Metadata Design and Best Practices

Introduction

Metadata is “structured information associated with an object for purposes of discovery, description, use, management, and preservation” (NISO, http://framework.niso.org/24.html). Good metadata facilitates identification and location of content, allows for coordination of instances of an item and supports long-term management and preservation of content.

Select examples of actions facilitated by metadata include:

- Discovering a copy of an article applicable to a class project
- Discover a book in Blacklight, the digital representation in HathiTrust and an image of illustration from the book in SharedShelf or Digital Collections Portal
- Discover a resource and then locate other resources from the same collection
- Uniquely identifying a resource and differentiating between resources
- Monitor fixity of digital resources to ensure
- Build a user experience leveraging logical order among resources or grouping images of the same resource

When trying to understand how best to describe items in a repository, it is important to take into consideration user expectations and needs as well as community standards and best practices; further, it is important to not allow metadata decisions to be determined based on system limitations as metadata inevitably outlives the current discovery environment.

Acknowledging that quality metadata requires involvement from multiple stakeholders, and understanding the intersection of differing priorities, allows for the creation of user focused metadata that is interoperable and reusable across collections both within and beyond the platform in which the metadata were originally created.

Metadata Services (within Library Technical Services) defines and provides guidance for metadata practice across CUL to ensure that metadata is interoperable and reusable across CUL’s discovery environments and beyond. This document is intended to help Repository and Collection managers consider the importance of proper metadata design and handling to overall functionality of your repository system. We list questions to consider throughout the process, but the intent is for your repository to work together with Metadata Services to design the most productive yet interoperable repository possible within the repository ecosystem at CUL. You can reach the metadata services group by emailing metadata_info@cornell.edu, and additional introductory information can be found on the metadata services website.

Because decisions and workflows around metadata practice are idiosyncratic and based on individual circumstances, case by case evaluations will be necessary. This document introduces repository managers and designers to high-level topics for consideration before consulting Metadata Services.

Contents in this section

- Metadata categories
- Metadata design
- Metadata production
- Metadata maintenance and clean-up
- Documentation

Metadata categories

Generally, metadata can be categorized into a few areas:

1. **Descriptive metadata** enables discovery, identification and retrieval of materials hosted in the repository. It is often the first and most commonly considered metadata type when implementing a repository, and includes elements such as title, author, subject, dates etc.

2. **Preservation metadata** captures details about the resource that is important for long-term retention and management. It may include information about the technical environment, custody and change history, processing, storage and status. It is important that preservation metadata is compatible with the collections management workflow of the archiving institution.

3. **Technical metadata** describes the electronic nature of the resources, such as information about format, file size, checksums, sampling frequencies, how the metadata themselves are collected, and other similar characteristics.

4. **Structural metadata** relates components of a compound resource and/or bundles related objects into a package by describing the way those components are organized. For example, how chapters or pages of a book are related, or in a collection of pictures about a piece of art, what view or perspective each image is of that work.

5. **Administrative metadata** facilitates the management of a resource. It can include such things as information about when and how an object was created, what processing activities have occurred on the object, the documentation of intellectual property ownership, usage and access restrictions and more.

Identifiers are important across all categories of metadata, intended to: uniquely identify a resource; differentiate between resources; coordinate resources across collections and systems; facilitate preservation; manage relationships between resources and collections; and more.

[Return to top]
Metadata design

While implementation practice and community norms vary greatly for each type of metadata and across standards, repository managers and metadata professionals should work carefully to build a metadata model that ensures stewardship of resources through the entire digital lifecycle.

Metadata practice versus application functionality

Metadata decisions should not be guided by limitations imposed by the application or user experience; metadata outlives user interfaces and functionalities of any single repository. While it is understandable to want to identify an easy solution by changing metadata practice, doing so has long-term implications for discovery of the resources.

Metadata Services, a central service provider at CUL, is available to help define metadata models and practice to ensure that the repository meets user requirements meanwhile ensuring adherence to standards. Prior to speaking with Metadata Services, it is helpful to consider the following:

- Have you outlined functional requirements based on known user needs? If so, have you considered what metadata elements might reasonably be provided by a self-service submission, versus which should be administratively applied?
- Have you considered a minimum level of metadata elements to be required for deposit (self-submission as well as mediated deposit)?
- Have you identified other repositories or infrastructures that may need to interact with your repository? If so, does your repository environment facilitate data exchange (e.g.: documented APIs, OAI-PMH, etc.)?

