

CESSDA Work Plan 2020

CESSDA Metadata Office (MDO) Task 1

D2b: Report on webinar

Webinar of the CESSDA Metadata Office – status quo, future developments & special focus on CMM – May 6th, 2020

Document info

Dissemination Level	PU
Due Date of Deliverable	31/05/20
Actual Submission Date	04/06/20
Type	Report
Approval Status	Approved (Technical and T&S WGL, 14/07/2020)
Version	V1.1
Number of Pages	p.1 – p.12

Version history

Version	Date	Comment	Revised by
V1.0	20/05/20		Kerrin Borschewski
V1.1	04/06/20		Kerrin Borschewski

Author List

Organisation	Name	Contact information
GESIS – Leibniz Institute for the Social Sciences	Kerrin Borschewski	kerrin.borschewski@gesis.org

Peer-review

Organisation	Name	Contact information
Gesis - Leibniz Institute for the Social Sciences	Levin Friedemann	Levin.Friedemann@gesis.org
UKDS	Sharon Bolton	sharonb@essex.ac.uk

Contents

Executive Summary	3
Abbreviations and Acronyms	4
1) General information on the webinar	5
a) Information on participants	5
b) Materials of the webinar	5
2) Webinar content	5
a) CESSDA ERIC	6
b) The CESSDA Metadata Office (MDO)	6
c) European Language Social Science Thesaurus (ELSST)	7
d) CESSDA Vocabulary Service (CVS)	7
e) The CESSDA Metadata Model (CMM)	7
f) DDI Profiles	8
g) CESSDA Metadata Validator (CMV)	8
3) Questions and Answers from Q&A session	9
4) Further reading	12

Executive Summary

This report aims to summarize the relevant information on the CESSDA Metadata Office webinar as an additional written record to the video recording. The “Webinar of the CESSDA Metadata Office – status quo, future developments & special focus on CMM” was conducted on the 6th of May 2020. This webinar is part of the CESSDA Metadata Office Task 1 Workplan for 2020.

The first section of the report contains overview information on the webinar. The second part provides details on the content that was presented during the webinar. The third part captures questions and comments of attendees, as well as the responses to those questions. Section number four comprises further reading proposals.

Abbreviations and Acronyms

CDC	CESSDA Data Catalogue
CMM	CESSDA Metadata Model
CESSDA MO	CESSDA Main Office
DDI	Data Documentation Initiative
EQB	CESSDA Euro Question Bank
MDO	CESSDA Metadata Office
SP	Service Provider

1) General information on the webinar

On Wednesday the 6th of May 2020 (11 am- 12 pm CEST), the CESSDA Metadata Office Task 1 conducted a project webinar "Webinar of the CESSDA Metadata Office – status quo, future developments & special focus on CMM" in the scope of its workplan for 2020.

The speakers were:

- Kerrin Borschewski (GESIS - Leibniz Institute for the Social Sciences), <https://orcid.org/0000-0002-0284-7024>
- Esra Akdeniz (GESIS - Leibniz Institute for the Social Sciences), <https://orcid.org/0000-0001-5022-2216>
- Darren Bell (UK Data Service), <https://orcid.org/0000-0002-4152-6549>
- Alexander Mühlbauer (GESIS - Leibniz Institute for the Social Sciences)
- Jeannine Beeken (UK Data Service),
- Sharon Bolton (UK Data Service), <https://orcid.org/0000-0003-3727-3689>
- Taina Jääskeläinen (Finnish Social Science Data Archive)

a) Information on participants

86 people from the following 27 countries registered: Australia, Austria, Bangladesh, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Lithuania, Mexico, Netherlands, Norway, Portugal, Romania, Serbia, Slovenia, Sweden, Switzerland, Uganda, United Kingdom, United States. Actual 55 participants joined the live webinar.

b) Materials of the webinar

Materials on the webinar are available on:

- Slides on Zenodo: <http://doi.org/10.5281/zenodo.3776809>
- Video on Zenodo: <http://doi.org/10.5281/zenodo.3776809>
- Video on CESSDA Training YouTubeChannel: <https://youtu.be/qYrZn8xUBy8>

2) Webinar content

The webinar contained three parts:

- 1) Introduction to the CESSDA Metadata Office (MDO) and MDO services
- 2) The CESSDA Metadata Model (general information, changes between CMM v0.1 and CMM v1.0, MDO's efforts to facilitate the provision of quality metadata)
- 3) Questions and Answers

