

Challenges and benefits of implementing the CESSDA Metadata Model in an existing environment

Alina Danciu, Alexandre Mairot

▶ **To cite this version:**

Alina Danciu, Alexandre Mairot. Challenges and benefits of implementing the CESSDA Metadata Model in an existing environment. EDDI 2019 TAMPERE – 11TH ANNUAL EUROPEAN DDI USER CONFERENCE, Dec 2019, Tampere, Finland. pp.12. hal-03621558

HAL Id: hal-03621558

https://sciencespo.hal.science/hal-03621558

Submitted on 15 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

SciencesPo

Challenges and benefits of implementing the CESSDA Metadata Model in an existing environment

Alina Danciu, Alexandre Mairot

EDDI19- 11th Annual European DDI Users Conference
Tampere, Finland
December, 3 - December, 4 2019

Creative Commons Attribution 4.0 International (CC BY 4.0)

Center for Socio-Political Data (CDSP)

- Service unit of Sciences Po and of the French National Center for Scientific Research (CNRS)
- SSH data curation since 2005 : more than 320 surveys
- Member of PROGEDO : national infrastructure for SSH data
- PROGEDO is the CESSDA French Service Provider

PROGEDO
DATA INFRASTRUCTURE

SciencesPo
CENTER FOR SOCIO-POLITICAL DATA

cessda

Review of procedures



- How to make data **FAIRer**?
- How to make them more **visible** and more **reused**?
- How to implement less **time-consuming documentation protocols**?
- How to **avoid information loss** (example : one colleague is leaving)
- How to communicate with stakeholders ?
- How do you exchange metadata protocols with other institutions (for example : the library of your university)

Roles defined at the CDSP

- Preservation responsible
- Metadata standards responsible
- Storage and archiving responsible
- Information system responsible
- Data users (ingest and publishing data) responsible

The DDI Alliance Controlled vocabularies

- Using 7 controlled vocabularies (Analysis Unit, General Data Format, Mode Of Collection, Sampling Procedure, Time Method, Topic Classification, Type Of Instrument)
- Translation in French of the CVs
- We worked with the CESSDA Service Provider (SODA) from Belgium to revise translation
- Published now in French : **Feel free to use them !**



The CESSDA Metadata Model

- CDSP not an official part of the project but **tested the different versions of the CMM**
- Find it useful for our particular case
- Our metadata are DDI-C with a “hint” of the DDI-L philosophy
- We have a **Nesstar server** and a in-house developed **DDI-C question bank**
- Currently implementing a new repository, Dataverse
- It was vital to have a “common language” and the CMM was it
- We used elements of the CMM to come up with a CDSP **metadata model** (mandatory, recommended elements)

Benefits

- Easier to speak to the management and partners: “CESSDA is recommending this”
- We had clear recommendations to follow and adapt to our particular situation
- We finally began to document thoroughly our metadata choices!

In some cases, we got a little bit lost in the multitude of recommandations and elements. CMM is pretty complex ! Also, we were not sure at a certain point that the CMM version we are using won’t change.

CDSP's metadata model (mandatory elements)

Title	Full title by which the Dataset is known.
Author	The person(s), corporate body(ies), or agency(ies) responsible for creating the work.
Contact	The contact(s) for this Dataset.
Description	A summary describing the purpose, nature, and scope of the Dataset.
Subject	Domain-specific Subject Categories that are topically relevant to the Dataset.
Keyword	Key terms that describe important aspects of the Dataset.
Topic Classification	The classification field indicates the broad important topic(s) and subjects that the data cover
Language	Language of the Dataset
Producer	Person or organization with the financial or administrative responsibility over this Dataset
Production Date	Date when the data collection or other materials were produced
Distributor	The organization designated by the author or producer to generate copies of the particular work including any necessary editions or revisions.
Distribution Date	Date that the work was made available for distribution/presentation.
Unit of Analysis	Basic unit of analysis or observation that this Dataset describes
Universe	Description of the population covered by the data in the file
Sampling Procedure	Type of sample and sample design used to select the survey respondents to represent the population
Collection Mode	Method used to collect the data; instrumentation characteristics

NEXT STEPS

- Finish documenting all our procedures
- Reviewing all our workflows and our internal documents
- Preparing for CoreTrustSeal

Get in touch!

alina.danciu@sciencespo.fr
alexandre.mairot@sciencespo.fr



Follow us on Twitter!
@cdsp_ScPo
@alina_danciu_
@alexandremairot

CDSP's metadata (CV)

Field	Definition	Controlled Vocabulary
Keyword	Key terms that describe important aspects of the Dataset.	ELSST
Topic Classification	The classification field indicates the broad important topic(s) and subjects that the data cover. Library of Congress subject terms may be used here.	CESSDA Topic Classification
Kind of Data	Type of data included in the file: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, or other.	DDI Alliance GeneralDataFormat
Unit of Analysis	Basic unit of analysis or observation that this Dataset describes, such as individuals, families/households, groups, institutions/organizations, administrative units, and more. For information about the DDI's controlled vocabulary for this element, please refer to the DDI web page at http://www.ddialliance.org/controlled-vocabularies.	DDI Alliance AnalysisUnit
Time Method	The time method or time dimension of the data collection, such as panel, cross-sectional, trend, time- series, or other.	DDI Alliance TimeMethod
Sampling Procedure	Type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction.	DDI Alliance SamplingProcedure
Collection Mode	Method used to collect the data; instrumentation characteristics (e.g., telephone interview, mail questionnaire, or other).	DDI Alliance ModeOfCollection
Type of Research Instrument	Type of data collection instrument used. Structured indicates an instrument in which all respondents are asked the same questions/tests, possibly with precoded answers. If a small portion of such a questionnaire includes open-ended questions, provide appropriate comments. Semi-structured indicates that the research instrument contains mainly open-ended questions. Unstructured indicates that in-depth interviews were conducted.	DDI Alliance TypeOfInstrument

CDSP's metadata model (recommended elements)

Alternative Title
Other ID
Related Publication
Production Place
Contributor (Related Person/Data Collector)
Grant Information
Depositor
Deposit Date
Time Period Covered
Date of Collection
Software
Related Material
Related Datasets

Geographic
Geographic Unit
Time Method
Frequency
Major Deviations for Sample Design
Type of Research Instrument
Characteristics of Data Collection Situation
Actions to Minimize Losses
Weighting
Cleaning Operations
Response Rate