

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

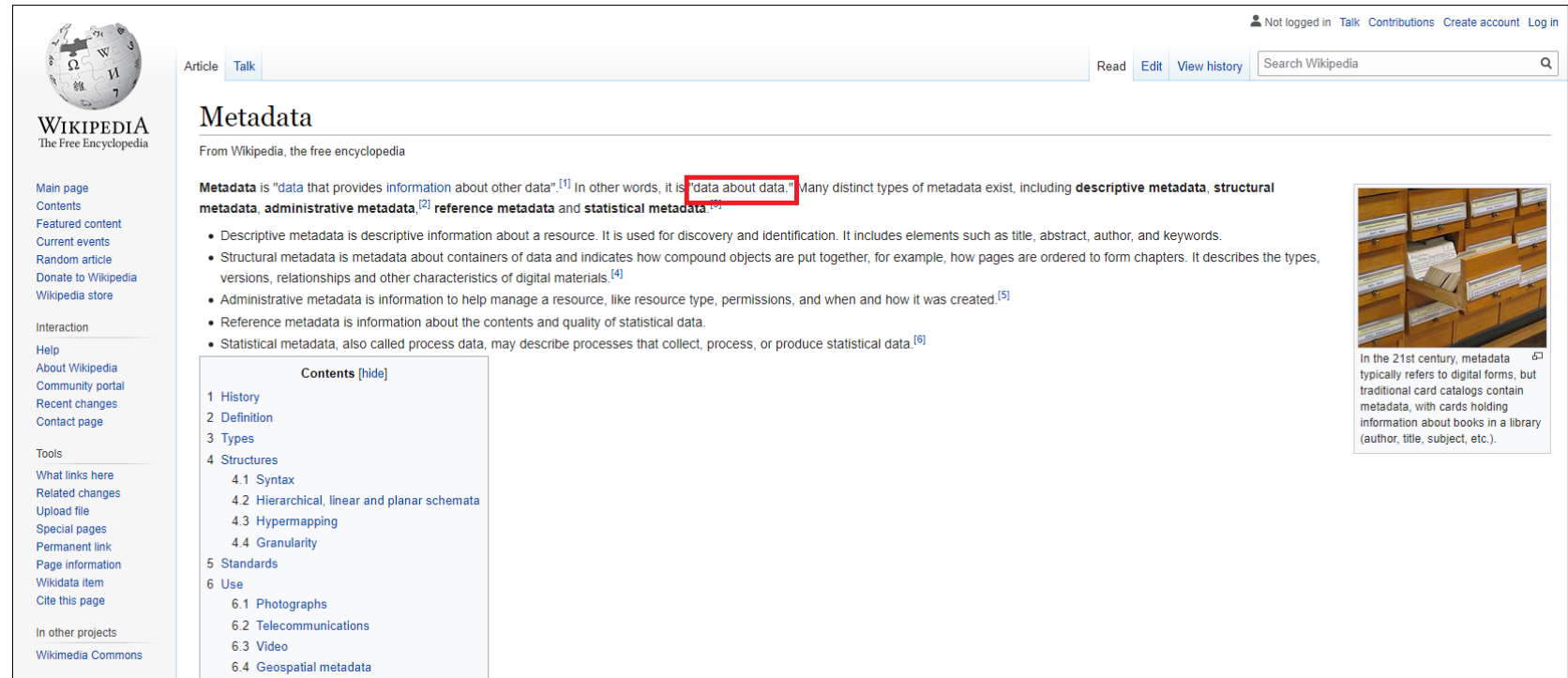
mcarruth@umich.edu

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.”



The screenshot shows the Wikipedia article for 'Metadata'. The page title is 'Metadata' and it is part of the 'WIKIPEDIA The Free Encyclopedia'. The article text defines metadata as "data that provides information about other data" and lists several types: descriptive, structural, administrative, reference, and statistical metadata. A red box highlights the phrase "data about data." in the definition. To the right of the text is an image of a library card catalog with a caption explaining that in the 21st century, metadata typically refers to digital forms, but traditional card catalogs contain metadata with cards holding information about books in a library (author, title, subject, etc.). The page also features a table of contents on the left side of the main content area.