The first part contained an overview of the CESSDA Metadata Office project. MDO presented an introduction to CESSDA ERIC and the CESSDA Metadata Office. Furthermore, the current work of MDO and in detail the MDO services European Language Social Science Thesaurus

(ELSST) and CESSDA Vocabulary Services (CVS) were presented. The second part focused on the CESSDA Metadata Model (CMM). It contained general information on the CMM, as for example, the information captured in the CMM, the use of DDI3.2 for the CMM, and the controlled vocabularies used by CMM. In the following, the Excel spreadsheet of the CMM (v1.0) was presented to the participants. There was a slide on the changes that had occurred between CMM v0.1 and CMM v1.0. The work on the DDI profiles and the CESSDA Metadata Validator (CMV) were introduced, and the last slide contained an outlook on the work the MDO intends to accomplish in 2020.

During the third part, the questions from the participants were addressed.

a) CESSDA ERIC

[CESSDA](#) stands for Consortium of European Social Science Data Archives, and ERIC stands for European Research Infrastructure Consortium. CESSDA is an alliance of social science data archives with currently 20 member countries (Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Netherlands, North Macedonia, Norway, Portugal, Serbia, Slovakia, Slovenia, Sweden, Switzerland, and United Kingdom). CESSDA aims to promote social science research results and to support national and international research and cooperation. CESSDA creates & runs services for the social sciences. It brings together social science data archives across Europe.

b) The CESSDA Metadata Office (MDO)

The CESSDA Metadata Office project oversees strategic developments for all metadata-related issues within CESSDA and the implementation of CESSDA tools and services. CESSDA ERIC has acknowledged the importance of metadata with several metadata projects over the years. The latest of these projects is the CESSDA Metadata Office project, which started at the beginning of 2019. It is a continuation of its predecessor, the CESSDA Metadata Management project. The CESSDA Service Providers (SPs) working on the MDO are UKDS (United Kingdom) and GESIS (Germany) with a joint lead, FSD (Finland), and NSD (Norway).

MDO forms a conceptual and strategic group to maintain and manage CESSDA's metadata-related material. The CESSDA Metadata Office is the principal point of contact for metadata issues within CESSDA. MDO stays up to date with metadata requirements of CESSDA SPs and CESSDA projects, while simultaneously providing recommendations to CESSDA SPs on metadata. Furthermore, MDO monitors metadata developments in other relevant consortia, institutions, initiatives, and projects. MDO defines strategic metadata requirements for CESSDA, to help plan as well as implement tools and services.

Additionally, MDO manages the content of the European Language Social Science Thesaurus (ELSST) and related multilingual vocabulary services (for further information, see below). The Metadata Office regularly updates the metadata materials it is responsible for (such as the

[CESSDA Metadata Model](#)). Those materials are based on the [Data Documentation Initiative \(DDI\)](#) metadata standard.

c) European Language Social Science Thesaurus (ELSST)

[European Language Social Science Thesaurus \(ELSST\)](#) is a broad-based, multilingual thesaurus for the social sciences. ELSST keywords describe in detail the actual subjects and concepts covered by data.

The latest version of the multilingual ELSST was released in September 2019. The new release contains two significant changes. Firstly, a new language, Dutch, was added to the thesaurus. Secondly, scope notes were added to the Slovenian version. ELSST is now available in 14 languages, Czech, Danish, Dutch, English (source language), Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Slovenian, Spanish and Swedish. Content development and structural work in 2019 have concentrated on reducing complex polyhierarchies and revising the POLITICS hierarchy. A full list of changes can be found on the [Changes to ELSST](#) webpage, and in a post on the [CESSDA ELSST blog](#).

d) CESSDA Vocabulary Service (CVS)

The [CESSDA Vocabulary Service \(CVS\)](#) provides a user-friendly source of standardized controlled vocabularies. Currently, there are 24 controlled vocabularies available in the CVS, including the CESSDA Topic Classification. The vocabularies can be browsed online and downloaded in different formats. The CVS also contains an editor, where authorized users create, manage and translate the vocabularies. The service provides Uniform Resource Names (URN) for both the controlled vocabulary and for each version of it, as well as an Application Programming Interface (API).

