2021-01-14

Metadata for Research Data

Carruthers, Matthew https://hdl.handle.net/2027.42/165328 http://creativecommons.org/publicdomain/zero/1.0/

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.
- 3. Metadata in a digital environment.
- 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.
- 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



crit	cal		
ary			
	American Institute for Poli The 1968 campaign: Washington, 1970.	tical Communication anatomy of a crue	ial election.
	vi, 125 p. 23 cm.		
	"An in-depth study of the candidates, issues, and the a month period in the Milwaukee L Elections-Milwaukee m	e evolution of voter att nass media carried out metropolitan area." etropolitan area. 2, Pu	itudes toward over a uine- blic opinion
	Wisconsin Milwaukee metrop kee metropolitan area. I. T	olitan area. 3, Mass m itle.	edia-Milwau.
	JS1117.A9P82	329.0237310923	74-25857 MARO
	Library of Congress	71 (2)	

Metadata in libraries: the present

Book	s		▼ <u>Rec stat</u> c <u>Entered</u> 19710713 <u>Replaced</u> 20140615223401.8			
<u>Type</u>	а		ELVI Srce Audn Ctri Lang eng			
<u>BLvi</u>	m		Form Conf 0 Biog MRec Ctry dcu			
			Cont s GPub LitF 0 Indx 0			
<u>Desc</u>			Ills Fest 0 DtSt s Dates 1970 ,			
010	Г	T	74025857			
040	H	t	DLC #b eng #c DLC #d CRU #d LGG #d NIALS #d OCLCO #d OCLCF #d OCLCQ			
019			13755790			
043			n-us-wi			
050		0) JS1117.A9 +b P82			
082	0	0	329.023/73/0923			
090			+b			
049			EYMG			
110	2		American Institute for Political Communication.			
245	1	4	The 1968 campaign: anatomy of a crucial election.			
260			Washington, ±c 1970.			
300			vi, 125 pages ‡c 23 cm			
336			text #b btt #2 rdacontent			
337			unmediated +b n +2 rdamedia			
338			volume +b nc +2 rdacarrier			
500			"An in-depth study of the evolution of voter attitudes toward candidates, issues, and the mass media carried out over a nine-month period in the Milwaukee metropolitan area."			
650		0	Elections +z Wisconsin +z Milwaukee Metropolitan Area.			
650		0	Public opinion +z Wisconsin +z Milwaukee Metropolitan Area.			
650		0	Mass media ‡z Wisconsin ‡z Milwaukee Metropolitan Area.			
650		7	Elections. +2 fast +0 (OCoLC)fst00904324			
650		7	7 Mass media. #2 fast #0 (OCoLC)fst01011219			
650		7	7 Public opinion. +2 fast +0 (OCoLC)/st01082785			
651		7	Wisconsin +z Milwaukee Metropolitan Area. +2 fast +0 (OCoLC)fst01351531			
776	0	8	3 41 Online version: 4a American Institute for Political Communication. 4t 1968 campaign: anatomy of a crucial election. 4t Washington, 1970 4w (OCoLC)755282559			

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)

```
<mets:structMap</pre>
w<mets:div TYPE="cd:compactDiscObject" DMDID="MODS1">
  ▼<mets:div TYPE="cd:disc">
   ▼<mets:div DMDID="DMD_disc01_tr001" ID="disc01_tr001" TYPE="cd:track">
      ▼<mets:div TYPE="cd:audio":
         <mets:fptr FILEID="FN10081"/>
         <mets:fptr FILEID="FN10320"/>
         <mets:fptr FILEID="FN10325"/>
       </mets:div>
     </mets:div>
    ▼<mets:div DMDID="DMD_disc01_tr002" ID="disc01_tr002" TYPE="cd:track">
      ▼<mets:div TYPE="cd:audio">
        <mets:fptr FILEID="FN10090"/
         <mets:fptr FILEID="FN10334"/>
         <mets:fptr FILEID="FN10339"/>
       </mets:div>
     </mets:div>
    ▼<mets:div DMDID="DMD_disc01_tr003" ID="disc01_tr003" TYPE="cd:track">
      ▼<mets:div TYPE="cd:audio">
         <mets:fptr FILEID="FN1009F"/>
         <mets:fptr FILEID="FN10348"/>
        <mets:fptr FILEID="FN1034D"/>
       </mets.div>
     </mets:div>
    ▼<mets:div DMDID="DMD_disc01_tr004" ID="disc01_tr004" TYPE="cd:track">
     ▼<mets:div TYPE="cd:audio">
        <mets:fptr FILEID="FN100AE"/>
         <mets:fptr FILEID="FN1035C"/>
         <mets:fptr FILEID="FN10361"/>
       </mets:div>
     </mets:div>
    ▼<mets:div DMDID="DMD_disc01_tr005" ID="disc01_tr005" TYPE="cd:track">
     ▼<mets:div TYPE="cd:audio">
         <mets:fptr FILEID="FN100BD"/>
         <mets:fptr FILEID="FN10370"/>
         <mets:fptr FILEID="FN10375"/>
       </mets.div>
     </mets:div>
    ▼<mets:div DMDID="DMD_disc01_tr006" ID="disc01_tr006" TYPE="cd:track">
     ▼<mets:div TYPE="cd:audio">
         <mets:fptr FILEID="FN100CC"/>
         <mets:fptr FILEID="FN10384"/>
         <mets:fptr FILEID="FN10389"/>
       </mets:div>
     </mets:div>
    ▼<mets:div DMDID="DMD disc01 tr007" ID="disc01 tr007" TYPE="cd:track">
     ▼<mets:div TYPE="cd:audio">
         <mets:fptr FILEID="FN100DB"/>
         <mets:fptr FILEID="FN10398"/>
         <mets:fptr FILEID="FN1039D"/>
       </mets:div>
     </mets:div>
    w<mets:div DMDID="DMD_disc01_tr008" ID="disc01_tr008" TYPE="cd:track">
      ▼<mets:div TYPE="cd:audio">
        <mets:fptr FILEID="FN100EA"/>
         <mets:fptr FILEID="FN103AC"/>
         <mets:fptr FILEID="FN103B1"/>
       </mets:div>
      </mets:div>
```

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

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

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



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

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 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

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?



Scholarly Communication:

- Metadata can help fight the "Digital Data Deluge".
- Make it discoverable or it might be lost forever.

"The Long Tail"

Research data can potentially be very useful for a long time.

"Long tail" by User:Husky – Own work. Licensed under Public Domain via Wikimedia Commons – http://commons.Wikimedia.org/wiki/File:L ong_tail.svg#mediaviewer/File:Long_tail.svg

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-OM **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

<u>Understanding Metadata: What is Metadata, and What is it For?: A</u> <u>Primer</u>

Dublin Core Metadata Initiative Metadata Basics

DataONE Metadata Best Practices

Miller, Steven J.,. *Metadata for Digital Collections : A How-to-Do-It Manual*. Neal-Schuman Publishers, 2011.

Questions?