WIKIPEDIA
The Free Encyclopedia

Article Talk

Read Edit View history Search Wikipedia

Metadata

From Wikipedia, the free encyclopedia

Metadata is "data that provides information about other data"^[1] In other words, it is "data about data."^[2] Many distinct types of metadata exist, including **descriptive metadata**, **structural metadata**, **administrative metadata**,^[2] **reference metadata** and **statistical metadata**.^[3]

- Descriptive metadata is descriptive information about a resource. It is used for discovery and identification. It includes elements such as title, abstract, author, and keywords.
- Structural metadata is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.^[4]
- Administrative metadata is information to help manage a resource, like resource type, permissions, and when and how it was created.^[5]
- Reference metadata is information about the contents and quality of statistical data.
- Statistical metadata, also called process data, may describe processes that collect, process, or produce statistical data.^[6]

Contents [hide]

- 1 History
- 2 Definition
- 3 Types
- 4 Structures
 - 4.1 Syntax
 - 4.2 Hierarchical, linear and planar schemata
 - 4.3 Hypermapping
 - 4.4 Granularity
- 5 Standards
- 6 Use
 - 6.1 Photographs
 - 6.2 Telecommunications
 - 6.3 Video
 - 6.4 Geospatial metadata

In the 21st century, metadata typically refers to digital forms, but traditional card catalogs contain metadata, with cards holding information about books in a library (author, title, subject, etc.).

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....

Historical
Library

JS American Institute for Political Communication.

1117 The 1968 campaign: anatomy of a crucial election.
A9 Washington, 1970.

P82 vi, 125 p. 23 cm.

"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."

1. Elections—Milwaukee metropolitan area. 2. Public opinion—Wisconsin—Milwaukee metropolitan area. 3. Mass media—Milwaukee metropolitan area. I. Title.

JS1117.A9P82

329.02373'0923

74-25857

MARC

Library of Congress

71 (2)

Metadata in libraries: the present

Books		Rec stat	c	Entered	19710713	Replaced	20140615223401.8
Type	a	ELvl		Srcr		Audn	
						Ctrl	
						Lang	eng
BLvl	m	Form		Conf	0	Blpg	
						MRec	
		Cont	s	QPub		LitF	0
						Indx	0
Desc		llls		Fest	0	DISt	s
						Dates	1970 ,
010		74025857					
040		DLC #b eng #c DLC #d CRU #d LGG #d NIALS #d OCLCO #d OCLCF #d OCLCQ					
019		13756790					
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 txt #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	Mass media. #2 fast #0 (OCoLC)fst01011219					
650	7	Public opinion. #2 fast #0 (OCoLC)fst01082785					
651	7	Wisconsin #z Milwaukee Metropolitan Area. #2 fast #0 (OCoLC)fst01351531					
776	0 8	#i Online version: #a American Institute for Political Communication. #t 1968 campaign: anatomy of a crucial election. #d Washington, 1970 #w (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 \(METS\)](#)

```
<?xml version="1.0" encoding="UTF-8" ?>
<structMap type="compactDiscObject" DMDID="MODS1">
  <div type="disc">
    <div DMDID="DMD_disc01_tr001" ID="disc01_tr001" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN10081"/>
        <fileptr FILEID="FN10320"/>
        <fileptr FILEID="FN10325"/>
      </div>
    </div>
    <div DMDID="DMD_disc01_tr002" ID="disc01_tr002" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN10090"/>
        <fileptr FILEID="FN10334"/>
        <fileptr FILEID="FN10339"/>
      </div>
    </div>
    <div DMDID="DMD_disc01_tr003" ID="disc01_tr003" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN1009F"/>
        <fileptr FILEID="FN10348"/>
        <fileptr FILEID="FN1034D"/>
      </div>
    </div>
    <div DMDID="DMD_disc01_tr004" ID="disc01_tr004" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN100AE"/>
        <fileptr FILEID="FN1035C"/>
        <fileptr FILEID="FN10361"/>
      </div>
    </div>
    <div DMDID="DMD_disc01_tr005" ID="disc01_tr005" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN100BD"/>
        <fileptr FILEID="FN10370"/>
        <fileptr FILEID="FN10375"/>
      </div>
    </div>
    <div DMDID="DMD_disc01_tr006" ID="disc01_tr006" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN100CC"/>
        <fileptr FILEID="FN10384"/>
        <fileptr FILEID="FN10389"/>
      </div>
    </div>
    <div DMDID="DMD_disc01_tr007" ID="disc01_tr007" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN100DB"/>
        <fileptr FILEID="FN10398"/>
        <fileptr FILEID="FN1039D"/>
      </div>
    </div>
    <div DMDID="DMD_disc01_tr008" ID="disc01_tr008" type="cd:track">
      <div type="cd:audio">
        <fileptr FILEID="FN100EA"/>
        <fileptr FILEID="FN103AC"/>
        <fileptr FILEID="FN103B1"/>
      </div>
    </div>
  </div>
</structMap>
```