The majority of the source (English) vocabularies included in the service have been created by the DDI Alliance. Translations of the DDI vocabularies have been provided by CESSDA members and associated organizations in different countries, as they use the vocabularies to describe research data. The other languages that can be found in the CVS are Danish, Finnish, French, Italian, Norwegian, Portuguese, Serbian, Slovenian, and Swedish.

MDO has also released the CVS User Guide. It includes information for the general user, and instructions for editors on how to add, translate, and maintain a controlled vocabulary (CV) in the tool. The User Guide is available online to registered CVS users. It will be revised and updated periodically.

e) The CESSDA Metadata Model (CMM)

MDO maintains CESSDA's official metadata schema, the CESSDA Metadata Model. The first version of the CMM was produced within the CESSDA Metadata Management project. It contains elements relevant for CESSDA, the CESSDA SPs, and CESSDA tools. The CMM is built from the viewpoint of quantitative social science data. Thus, it aims to help CESSDA Service

Providers to make their quantitative data more discoverable and understandable to users. The CMM is based on the DDI Lifecycle metadata standard because it is currently the most comprehensive standard for the social sciences and promotes interoperability. The CMM was produced because the different CESSDA Service Providers use different kinds of standards and have different ways to document their data. In CMM, CESSDA agreed on one way of documentation. Service Providers who want to have their metadata represented in CESSDA tools, like CDC and EQB, need to make their metadata compatible with the metadata schemas of those tools. The tools metadata schemas are consistent with the CMM.

The information covered in the CESSDA Metadata Model is information on Study, Persons, Institutions, Dataset, Survey Instrument, Questions and Responses, Concepts, Group of Studies, Further documents, publications, and the document description. The new version of CMM also includes mapping for the elements of the current version of the CESSDA Data Catalogue metadata schema. The CMM Metadata Schema structure is based on the principle for the reusability of metadata elements.

The CMM is a model; this means it defines a global set of objects and metadata supported by CESSDA. The CMM is useful because it reduces the complexity of the whole DDI specification and allows the representation of the DDI objects that are important for CESSDA.

The CMM is updated on an annual basis. The first versions of the CMM and the CMM User Guide were published (in May/June 2019) via Zenodo. The [latest version of the CMM](#) was published (in November 2019) via Zenodo. The document can be downloaded and cited using this DOI: <https://doi.org/10.5281/zenodo.3236171>.

From the first published version, version 0.1, to the current version of the CMM "version 1.0," there have been 481 changes in total; 227 of those were significant changes, and the remaining 254 were minor changes.

f) DDI Profiles

For a better understanding of the CDC and EQB metadata schemas, the CESSDA Metadata Office is currently working on [DDI profiles](#) for the metadata schemas of the CESSDA Data Catalogue (CDC) and the CESSDA Euro Question Bank (EQB).

MDO will produce DD1.2.2, DDI2.5, and DDI3.2 profiles for the CDC, as well as DDI2.5 and DDI3.2 profiles for EQB. The Metadata Office is also working on a UML class model based on the CMM.

The profiles and the UML model are intended to promote the understanding and the implementation of CESSDA's metadata schemas for CESSDA SPs.

g) CESSDA Metadata Validator (CMV)

To further assist the production of quality metadata by the SPs, the Metadata Office is currently working on a CESSDA Metadata Validator (CMV).

The CESSDA Metadata Validator will be an open-source tool for CESSDA's metadata providers and CESSDA tools. The CMV's functions are aimed to cater to the needs of CESSDA. With the CMV, it will be possible to validate metadata documents and check if they correspond to the metadata specifications of CESSDA tools, like the CESSDA Data Catalogue and the CESSDA Euro Questions Bank.

3) Questions and Answers from Q&A session

The participants had several opportunities to pose questions or comments concerning the webinar, the work of MDO, and the webinar's content. Before the webinar, it was possible to ask a question during the registration under the section "Do you have any questions for the speaker?" or by sending an email to the Metadata Office. MDO received two questions from the registration form. Furthermore, the participants could pose their questions and comments during the whole webinar, using the Chat/Questionbox that GoToWebinar provides for participants. All issues were discussed and responded to during the webinar. Those questions, comments and the according answers will be displayed in the following (questions and comments are displayed verbatim):

- 1) Question: "What metadata is mandatory? Any tools for publishing/searching research data with CESSDA metadata? For Social Science only?"

