ETH zürich

ETH Library

Coffee Lectures

Metadata Standards for Documenting your Research

Dr. Fabian Schmid

9 February 2022, Webinar via Zoom

DISCLAIMER

Goals

You will understand the importance of metadata for ensuring reproducibility in research.

You know where to find metadata standards suitable for your scientific discipline.

What are metadata?

Image: Adobe Stock #96808114

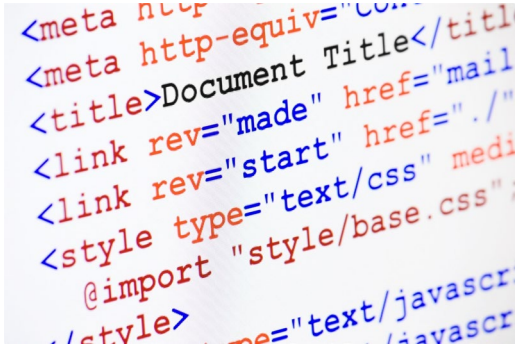


Image: Adobe Stock #174698805

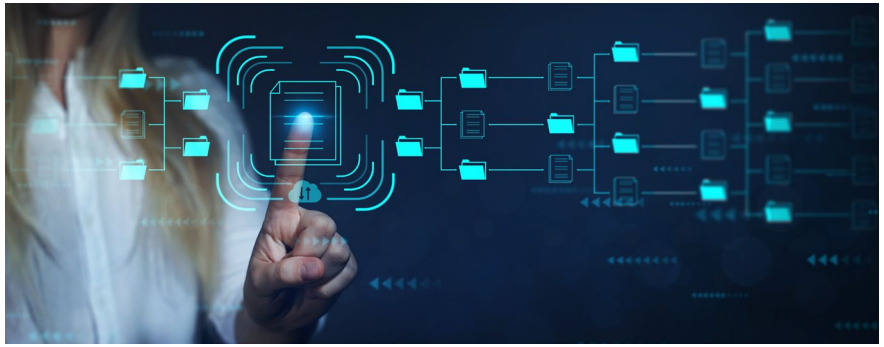


Image: Adobe Stock #465625340

- The term '**metadata**' comprises all **auxiliary information** which describe the **characteristics** of a set of data (i.e. data about the actual dataset).
- Metadata were initially **used by librarians** and refer to descriptions of **digital objects**
- Widely-used standard for electronic resources: **Dublin Core Metadata Basics** (15 generic elements: e.g., title, creator, persistent identifier, publisher, …)
- Metadata make a set of data **intelligible**, **findable** and control data **accessibility.**


Image: Adobe Stock #96808114

Scientific metadata: ensure **verification**, **validation** and **reusability of research data** and contribute to **integrity in research**

Image: Adobe Stock #206233438

SI Base units as a good example for a commonly used metadata standard

The International System of Units (SI): Base units
https://www.bipm.org/en/measurement-units/si-base-units

Name	Typical symbol	Name	Symbol
time	*t*	second	s
length	*l*, *x*, *r*, etc.	metre	m
mass	*m*	kilogram	kg
electric current	*I*, *i*	ampere	A
thermodynamic temperature	*T*	kelvin	K
amount of substance	*n*	mole	mol
luminous intensity	I_v	candela	cd

Most of the scientific metadata standards are more extensive but should still be consistently and widely-used in the research community and beyond.

Scientific Metadata can be more extensive and less standardised

Image: Adobe Stock #98808114

Challenges

- **Scientific data** are created by **experiments** and **observations** and thus require **more diverse types of metadata** (e.g., normal image vs. microscopy image).
- **Scientific user-communities** are very **specialised and the scientific landscape is highly dynamic**.
- **Tacit knowledge** is often necessary for **interpretation and use** of the data **outside of the community** (e.g., preserved for an extended time period).
- **Considerable effort** may be required to define sufficient **metadata for ensuring reusability** of these data, (i.e., making implicit knowledge explicit).

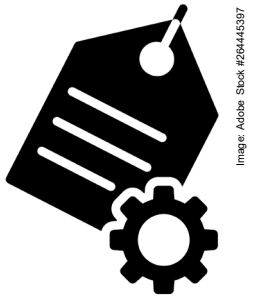
Solutions

- **Disciplinary metadata standards** which should ideally be **unique**.
- Standards should be kept **traceable** and **comprehensible** since they can evolve as fast as the community from which they have initially emerged.
- **Machine-actionable metadata** gathering: Automated process enabling completeness and consistency.

