 pixels

Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Access Files:

Pixel dimensions:	long side of 600 pixels
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Thumbnail Files:

Pixel dimensions:	long side of 200 pixels
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Transparency film

Transparency film is generally not scanned at a fixed resolution but instead at a fixed number of pixels across the long side, so that a 35mm transparency and a 6x9 cm transparency would yield similarly-sized digital files.

Due to inherent color casts of transparency film, color film is generally scanned in RGB color and black & white film in grayscale.

Master Files:

Pixel dimensions:	long side of 3000-5000 pixels. Commercial-grade transparencies may benefit from higher values in this range
Resolution:	sufficient to achieve desired pixel dimensions
File format:	uncompressed TIFF, Intel byte order
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Full-screen Files:

Pixel dimensions:	long side of 1000 pixels
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Access Files:

Pixel dimensions:	long side of 600 pixels
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Thumbnail Files:

Pixel dimensions:	long side of 200 pixels
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Negative film

Photographic negatives are generally not scanned at a fixed resolution but instead at a fixed number of pixels across the long side, so that a 35mm negative and a 6x9 cm negative would yield similarly-sized digital files. Images from negatives should be stored as positives, a conversion which scanning software will do for you.

Due to inherent color casts of negative film, color film is generally scanned in RGB color and black & white film in grayscale.

Master Files:
        Pixel dimensions: long side of 3000 pixels. Commercial-grade negatives may benefit from higher values in this range
        Resolution: sufficient to achieve desired pixel dimensions
        File format: uncompressed TIFF, Intel byte order
        Bit depth: 24-bit RGB color or 8-bit grayscale, as appropriate

Full-screen Files:
        Pixel dimensions: long side of 1000 pixels
        Resolution: 72 ppi
        File format: JPEG
        Bit depth: 24-bit RGB color or 8-bit grayscale, as appropriate

Access Files:
        Pixel dimensions: long side of 600 pixels
        Resolution: 72 ppi
        File format: JPEG
        Bit depth: 24-bit RGB color or 8-bit grayscale, as appropriate

Thumbnail Files:
        Pixel dimensions: long side of 200 pixels
        Resolution: 72 ppi
        File format: JPEG
        Bit depth: 24-bit RGB color or 8-bit grayscale, as appropriate

MUSICAL SCORES AND SHEET MUSIC

The guiding principle behind scanning sheet music is to capture the small details of the musical notation accurately. However, sheet music often has screen-printed color covers that also need to be captured accurately, requiring a resolution high enough to avoid moiré patterns from appearing as a result of the printed halftones. The recommendations below are designed to meet both goals for publications with standard printing sizes.

The access file dimensions recommended below are designed to allow an entire image to fit on a screen, and the full-screen file dimensions are intended to effectively show musical notation on-screen. Thumbnail images for web delivery are generally only desired for graphic covers; thumbnail images of pages of musical notation are not for the most part useful to end-users.

Master Files:

Pixel dimensions:	dependent on size of original
Resolution:	300ppi
File format:	uncompressed TIFF, Intel byte order
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Full-screen Files:

Pixel dimensions:	1200 pixels tall
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Access Files:

Pixel dimensions:	600 pixels tall
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Thumbnail Files:

Pixel dimensions:	200 pixels tall
Resolution:	72 ppi
File format:	JPEG
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

Maps

The guiding principle behind scanning maps is to adequately capture the small characters and lines drawn on the map. Maps with smaller details require higher scan resolutions, and those with larger details require lower resolutions. Calculation of appropriate resolution based on the size of the smallest detail can be done with UIUC's Digital Imaging & Media Technology Initiative online Image Quality Calculator (UIUC), based on principles put forth in Kenney & Rieger's *Moving Theory into Practice* (Kenney, 44-47). The large physical dimensions of many maps can require specialized scanning equipment to digitize at required resolutions. Outsourcing digitization of large maps with small details may be appropriate.

Web-deliverable images of maps can be challenging as well. Small- and medium-sized maps might be delivered appropriately in JPEG formats at sizes comparable to photographic materials. Larger maps might benefit from presentation in a "zoomable" image format such as MrSid or JPEG2000. However, these formats are not natively viewable in today's web browsers, but require users to download and install third-party plug-ins.