Answer: The list of the mandatory elements of CMM can be found in the published version of CMM (<https://doi.org/10.5281/zenodo.3236171>). The mandatory elements are marked with an M for "mandatory" in column D "Status" on the second Excel Sheet "CESSDA Metadata Model". Some elements are only mandatory for specific tools, e.g., the Euro Question Bank or the CESSDA Data Catalogue. This is illustrated in the model with additional information in the field (e.g., "M for EQB").

- 2) Question: "How can we adapt the CMM to Nesstar (for example to have PID there - in which field can we put it? - as it should appear in CDC)."

Answer: According to the CESSDA Data Access Policy and the CESSDA PID Policy (<https://doi.org/10.5281/zenodo.3611316>) the use of global PIDs for CESSDA Service Providers will be mandatory by the end of 2020. Therefore, PID will be a compulsory element in the metadata schemas of CESSDA tools (EQB and CDC) and the CMM by then. If it is not possible to document a PID element in Nesstar, other options need to be considered (for example, the adaption of other DDI formats that allow the provision of a PID, e.g., DDI Codebook).

- 3) Comment from attendee: "Hi all, what we did in order to have a PID using DDI 2 was to use Dataverse instead of Nesstar (Center for Socio-Political Data)."

- 4) Question: "Is ELSST available as LOD? rephrased later: Sorry, Linked Open Data (in RDF Format for example)."

Answer: The current ELSST software is nearing the end of its life. ELSST doesn't have an API, and its linked open data opportunities are limited at the moment. But ELSST is being moved to a new ontology management platform (VocBench) this fall. Once

ELSST is based in VocBench it will be linked open data, and there will be an API as well. It will be available in RDF and SKOS format. The license for ELSST will be a CC-license.

- 5) Question: "I believe ELSST used to have restricted access, but it is now publicly accessible. Is that correct?"

Answer: Currently, ELSST is available publicly to everyone for browsing. Its access is restricted if an institution wants to use it in its own systems or do translations. For this, a license agreement (without any charge) needs to be entered. The license agreement is currently with the University of Essex, but when ELSST is moved to the new software and is published on the CESSDA website, it will move to a CC-license.

- Question: "Thank you for the webinar. One question re: the Cessda Vocabulary Service; I didn't quite get whether or not its content is available via API now, or when/if it is planned."

Answer: There is an API in place for the CVS (CESSDA Vocabulary Services). For ELSST, there will be both a RESTful endpoint and an RDF-XML endpoint as well. In terms of the CESSDA Metadata Model, there is no API.

- 6) Question: "Could I hear more on the CMM XSD / UML model?"

Answer: The information for the CMM was collected over the last few years in an Excel Spreadsheet, as a starting point. The simplest way to standardize the mappings, later on, is to start with an XSD specification (should be ready by the end of June). Since the XSD isn't going to be used in any practical sense to generate XML documents, MDO will create a UML model that will be able to convert into XMI and automatically generate documentation as needed. A UML model is a mechanism to define all the possible object attributes that CESSDA would like to use in the model. It can map to different DDI schemas. It's not a schema that is implemented, but a model, that makes it possible to make associations between, e.g. DDI2.5 or DDI3.2 elements and demonstrate equivalences.

- 7) Question: "Are the profiles published somewhere?"

Answer: The profiles will be published as soon as they are finished and approved.

- 8) Question: "I'm not entirely sure I understand the rationale of the CMM. I guess it is twofold: 1) provide some CESSDA-specific guidelines regarding dataset/study metadata in general, drawing on the opportunity of having several social science experts working together thanks to CESSDA, and 2) prepare a technical standard that can help with later developments of tools such as the CESSDA Data Catalogue. Is that correct?"

Answer: The CESSDA Metadata Model acts as a definition of supported elements from various DDI specifications. The model is an abstraction of all the different DDI elements that CESSDA supports. Since DDI Lifecycle is an extensive specification, it is helpful to define a supported subset of DDI elements. Rather than making it a DDI3.2 specification, CESSDA decided to abstract the model for its uses. The CMM is more documentation after the fact than a specific artifact to help with the implementation

of metadata. The practical implementation and guidance of metadata is done by the DDI-profiles.

- 9) Question: "Is any qualitative data archive already using the CMM? or would it be possible to use it also for including Metadata Records from qualitative Data Archives in CESSDA?"

Answer: Many of the elements in CMM are common between qualitative and quantitative datasets. FSD has been using the CMM / DDI for years to document qualitative data as well as UKDS. The CESSDA Data Catalogue contains already some qualitative datasets.