Image: Adobe Stock #280886010 (educational license)

Scientific Metadata Standards for various research fields

Scientific Metadata Standards (examples)

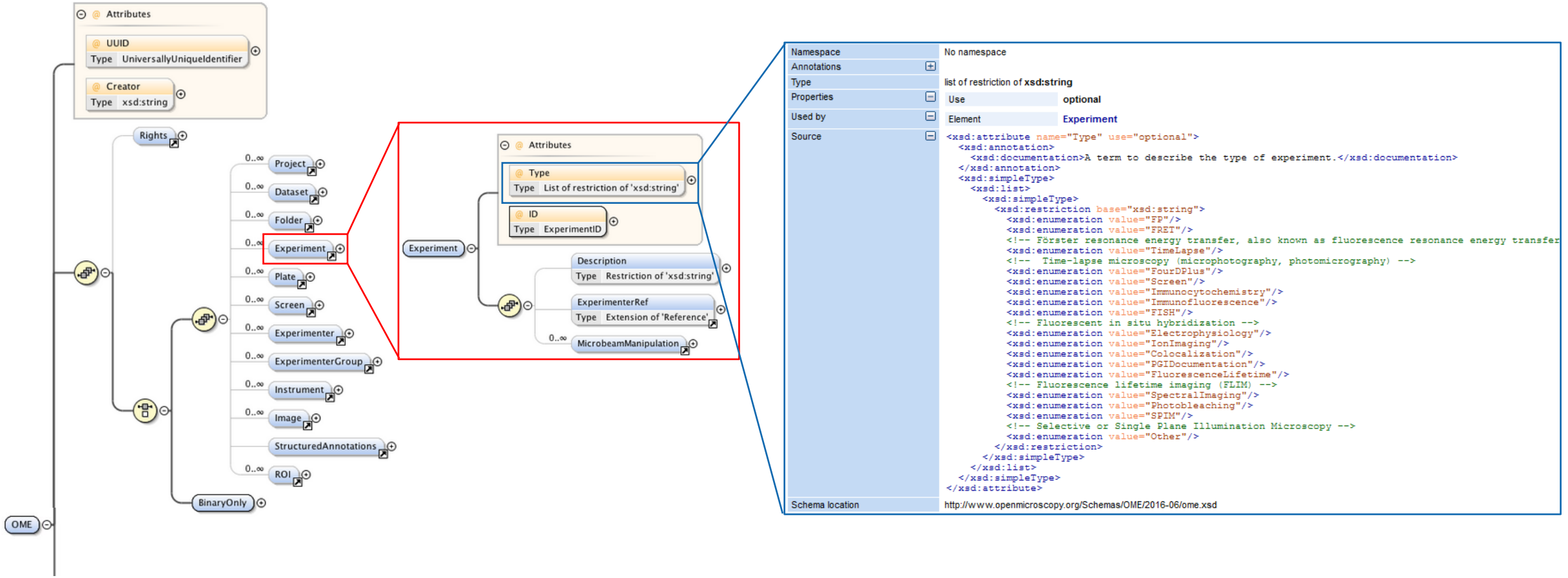


- Biology → Gene ontology, NCBI taxonomy, etc.
- Physical sciences → IUPAC, InChI
- Earth science and ecology → USGS Thesaurus, GIS dictionary, etc.
- Math & computer science → Mathematics Subject Classification, ACM Computing Classification System

Databases for available disciplinary metadata standards

- DCC website (Digital Curation Centre)
- RDA Metadata Standards Directory
- https://fairsharing.org/standards/

Open Microscopy Environment (OME) provides open-source software and standards for microscopy data in TIFF format



Source: https://www.openmicroscopy.org/Schemas/Documentation/Generated/OME-2016-06/ome.html

Selected references

Articles & book chapters

- Allison B. Zhang, Don Gourley, 4 - Metadata strategy, In Chandos Information Professional Series, Creating Digital Collections, Chandos Publishing, 2009, Pages 31-53. https:/doi.org/10.1016/B978-1-84334-396-7.50004-3
- Allison B. Zhang, Don Gourley, 6 - Creating metadata, In Chandos Information Professional Series, Creating Digital Collections, Chandos Publishing, 2009, Pages 73-88. https://doi.org/10.1016/B978-1-84334-396-7.50006-7
- Goldberg, I.G., Allan, C., Burel, JM. et al. The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. Genome Biol 6, R47 (2005). https://doi.org/10.1186/gb-2005-6-5-r47

Websites

- Digital Curation Centre (DCC), Scientific Metadata, Text: Clive Davenhall from national e-Science Centre, https://www.dcc.ac.uk/resources/curation-reference-manual/chapters-production/scientific-metadata
- Dublin Core Metadata Basics: https://www.dublincore.org/resources/metadata-basics/
- Databases for disciplinary metadata standards:
  - Digital Curation Center (DCC), UK: http://www.dcc.ac.uk/resources/metadata-standards/list
  - Research Data Aliance (RDA) Metadata Standards Directory Working Group: http://rd-alliance.github.io/metadata-directory/
  - Fairsharing: https://fairsharing.org/standards/
- Open Microscopy Environment (OME): https://www.openmicroscopy.org/index.html

Dr Fabian Schmid
Data Management Consultant
fabian.schmid@library.ethz.ch

ETH Library
Research Data Management and Digital Curation
HG H 31.6
Rämistrasse 101
8092 Zurich

www.library.ethz.ch/coffee-lectures

Thanks for your attention!

Next Coffee Lecture:

New Open Access Funding Options at the ETH Library

16 February 2022, Webinar via Zoom
www.library.ethz.ch/coffee-lectures