Metadata

Metadata

York University Libraries

Contents

Types of metadata[1]

Descriptive

- Describes a resource for purposes such as discovery and identification
- Can include elements such as title, abstract, author, subject, and keywords

Structural

- Indicates how compound objects are put together
- Example:
  - Show relationships between digital object and page number of book
  - The first scanned page of a book is rarely marked as page # 1 of the book itself

Administrative and Technical

- Provides information to help manage a resource such as:
  - when and how it was created, file type and other technical information, and who can access it
- Subsets of administrative data:
  - Terms and Conditions – deals with intellectual property rights
  - Preservation Metadata – contains information needed to archive and preserve a resource

1. http://www.niso.org/standards/resources/UnderstandingMetadata.pdf

Metadata: standards, schemata, and profiles

There are many metadata standards, each providing ways to structure data. When data is structured we can discover, trace, and visualize relationships between data. Keep in mind that metadata standards, like any knowledge structures, reflect the particular biases and mentalities of the institutional organizations, individuals or communities of practice that develop and use them.

Dublin Core is a type of metadata standard. Dublin Core (DC) is a small set of vocabulary terms developed for describing digital objects. DC is the minimum standard for describing digital objects. Dublin Core comes in a simple (15 elements) and a larger set. The Simple Dublin Core Metadata Element Set (DCMES) includes 15 metadata elements (terms). Each element describes a property of a resource.

Simple Dublin Core set:

1. Title	6. Contributor	11. Source
2. Creator	7. Date	12. Language
3. Subject	8. Type	13. Relation
4. Description	9. Format	14. Coverage
5. Publisher	10. Identifier	15. Rights

All elements are optional and repeatable.

Uploading Digital Objects

At York University Libraries, born digital or digitized objects are deposited in YorkSpace, our institutional repository, or York University Digital Library (YUDL), our preservation repository, depending on their audience, format, and their copyright/rights status.

In YorkSpace, metadata is stored in a Dublin-Core (DC) form. In YUDL, metadata is stored in a MODS form. For both platforms, librarians, archivists, and research team members can hand-enter descriptive metadata through an online form.[12] For research projects, we recommend teams work in shared spreadsheets and then uploading digital objects and their accompany metadata in bulk.

Adding Descriptive Metadata

YUL uses the MODS metadata schema since it supports the system we are moving towards. This systems supports a linked data environment. Linked data is a process by which data is shared and connected on the Semantic Web. Linked data is made possible through the use of Uniform Resource Identifiers (URIs). URIs are characters that identify resources.[3] Linked data is also made possible by

---

1. For step-by-step instructions for YorkSpace see: https://docs.google.com/presentation/d/14OdnFq9gU2NiJNu6o7xmIfpUAaZCoStIjwazqvKpAgs/edit?usp=sharing .
2. For step-by-step instructions for YUDL, contact York University Libraries' Digital Scholarship Centre .
3. https://en.wikipedia.org/wiki/Uniform_Resource_Identifier

the Resource Description Framework (RDF), "a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialization formats."[4]

MODS is a subset of MARCXML elements. MODS is embedded in METS (Metadata Exchange) records for item level description. "The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium."[5]

Taxonomies, Ontologies, Controlled Vocabularies

What information you decide to capture about your content in a first step. *How* you decide to describe, categorize and tag up your content is also important. Generally within the GLAM (Galleries, Libraries, Archives & Museums) we embrace collaborative solutions to information challenges. Why recreate the wheel when there is a community of practice that will allow the creation of a shared vocabulary.

Controlled vocabularies can be used to ensure consistency and can be structured in such a way that related or similar terms can be linked to each other.[6]

Taxonomies are controlled vocabularies that are structured into simple parent>child hierarchies. If you look at the Library of Congress Thesaurus for Graphic Materials, you can see how the term Photographs is part of the broader category of Pictures and includes a number of narrower terms related to format (i.e. Daguerreotypes) or genre (Fashion photographs, Marine photographs).

Ontologies are a taxonomy with more complex structures and specific relationship between terms, so that relationships between terms are more complex (and are qualified beyond a simple hierarchy).

Schema and Data Profiles

Application Profiles are policy documents that instruct how particular project teams, institutions, or communities of practice use, describe and preserve metadata.[7] At York University Libraries, librarians and archivists have developed application profiles for digitized projects over many years, both formally, and informally. Currently, we are in the process of implementing a metadata profile using MODS with the intention to normalize as much metadata in anticipation of migrating to an RDF standard.

Here is our current documentation on how York University Libraries goes about it's preservation of digital and digitized objects. The section on metadata is brief, with more detailed policy and documentation currently in development.

Here is a sample spreadsheet of how one might track and structure the metadata for the objects being generated as part of the digitization project: Metadata_YUL_Template_for_AIFProjects .

Linked Data

"Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods."[8] In practical application, linked data uses URIs and RDF to create machine-readable information that will follow best practices to surface, share and connect information.

---

4. https://en.wikipedia.org/wiki/Resource_Description_Framework
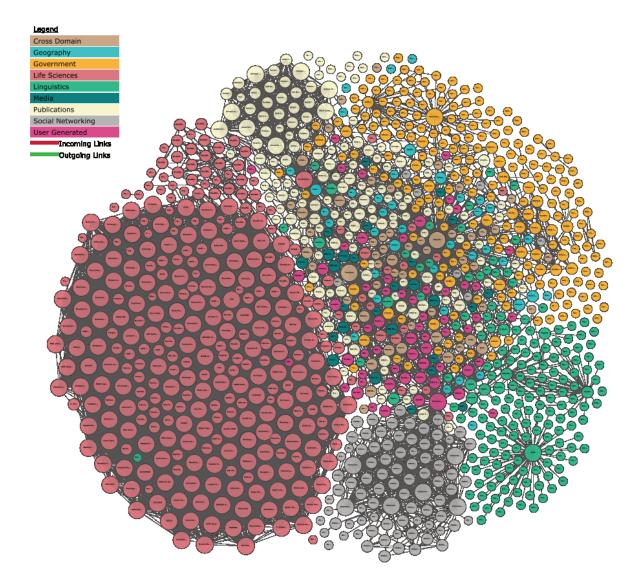
5. http://www.loc.gov/standards/mets/

6. See for example the definition by the American Society for Indexing here: http://www.taxonomies-sig.org/about.htm#cv .

7. See for example, the Digital Public Library of America's Metadata Profile, available at http://dp.la/info/wp-content/uploads/2015/03/MAPv4.pdf .

8. See: http://linkeddata.org/

Linked data requires collaboration among different creators of metadata, and is based on open vocabularies.

Publishing structured data so that it can be interlinked and become more useful through semantic queries… in other words, breaking up your information into discrete "bits" to create strings of interrelations. So individuals are assigned a unique identifier (see VIAF, ORCid), a published work is assigned a unique identifier (DOI, ISBN), places get a unique identifier (GeoNames), discrete objects get unique identifiers (like file formats).

Linked data shifts value from discrete scholarly online projects to more open projects that share their own data, mobilize the work of others, and connect data sets with others.

Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak.
http://lod-cloud.net/

9

9. Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/. For interactive version, see: http://lod-cloud.net/versions/2017-08-22/lod.svg .

Acknowledgements