



Overview and Importance of Metadata and Vocabularies

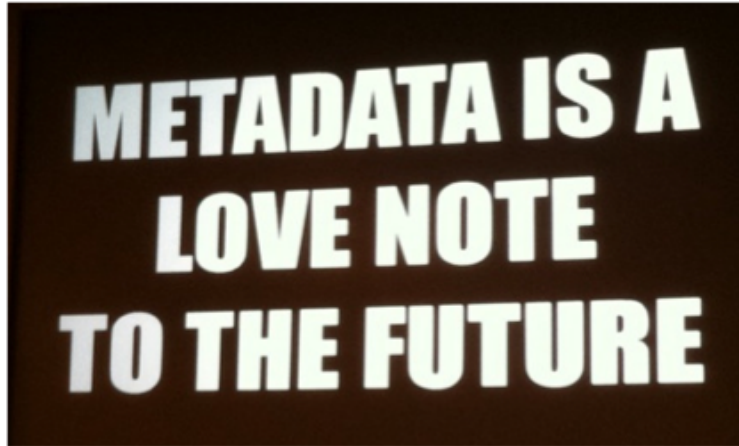
Jochen Schirrwagen, 16-May-2018,
COAR Annual Meeting 2018, Hamburg

Why Metadata



„Metadata, the information we create, store, and share to describe things, allows us to interact with these things to obtain the knowledge we need.“

Source: Jenn Riley (2017). Understanding Metadata: What is Metadata, and What is it For?: A Primer. NISO, ISBN 978-1-937522-72-8, url: <https://www.niso.org/publications/understanding-metadata-2017>



source: cea + from The Netherlands

([https://commons.wikimedia.org/wiki/File:Metadata_is_a_love_note_to_the_future_\(8071729256\)_cropped.jpg](https://commons.wikimedia.org/wiki/File:Metadata_is_a_love_note_to_the_future_(8071729256)_cropped.jpg)) , „Metadata is a love note to the future (8071729256) (cropped)“, <https://creativecommons.org/licenses/by/2.0/legalcode>

Types of Metadata (for knowledge resources)



Type	Goal	Example
Descriptive	For discovery and identification of information resources; interoperability	Dublin Core
Structural	Facilitates navigation and presentation of digital resources	XML
Administrative <ul style="list-style-type: none">• Rights management• Preservation management	Processing and management of digital collections; interoperability	Premis
Technical	How a system functions or metadata behaves; interoperability	METS, JHOVE (tool)

Types of Meta Formats

According to the W3C Web Architecture we can distinguish

- XML – e.g. to markup content; labelled trees with annotations and cross-ref.
- RDF – for expressing graph structured information
- JSON – for expressing tree based data structures

Meta formats can have a schema language to define vocabularies

- E.g. XML Schema or RELAX NG to validate XML documents
- RDF Schema or OWL Ontology to describe RDF vocabularies

Meta formats can have a query or transformation language to retrieve or process information described in such formats

- XPath, XSLT, XQuery for XML
- SPARQL for RDF

Sharing Metadata

- OAI-PMH: exposing metadata records
- ResourceSync: synchronizing web-resources based on sitemap extension
- Linked Data
 - Publishing of structured data in the web
 - Data can be linked to other's data to become a global information network
 - Example: wikidata,
https://www.wikidata.org/wiki/Wikidata:Data_access#Linked_Data_interface
 - Design issues (<http://www.w3.org/DesignIssues/LinkedData.html>)
 - Use URIs as names for things
 - Use HTTP URIs so that people can lookup those names
 - Useful information should be provided when an URI is looked up
 - Links to other URIs to discover more and other things

Influence of the Linked Data concept



- Towards (meta-)data openness
- Focus on the graph as a whole vs. bounded and distinct sets of data from single sources
- (deeply) connecting data(sets) from multiple sources resulting in larger knowledge graphs

Why Controlled Vocabularies



- An organized arrangement of words and phrases used to tag information in a consistent way, to organize knowledge for subsequent retrieval (sources: getty, wikipedia)
- May have a defined scope or describe specific domain
- May have preferred, alternative and variant terms

Types of Controlled Vocabularies

- Relationships, e.g. between things
- Subject heading lists
- Controlled lists to control terminology
- Synonym ring lists
- Authority Files
- Taxonomies
- Classification schemes, e.g. DDC
- Thesauri
- Ontologies
- Folksonomies

Motivation & Challenges regarding Controlled Vocabularies in Repository Metadata



- Aimed for improving quality and consistency of metadata
 - To capture & tame the proliferation of heterogeneous terms & standards

Challenge	Approach
International context <ul style="list-style-type: none">• but regional distinctions• E.g. by language, communities	Consideration of <ul style="list-style-type: none">• Multilinguality• Terms may have limited scope
Numerous vocabulary/ontology initiatives referring to the same/similar thing	Support of of linking, mapping, crosswalking
Interoperability	Adoption of modern (web) standards to describe vocabularies

Guidelines and Application Profiles for Repository Metadata



- Co-existence of a metadata-zoo
 - Platforms with a distinct scope, e.g. for research information (CERIF), for publications and research data (DC, DataCite)
 - Specific local or regional requirements, (JPCOAR, LA Referencia)
 - Different technological levels and consumers, e.g. for machines, bots (triples) or humans (descriptive)
 - Evolutionary process to incorporate new developments and requirements
- Controlled Vocabularies are an essential building block to align application profiles, to make them interoperable