2021-01-14

Metadata for Research Data

Carruthers, Matthew

http://hdl.handle.net/2027.42/165328

Downloaded from Deep Blue, University of Michigan’s institutional repository
For Research Data

Matt Carruthers
mcarruth@umich.edu
What we’ll cover:

1. Defining metadata.
2. A brief history of the evolution of metadata.
4. Metadata and the research data lifecycle.
5. Identifying the right metadata standard.
What is metadata?

“Data about data.”
What is metadata?

1. Metadata describes the content, quality, condition, and other characteristics of data.
2. Metadata is standardized, structured information about an object that facilitates functions associated with that object. (Discovery, management, rights and access control, preservation, reuse.)
Metadata in libraries: the past

A long time ago....
Metadata in libraries: the present

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>74225557</td>
<td>The 1988 campaign: anatomy of a crucial election.</td>
</tr>
<tr>
<td>3260237309523</td>
<td>120 pages, 30 cm.</td>
</tr>
<tr>
<td>110 2</td>
<td>American Institute for Political Communication.</td>
</tr>
</tbody>
</table>

The table includes detailed information about various metadata entries related to books, including titles, authors, publication dates, and other relevant details. This data is crucial for library management and user search functionalities.
Metadata in libraries: the present

New things to consider with digital objects and collections:
• How do you represent context in a digital environment?
• How do you facilitate long-term preservation?
• How do you track changes to digital objects over time?
  • Format migration
  • Versioning of files
Metadata in libraries: the present

Describing collections of objects rather than single objects/resources

- Archival collections
- Digital collections
- Digital exhibits

*Encoded Archival Description (EAD)*

*Text Encoding Initiative (TEI)*

*Dublin Core (DC)*

*Metadata Object Description Schema (MODS)*

*Metadata Encoding & Transmission Standard (METS)*
Types of Metadata

• Descriptive metadata
  • For finding or understanding a resource

• Administrative metadata
  • For long-term management of files
  • For decoding and rendering files
  • For documenting use and access rights

• Structural metadata
  • For documenting the relationships of parts of resources to one another
## Types of Metadata

<table>
<thead>
<tr>
<th>Metadata Type</th>
<th>Example Properties</th>
<th>Primary Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive metadata</td>
<td>Title, Author, Subject, Genre, Publication date</td>
<td>Discovery, Display, Interoperability</td>
</tr>
<tr>
<td>Technical metadata</td>
<td>File type, File size, Creation date/time, Compression scheme</td>
<td>Interoperability, Digital object management, Preservation</td>
</tr>
<tr>
<td>Preservation metadata</td>
<td>Checksum, Preservation event</td>
<td>Interoperability, Digital object management, Preservation</td>
</tr>
<tr>
<td>Rights metadata</td>
<td>Copyright status, License terms, Rights holder</td>
<td>Interoperability, Digital object management</td>
</tr>
<tr>
<td>Structural metadata</td>
<td>Sequence, Place in hierarchy</td>
<td>Navigation</td>
</tr>
<tr>
<td>Markup languages</td>
<td>Paragraph, Heading, List, Name, Date</td>
<td>Navigation, Interoperability</td>
</tr>
</tbody>
</table>

Riley, Jenn. Understanding Metadata: What is Metadata, and What is it For: A Primer. NISO, 2017
Metadata and the Research Data Lifecycle

Adapted from Hüser, Falco Jonas; Elbæk, Mikael K.; Martinez lavanchy, Paula (2016); DTU Research Data Life Cycle. figshare. Figure. https://doi.org/10.6084/m9.figshare.4258019.v1
Metadata and the Research Data Lifecycle

Levels of Metadata and Documentation:

1. **Study-level**: provides an overview of the research context and design, data collection methods, data preparation and results or findings.

2. **Data-level**: provides labeling and documentation of individual items, such as names and descriptions of variables, and explanations of codes and classification schemes used. It can be embedded within a data collection or recorded in an accompanying document.
Metadata and the Research Data Lifecycle

Metadata that can facilitate reuse will include information on:

- What research data exists
- Where it can be found
- How, when, and why the data was created
- Who created the data
- How to access the data
- What individual data points represent
Metadata and the Research Data Lifecycle

The difference metadata can make for comprehension and reuse:

Example 1
Example 2

Which example is easier to understand?
Which do you think has enough metadata to allow for reuse by other researchers?
Why should researchers care?

And I should care

Why?
Why should researchers care?

Scholarly Communication:
• Metadata can help fight the “Digital Data Deluge”.
• Make it discoverable or it might be lost forever.
Why should researchers care?

“The Long Tail”

Research data can potentially be very useful for a long time.
Why should researchers care?

The practical reasons:

• Potential for increase in data citations
• Funding agency requirements
• Can aid in the organization of your research data while you are creating/analyzing it
Identifying the right metadata standard

Too much to cover

ABCD - Access to Biological Collection Data
AgMES - Agricultural Metadata Element Set
AVM - Astronomy Visualization Metadata
CF (Climate and Forecast) Metadata Conventions
CIF - Crystallographic Information Framework
CIM - Common Information Model
CSMD-CCLRC Core Scientific Metadata Model
Darwin Core
DataCite Metadata Schema
DCAT - Data Catalog Vocabulary
DDI - Data Documentation Initiative
DIF - Directory Interchange Format
Dublin Core
EML - Ecological Metadata Language
FGDC/CSDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata
FITS - Flexible Image Transport System
Genome Metadata
IVOA Photometry Data Model (PhotDM)
Simulation Data Model (SimDM)
Space-Time Coordinate (STC) Metadata for the Virtual Observatory
Astronomical Dataset Characterization Data Model (CharDM)
Simple Spectral Lines Data Model
IVOA Spectral Data Model
Observation Data Model Core Components (ObsCoreDM)
ISA-Tab
ISO 19115
MIBBI - Minimum Information for Biological and Biomedical Investigations
MIDAS-Heritage
OAI-ORE - Open Archives Initiative Object Reuse and Exchange
Observ-DM
Observations and Measurements
OME-XML - Open Microscopy Environment XML
PDBx/mmCIF – Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework
Protocol Data Element Definitions
PROV
QuDEx - Qualitative Data Exchange Format
RDF Data Cube Vocabulary
Repository-Developed Metadata Schemas
SDMX - Statistical Data and Metadata Exchange
SPASE Data Model
Identifying the right metadata standard

Questions to consider:
1. Who is the intended audience?
2. What are the research norms of the discipline?
3. Is there already an established standard for that discipline?
4. What legal or ethical requirements?
Identifying the right metadata standard

Resources to help you identify appropriate metadata standards:

- Research Data Alliance Metadata Directory
- Linked Open Vocabularies
- Open Metadata Registry
- FAIRsharing
- BioPortal
Further reading

Research Data Curation Bibliography

Understanding Metadata: What is Metadata, and What is it For?: A Primer

Dublin Core Metadata Initiative Metadata Basics

DataONE Metadata Best Practices

Questions?