Master Files:

Pixel dimensions:	dependent on size of original
Resolution:	300-600ppi
File format:	uncompressed TIFF, Intel byte order
Bit depth:	24-bit RGB color or 8-bit grayscale, as appropriate

TEXTUAL MATERIALS

Special considerations for textual materials

Often, scanning is an interim step towards the goal of creating machine-readable text files for searching and display. Decisions must be made whether to treat textual materials for capture simply for the textual content they contain, or whether to consider them archival documents whose current appearance it is important to capture faithfully.

When converting scanned images to text files, whether through Optical Character Recognition or re-keying, markup of the text file is an important consideration. Encoding of text for current display purposes only in HTML is not a sustainable approach. Time is rather better spent in encoding text files in a content-based markup scheme in XML, then converting these source files into versions for web display with XSLT. The Text Encoding Initiative guidelines (TEI) are appropriate for encoding of many types of textual documents.

Printed texts

Printed texts do not generally hold the same artifactual value as graphic materials or archival documents. The following recommendations are largely based on the Digital Library Federation's *Benchmark for Faithful Digital Reproductions of Monographs and Serials* best practices document (Digital Library Federation). They hold for projects where searchable full text is the primary objective, and page images are only secondary.

The size recommendation listed below for access files is appropriate for single page images (1-up) at standard print sizes. 2-up pages and unusually small print sizes may require differently-sized access files to show text at a readable size on a computer screen. For printed texts, thumbnail images do not generally show enough detail to be useful.

Master Files from paper originals:
      Pixel dimensions: dependent on size of original
      Resolution: 600 ppi
      File format: CCITT Group 4 compressed TIFF, Intel byte order
      Bit depth: Bitonal (1 bit per pixel)

Master Files from microforms:
      Pixel dimensions: dependent on size of original
      Resolution: sufficient to yield a 600 ppi image at original page size
      File format: CCITT Group 4 compressed TIFF, Intel byte order
      Bit depth: Bitonal (1 bit per pixel)

Access Files:
      Pixel dimensions: long side of 600 pixels
      Resolution: 72ppi
      File format: LZW-compressed GIF
      Bit depth: 8-bit (native to GIF)

Archival documents

Archival documents are generally of artifactual value and are not good candidates for optical character recognition. They often pose interesting digitization problems as they may be handwritten, on onionskin paper, or have watermarks of interest. Images captured for the purposes of preserving these properties of the original may or may not be appropriate for Optical

Character Recognition. The resolution recommendations indicated below are appropriate for normal-sized handwriting and typesetting; unusually small characters may require higher resolutions.

Most documents only require scanning in 8-bit grayscale to capture all important information from the paper original; however, documents with color ink, raised colored seals, and the like, will require scanning in 24-bit RGB color. The size recommended for access files would display normal-sized text at a size readable on a computer screen, but smaller characters may require higher resolutions. Depending on the nature of the paper original, thumbnail images may or may not be useful.

Master Files:
      Pixel dimensions: dependent on size of original
      Resolution: 300 ppi
      File format: uncompressed TIFF, Intel byte order
      Bit depth: 24-bit color, 8-bit grayscale

Access Files:
      Pixel dimensions: long side of 600 pixels
      Resolution: 72 ppi
      File format: JPEG or GIF
      Bit depth: dependent on file format

Thumbnail Files:
      Pixel dimensions: long side of 400 pixels
      Resolution: 72 ppi
      File format: JPEG or GIF
      Bit depth: dependent on file format

<u>Published Standards and Best Practices</u>

These online and print publications were cited in this document.

Digital Library Federation Benchmark Working Group. Benchmark for Faithful Digital Reproductions of Monographs and Serials, Version 1. December 2002 <http://www.diglib.org/standards/bmarkfin.htm>.

Kenney, Anne R., and Oya Y. Rieger. Moving Theory into Practice: Digital Imaging for Libraries and Archives. Mountain View, CA: Research Libraries Group, 2000.

TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices, Version 1.0. July 30, 1999 <http://www.diglib.org/standards/tei.htm>.

UIUC Digital Imaging & Media Technology Initiative. Image Quality Calculator. 2000 <http://images.library.uiuc.edu/projects/calculator/>.

Western States Digital Standards Group, Digital Imaging Working Group. *Western States Digital Imaging Best Practices*, *Version 1.0*. January 2003 <http://www.cdpheritage.org/resource/scanning/documents/WSDIBP_v1.pdf>.