- 10) Question: "There seem to be mandatory CMM elements and others. Is there a distinction in the others group like strongly recommended and others?"

Answer: The CMM differentiates between mandatory, recommended, and optional elements. Within the mandatory elements, there is a distinction between mandatory – meaning mandatory in any case – mandatory for DDI3.2 – which refers to elements that are specific to the DDI3.2 specification – mandatory for CDC – elements that are compulsory if the metadata is to be included in the CESSDA Data Catalogue – and mandatory for EQB – which means that the use of those elements is mandatory if the metadata is to be included in the CESSDA Euro Question Bank. There are no differentiations for different recommended types or optional types. Please note, if a mandatory element in question is a child element, its use is only mandatory if the parent element is used.

- 11) Question: "I'm struggling to understand, in CMM, the difference between element 1.1.1 "Study ID" and element 1.1.2 "Study Number." Maybe because it's DDI-Lifecycle-specific? We use DDI 2.5 (Dataverse) in Belgium."

Answer: Element 1.1.1 is an element specific to the DDI3.2 structure (Definition "Identifier of the study according to DDI3.2 structure"). Element 1.1.2 however, refers to the inhouse ID of an archive (Definition: "Unique archival number - if there is no unique archival number available, a PID may be used alternatively (if a PID is used as study number, the information for this PID should however still be captured in the element 1.1.10.4 Study PID)").

- 12) Comment from attendee: "CVs API link? <https://vocabularies.cessda.eu/swagger-ui.html#>".

- 13) Question: "Is Dataverse compatible with ("ready for") CMM mandatory fields? That is, if we use Dataverse, will metadata be compatible with CMM?"

Answer: Written answers from attendee who is working in the SSHOC project to these questions (verbatim):

- "Within the SSHOC project there are tools available to convert between NESSTAR and Dataverse.
- Yes, Dataverse is compatible with CMM and the CESSDA data catalogue. We are working within the SSHOC.
- Within SSHOC we are working on the languages.

- Please send questions about Dataverse to marion.wittenberg@dans.knaw.nl".
- See also documents of webinar from SSHOC: Webinar slides: <https://doi.org/10.5281/zenodo.3714988>; Webinar recording: <https://youtu.be/f7-r-80M-Fk>
- 14) Comment from attendee: "Thanks, I am from Qualiservice (Qualitative Data Archive in Germany) and we also use DDI 3.2 for our records so a mapping with the CMM is encouraged, we use these IDs."

4) Further reading

Barbalet, Suzanne; Bolton, Sharon (2019): Sustainable European Multilingual Vocabularies: A Model for Cooperation in Metadata Management among European Data Archives. Available online at https://ukdataservice.ac.uk/media/622472/iassist19_sbarbalet_sbolton_02.pdf, checked on 05/14/2020.

Borschewski, Kerrin; Förster, André; Friedrich, Tanja; Zenk-Möltgen, Wolfgang; Miranda, Patrícia; Moura Ferreira, Pedro et al. (2019): CMM CESSDA Metadata Model. DOI: 10.5281/zenodo.3236171.

Consortium of European Social Science Data Archives (CESSDA) (2020): The CESSDA Consortium. Available online at <https://www.cessda.eu/About/Consortium>, checked on 05/14/2020.

Förster, André, Kerrin Borschewski, Sharon Bolton, and Taina Jääskeläinen. 2020 (Forthcoming). "The matter of meta in research data management: Introducing the CESSDA Metadata Office Project." IASSIST Quarterly (44, 3).

Green, Ann E.; Humphrey, Chuck (2014): Building the DDI. In IASSIST Quarterly 37 (1-4), pp. 36–44. DOI: 10.29173/iq500.

Gregory, Arofan; Heus, Pascal; Ryssevik, Jostein (2009): Metadata. In RatSWD Working Paper Series 57, pp. 1-22. DOI: 10.2139/ssrn.1447866.

Hoyle, Larry; Castillo, Fortunato; Clark, Benjamin; Kashyap, Neeraj; Perpich, Denise; Wackerow, Joachim; Wenzig, Knut (2011): Metadata For The Longitudinal Data Life Cycle. DOI: 10.3886/DDILongitudinal03.

Jääskeläinen, Taina; Bolton, Sharon (2019): CESSDA Vocabulary Service – For managing vocabulary content. DOI: 10.5281/zenodo.3600174.