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 crosswalking, 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