Granularity and impacts on scalability

Before talking with metadata services, it may also be helpful to consider the level at which your users will expect resources to be described at and the level at which you can sustainably provide item description. This "granularity" of detail can be thought of two ways:

1. Granularity could be considered as level-of-description, i.e.: should you describe each item, a related set of materials or the entire collection? For example, a series of photographs could be described as "Photos of Ithaca, NY" or as "Photo 1: The Commons, Photo 2: Ithaca Falls, etc".
2. Granularity could also be considered as detail-of-describe. For example, you could describe an album as-a-whole or describe individually the songs on the album.

Granularity impacts scalability regarding production capacity as well as how well the metadata interacts outside of the original collection and repository; this must be considered as metadata requirements are defined for a set of materials.

Metadata modelling/Metadata strategy

While many repositories at Cornell have established metadata models pre-packaged with the environment, nearly all can be customized; this customization should adhere to documented models as it will create complexity when sharing data beyond the native repository. Further, as CUL continues to move toward open source repositories, such as Samvera-based solutions, underlying data models become more open and flexible. When defining metadata models, it is important to look across and beyond the library community; stakeholders and domain-based communities of practice can offer use cases as well as solutions to guide modeling for different material-types.

In undertaking this work, it is important to understand the communities being served by the repository; for more information on this, see the Curation principle document.

Metadata Services should be involved in the design of metadata models for repositories at CUL. As part of this process, Metadata Services works with repository managers and identified stakeholders to:

- Determine specifications required or recommended by the repository platform or infrastructure
- Examine elements and content standards for use in describing your data
- Identify key pieces of descriptive information that facilitate documented functional requirements (e.g.: data exchange across repositories, metadata faceting, preservation needs, etc.)
- Create identifier pattern that can be used across materials and collections within the repository

Interoperability

A primary goal of most repositories is for discovery both within and outside of its home environment; due to this need, repositories must support interoperable metadata models. Interoperability, or facilitation of data exchange across systems, is essential for ensuring discovery across environments (e.g.: repository and aggregators) and facilitating metadata synchronization across systems. Further, interoperable models and standardized practice is important in facilitating discovery across collections even within a single repository. Adhering to defined standards and using adequately featured repository environments will facilitate the requirement for interoperability and substantially facilitate a better user experience.

Metadata production

Balancing user needs with best practices

While difficult to say "No" to a user or depositor, it is important to remember that the user who created the set of materials, or who is complaining the loudest, is not necessarily representative of the user base for a collection. Altering metadata practice based on very specific, individual needs...
creates inconsistent metadata, and can limit possibilities for the reuse of that metadata beyond the original collection.

To aid in responding to user needs, repository managers should consult with metadata services to:

- Determine user needs and how this may impact metadata creation.
- Identify who in the repository management group will respond to user requests.

**Guidelines/application profile**

As part of metadata production, metadata guidelines should be produced, outlining fields used and how those fields should be populated; this documents aspects of the project such as required fields, repeatable fields, expected values and formatting requirements (e.g., Extended Date Time Format (EDTF)), and more. This information should be documented within a metadata application profile for management of the repository. Further, clearly articulated guidelines aid the metadata capturer in creating consistent and well-formed metadata.

**User-generated vs. library-generated metadata**

Repository systems often contain a combination of user-generated and librarian or repository manager created metadata. There are benefits to both workflows, but issues related to maintenance and curation of user generated metadata should be specifically considered. CUL Metadata Services can help mitigate problems and will discuss with you things like:

- How best to present metadata elements to encourage their use and communicate importance
- How to teach standard and/or best practices
- Implications of user failure to submit controlled terms (for example how this might limit faceting and browsing within or beyond local environment)

**Metadata maintenance and clean-up**

All metadata eventually demands clean-up or enhancement; no metadata should be considered static. Normalizing, updating and enhancing metadata is common as we learn more about our resources, as repositories migrate between systems and as the evolution of data affords different models for description.

Metadata Services routinely performs maintenance and clean-up on existing metadata across repositories; these efforts are performed using manual workflows (item-by-item editing) as well as batch processes using scripts and extract, transform and load (ETL) tooling. Metadata services is available to consult on maintenance issues and might be available to perform these tasks.

**Documentation**

As with all aspects of repository management, documentation around data models and metadata decisions is an essential component supporting long-term sustainability and interoperability. Clear, complete and transparent documentation is important across the entire spectrum of metadata, including the model, guidelines, workflows and decisions made as part of production efforts.

Metadata services can again help you define and begin to create proper metadata about your metadata, but the following factors should be considered when drawing up good documentation:

- What organizational structure / naming schemes can be consistently used to facilitate creation and finding of the documentation?
- Where should the documentation be stored so that it is available to all necessary producers/consumers?
- Who (which specific individual or group of individuals) is responsible for keeping the documentation up to date, and how will they be held accountable?
- What are the triggers for review and update?
- What parts of your documentation can or should be made publicly available?