Collaborative Notes

Data Description and Metadata - What it takes to produce a good one?

Date: Wednesday 8 December 2021 10:00 - 12:30 CET
https://www.fsd.tuni.fi/en/news/events/webinar2021-12-08/

Programme

- Introduction
    -
- CESSDA and its services related to data description and metadata production and use
    - If your organisation dont use ISO 8601, make the move right now :)
    - https://xkcd.com/1179/
- Significance of data description quality for cross-national data catalogues
    -
- Describing data in practice: FSD as a case example
    -
- Metadata in (re)use: harvesting, licensing, repurposing and FAIR
    -
- Discussion and Q & A

Questions covered during the discussion

Q: Are there any good templates to use to assist in describing metadata?
A: Not that we know of. Various submission forms provide a structure, but they are not usually open. Best suggestion is to look at any catalogue entry, for example in Aila, where one can see all the elements used and the content entered. E.g. FSD2316 Electoral Alliances in Parliamentary Elections 1945-2007

Q: What is the authoritative source for DDI vocabularies?
A: When the pipeline is finalised it will be DDI Alliance Vocabulary Service. At this point DDI website has older versions, latest versions are on cessda's vocab service.

Q: Is the standard for service the same as for the dataset ?

A: An API service is capable of serving harvesters with many types of metadata. It does not care if the metadata is about data or books etc. It may be useful that you serve metadata in various formats via your API service because different aggregators may use different metadata models. Therefore some mapping is usually needed. With OAI-PMH providing DublinCore is required.

[Protocol for Metadata Harvesting - v.2.0](#) is used in many archives regardless of content. It can express records in multiple metadata schemas.

Q: How long does it take to produce a metadata description at FSD?
A: Depends on the dataset. Metadata production takes a day or two. This is not including any other things that need to be done, like anonymisation, data cleaning.

Q: Who assigns the Topic Classification - authors themselves or database administrators?
A: Data producers at FSD. Depositors approve the final metadata description and therefore the topics too.

Q: How do you define the owner/author of a dataset?
A: Advised to decide and agree on in advance. Definition for the owner: "The person, corporate body, or agency responsible for the work's substantive and intellectual content." FSD requires contact details to someone who can act on the dataset, and decide that it can be archived (and in the case of those dataset that are disseminated only on a request, contact details of a person granting access.
About authorship: https://tenk.fi/en/advice-and-materials

Q: Who curates the data sets on the variable level? FSD? The researchers? FSD together with the researchers?
A: Together, but FSD takes the main responsibility. FSD also has the responsibility to produce metadata needed for long term preservation and relevant technical metadata. Remember to describe the data, not the study or research results.

Q: Are there best practices regarding data ingestion and curation within the CESSDA consortium? Regarding workflows specially.
A: See CESSDA's DMEG ([Data management expert guide](#)) for valuable information. Most archives have similar ingest and curation principles, but there is variation in depth. There has been training and learning from others. This could be elaborated. Repositories need to be open about their policies, practices and use of DMPs. Different resources in different archives play a central role in how detailed or extensive the processes can be. Certified trusted digital repositories need to offer, preferably public, documentation about these processes. See also eg.
https://www.fsd.tuni.fi/en/data-archive/documents/records-management-and-archives-formation-plan/

Q: Preservation metadata, how is it done? How do you update these datasets?
A: FSD uses a national digital preservation solution and relies on digital preservation specifications set by the service. Descriptive DDI metadata is included in the METS document descriptive metadata section as it is. In addition we produce technical metadata, like size of the tile, checksums, image or video file information (MIX, MODS). We use a very

minimal provenience section (PREMIS) and rights metadata (because data are only disseminated via FSD and extensive rights metadata doesn't have to travel with the archival package. Process is semi automated at the moment. Updates are incremental and it would be possible to pull previous versions from digital storage too.
For specifications and details: https://digitalpreservation.fi/en

Q: CESSDA Data Catalogue metadata model?
A: CESSDA Data Catalogue - DDI Codebook Profile This is the CESSDA Data Catalogue metadata model in a table format for DDI 2.5 version but please note that this is only at study level (excluding variable information). Archives often have and need more information,so the catalogue metadata is only a sub-set. It lacks some very relevant elements.
https://cmv.cessda.eu/profiles/cdc/ddi-2.5/1.0.4/profile.xml

The full CESSDA metadata model: CMM CESSDA Metadata Model