Types of Metadata

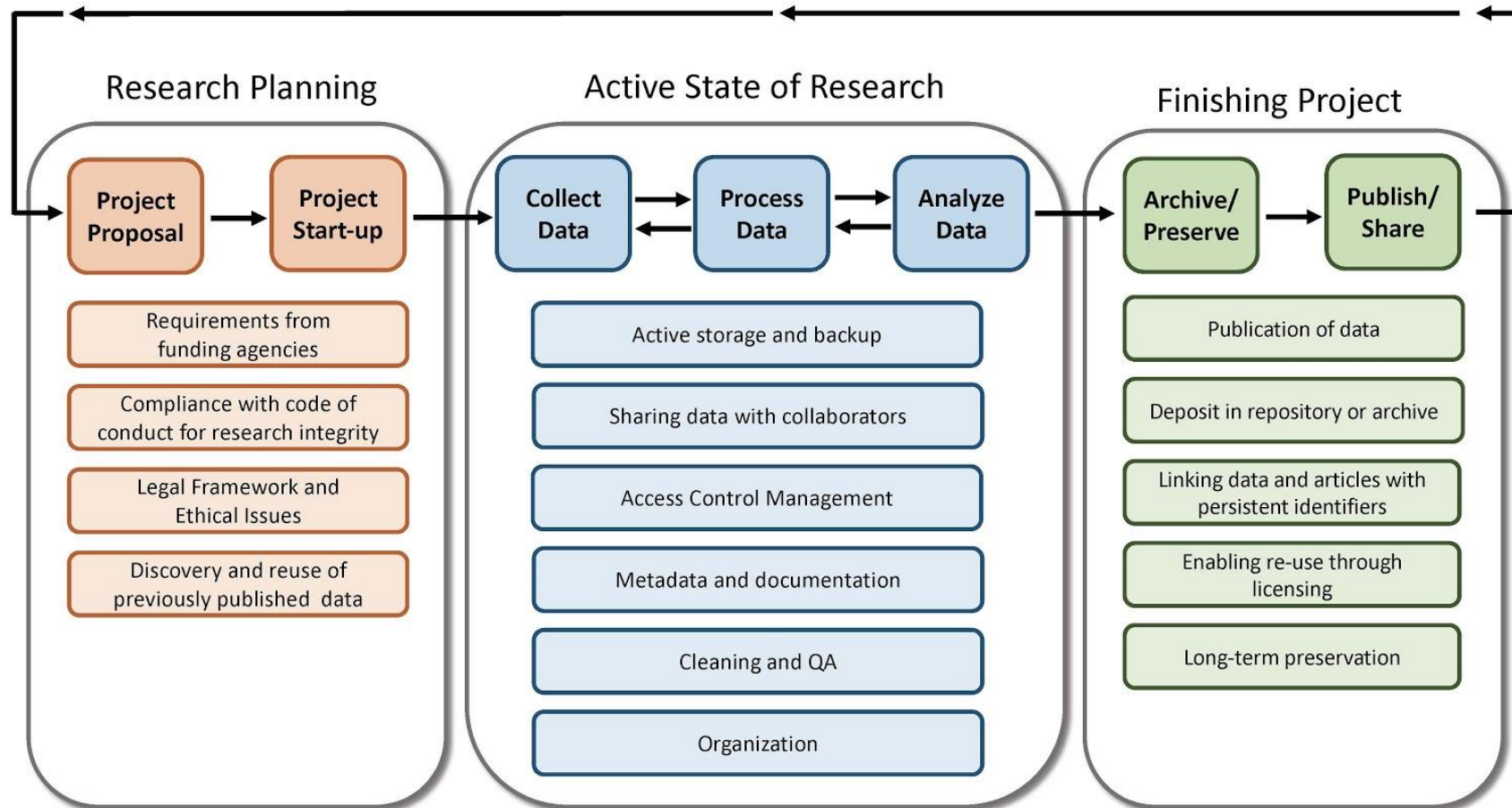
- Descriptive metadata
 - For finding or understanding a resource
- Administrative metadata
 - For long-term management of files
 - For decoding and rendering files
 - For documenting use and access rights
- Structural metadata
 - For documenting the relationships of parts of resources to one another

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

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?

