

Specific requirements on the da | ra metadata schema

Nicole Quitzsch

GESIS - Leibniz Institute for the Social Sciences

Workshop: “Metadata and Persistent Identifiers for Social
and Economic Data”

07-08 May 2012, Berlin

Topics

1. Metadata for Research Data
2. Development of the da | ra metadata schema
3. Research Data for Social and Economic Sciences - Specific requirements
4. Properties of the da | ra metadata schema
5. Conclusion

Metadata for Research data

Objectives:

- Improving the retrieval of documents / objects
- Increase the retrieval quality
- facilitate data exchange and interoperability with other metadata standards
- Faster location of the primary documents using reliable resource indicators (DOIs, URNs,...)

Development of the da | ra metadata schema

Conditions:

- Membership in the consortium DataCite
- Delivery of all required information on DataCite
- Mapping to DataCite metadata schema

da | ra schema:

- Contains elements for content and technical description of social and economic research data

Research Data for Social and Economic Sciences - Specific requirements

DataCite schema: "lowest common denominator" of different disciplines

da | ra-schema:

- Extended according to the DataCite metadata schema
- DataCite elements + specific elements of social science + economic research data
- Domain specific possibilities for retrieval and research
- Provides more extensive description of research data
- lays foundations for consistent citation of data

Research Data for Social and Economic Sciences - Specific requirements

- Adaptation to conditions of the disciplines in terms of specific properties → **specific metadata elements**
- **specific development tools** → allowed vocabularies: thesauri and classifications of social and economic sciences: GESIS Classification Social Sciences, Standard Thesaurus Economics, Social Science Thesaurus

Properties of the da | ra metadata schema

36 elements:

- 8 descriptive mandatory elements
Title, Principal Investigator, Publication Agent, Registration Agency, DOI, URL, publication date, availability
- +4 Administrative elements
- 24 optional items, e.g.:
Keywords, classifications, dataset, collection mode, geographic coverage, temporal coverage, selection method, time dimension, other titles, other identifiers, related publications, technical description of the data

da | ra mandatory properties

ID	Property Name	Mapping to DataCite	Definition	Occ
1	Title	Title	Title of the dataset.	1
3	Principal Investigator (Name and/or Institution)	Creator (type = Data Collector)	Name and/or Institution of the Principal Investigators	1-n
4	Publication Agent	Publisher	Name of the publication agency for the resource.	1
5	Registration Agency (Homepage, Contact, E- mail)	<i>Contributor (type = Registration Agency)</i>	Name of the registration agency ("da ra").	1

da | ra mandatory properties

ID	Property Name	Mapping to DataCite	Definition	Occ
6	DOI	Identifier (type = DOI)	Persistent Identifier (DOI) assigned to the resource.	1
7	URL		Uniform Resource Locator that will be registered with the DOI.	1-n
11	Publication Date	Publication Year	Date the dataset was made publicly available by the publication agent.	1
28	Availability (controlled)	<i>Rights</i>	Description under which conditions the data is available.	1

Properties of the da | ra metadata schema

Principles

For fields with controlled content always an extra field for free content

17.1	Geographic Coverage (controlled)	Universe.areaControlled	Geographic units on which the study focuses. These are taken from a controlled vocabulary geographic names authority list.		ISO 3166-2/3, UN/LOCODE
17.2	Geographic Coverage (free)	Universe.areafree	Geographic units on which the study focuses (free).	Ability to assign geographic units free if they are not available in the controlled vocabulary, eg West Berlin	

Controlled vocabularies/syntax: standards (DataCite vocabularies, DDI vocabularies, ISO/DIN), own vocabularies have been developed only where necessary.

Properties of the da | ra metadata schema

Information Schema documentation:

- identifier of the elements in German and English
- Definitions of the elements
- Details of the commitment (M, MA, O)
- Repeatability of the elements
- Vocabulary encoding schemes (controlled vocabularies)
- Syntax Encoding Schemes (Controlled syntax)
- Language dependence of the element
- Data type of element
- Editing of fields by da | ra or data encoder

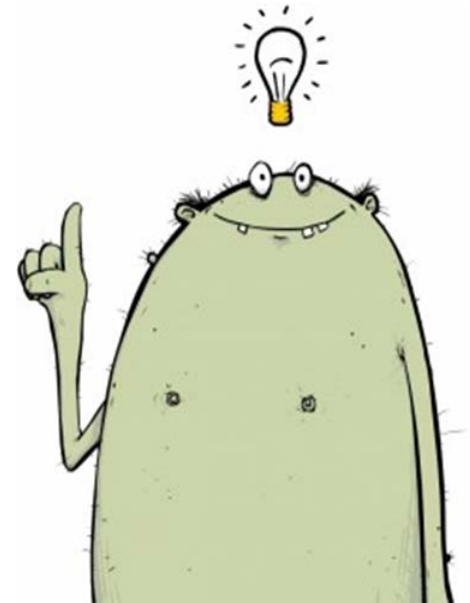
Properties of the da | ra metadata schema

Goals:

- Ensure the quality of metadata
- Interoperability (to support interoperability between different metadata formats and different technical systems)
- Further development of the mappings to DataCite, DDI and Dublin Core

Conclusion

- includes mandatory fields of the DataCite schema
- da | ra is compatible with the standard DDI
- Interoperability → controlled vocabularies and mappings to DDI and Dublin Core
- Improves the visibility of research data from the Social and Economic Sciences



Thank you for your attention!

Nicole Quitzsch GESIS–Leibniz-Institute for the Social Sciences
nicole.quitzsch@gesis.org