

Finding Ways Between Metadata Models

Mari Kleemola (FSD), Katja Moilanen (FSD),
Maja Dolinar (ADP), Irena Vipavc Brvar (ADP)



YHTEISKUNTATIETEELLINEN
TIETOARKISTO



EDDI18 – 10th Annual European DDI User Conference

4.12.2018, Berlin

This work is licensed under Creative Commons
Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)) License.
Exceptions: see last slide.

Contents

- Metadata jungle
- The need for FAIR metadata
- Richness of different metadata models (relevant to CESSDA)
- Approaches to metadata
- Examples
- Criteria for selecting metadata standard
- Surviving Metadata Model Jungle

The presentation builds on the work of CESSDA's Metadata Management 1&2 and DataverseEU 2018 projects, and Data Without Boundaries project.

Metadata jungle



List of metadata standards:
<http://www.dcc.ac.uk/drupal/resources/metadata-standards>

The need for FAIR metadata

FINDABILITY

- F3. (meta)data are registered or indexed in a searchable resource.
- F2. data are described with rich metadata.
- F4. metadata specify the data identifier.
- F1. (meta)data are assigned a globally unique and eternally persistent identifier.

ACCESSIBILITY

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

INTEROPERABILITY

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

REUSABILITY

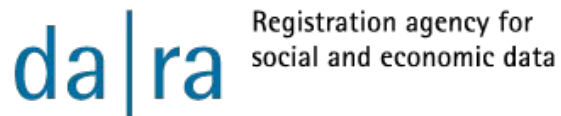
- R1.3. (meta)data meet domain-relevant community standards.
- R1. (meta)data have a plurality of accurate and relevant attributes.
 - R1.2. (meta)data are associated with their provenance.
 - R1.1. (meta)data are released with a clear and accessible data usage license.

Richness of different metadata models

CITATION METADATA



+ METHODOLOGY & DATASET DESCRIPTION, PUBLICATIONS



+ VARIABLE DESCRIPTIONS



Less rich metadata

Rich metadata

DISCOVERY

UNDERSTANDABILITY

REUSABILITY

Different approaches for different purposes

Lowest common denominator

- 👍 relatively easy to produce
- 👍 enables discovery and citation
- ❑ limits on resource discovery for more specialised disciplinary concepts
- ❑ desperate users due to massive search result lists
- ❑ very limited support for re-use or understanding data

Rich disciplinary metadata

- ❑ more effort to produce
- 👍 includes discovery metadata
- 👍 enables granular searches (incl. question and variable level)
- 👍 happy users :)
- 👍 ensures re-usability (also over time)

Some examples

- CESSDA Catalogue <https://datacatalogue.CESSDA.eu/>
- OpenAire <https://explore.openaire.eu/search/find/datasets>
- Google data search <https://toolbox.google.com/datasetsearch>
- CLOSER Discovery
<https://discovery.closer.ac.uk/item/uk.cls.ncds/be89bbd8-4708-4539-a2c5-cc1e073c30fd>

Selecting metadata standard

Provenance and reputation	<p>A stable and proactive organisation should maintain the standard.</p> <p>The background of the standard needs to be documented (who created the standard, what are the governance procedures, organisational viability and stability, what methods are used to update and maintain the standard, history of the standard).</p> <p>The standard should have a documented history of success in relevant fields.</p>
Accessibility	<p>The standard should have a user community (a standard that is not used has no value).</p> <p>The standard should be openly available.</p> <p>The standard should be well documented, and there should be practice guides available.</p>
Completeness and conformance to expectations	<p>The standard should contain all the relevant information elements and functions that are needed to describe the resource.</p> <p>The standard should contain the elements that the user community could reasonably expect to find there.</p>
Granularity	<p>The standard should make all the necessary distinctions and be sufficiently detailed.</p>
Accuracy	<p>The standard should have a way to control the contents of the elements (for example, a mechanism to restrict the values of the elements to match a given regular expression).</p>
Logical consistency and coherence	<p>The standard should support use of controlled vocabularies.</p>
Timeliness	<p>The standard should support updating and versioning of both metadata and the described objects.</p> <p>The standard should support using PIDs.</p>
Interoperability	<p>The standard should be mapped to other relevant standards.</p> <p>The standard should be well documented to allow its integration in larger automatised infrastructures.</p>
Adaptability to local needs	<p>The standard should allow customisation without losing its integrity.</p> <p>Optionality should be a feature of most parts of the standard.</p>
Surrogacy	<p>The standard should allow for both embedded and linked metadata.</p>
Training	<p>There should not be an unreasonable amount of training required to be able to use the standard.</p>
Tools	<p>There should exist tools that fulfil the requirements of potential users (researchers as well as metadata creators).</p>

Source: Data without Boundaries Integrated DELIVERABLE D7.2 - D7.3: [Standards with future relevance for European Social Science data infrastructure & Needs, Key Areas, Rules & Best Practices in Metadata Standard selection and usage](#) (2014), p. 24-28.

Surviving Metadata Model Jungle

- Collect metadata that your organization and your partners need
- Remember: You probably need to provide different metadata with different models for different purposes
- Metadata is the thing, not the used model!
- “Master Metadata”
 - the original metadata
 - not (necessarily) dependent on the used model(s)
 - export the required parts of the master metadata in the wanted metadata formats (derived metadata)

Thank you for your
attention!

Exceptions to CC BY 4.0 license: logos in slides 1 and 5.