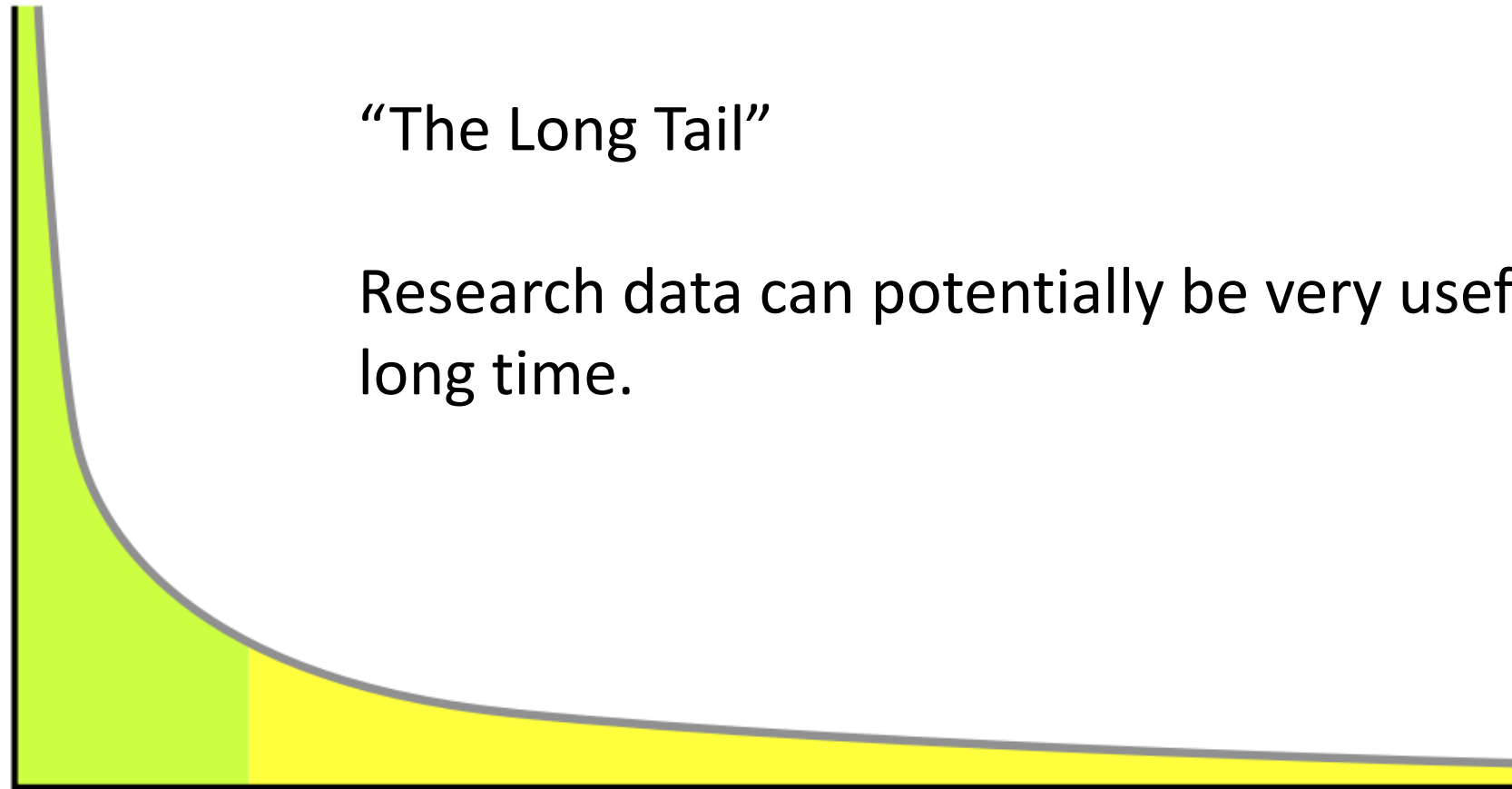


Why should researchers care?

Scholarly Communication:

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

Why should researchers care?



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

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-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](#)

[Understanding Metadata: What is Metadata, and What is it For?: A Primer](#)

[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?