



CESSDA Metadata Office

November 2020

*****In case you have missed the newsletter from march, you can find it at the end of the November newsletter*****

1 CESSDA Metadata Office (MDO)

The CESSDA Metadata Office (MDO) project oversees strategic developments for all metadata-related issues within CESSDA and the implementation of CESSDA tools and services. The CESSDA Service Providers 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. For more detailed information on the MDO, see the newsletter of March 2020 attached to the end of this newsletter.

2 CESSDA Metadata Office Webinar (May 2020)

On the 6th of May 2020, the CESSDA Metadata Office (MDO) conducted a webinar on the status quo and future developments with a special focus on the CESSDA Metadata Model ([CMM](#)).

The webinar was divided into two parts. The first part contained an overview of the CESSDA MDO project, the current work, and services of the CESSDA MDO. The second part is especially interesting for those working with the CMM or interested in the CMM. This part contained general information about the CMM and the changes that have been made between the first published version of CMM (v0.1) and the current published version of CMM (v1.0). MDO also presented information on how it tries to facilitate the work with CMM in the future, e.g., by producing DDI profiles and the CESSDA Metadata Validator (CMV).

The (video) [recording and the slides](#) of the webinar can be found on Zenodo using the following DOI: 10.5281/zenodo.3776808. The video of the webinar can also be found on [youtube](#). MDO has produced a [webinar report](#) with the DOI: 10.5281/zenodo.4072183. This report summarizes the relevant information on the CESSDA MDO webinar as an additional written record to the video recording. The report contains overview information on the webinar, provides details on the content presented during the webinar, captures the questions and comments of attendees as well as the responses to those questions, and contains information on further reading proposals.

3 ELSST (European Language Social Science Thesaurus)

The [European Language Social Science Thesaurus](#) (ELSST) is a broad-based, multilingual thesaurus for the social sciences, owned and published by CESSDA. Freely accessible and covered by a CC-BY-SA 4.0 license, it is now available on the CESSDA Platform as one of a suite of web-based tools. Previously hosted by the UK Data Service, ELSST moved to CESSDA in November 2020.

ELSST consists of over 3,000 concepts (keywords) and covers the core social science disciplines: politics, sociology, economics, education, law, crime, demography, health, employment, information and communication technology, and increasingly, environmental science. It is used for data discovery within CESSDA and facilitates access to data resources across Europe, independent of domain, resource, language, or vocabulary.

ELSST provides a structured, controlled vocabulary of keywords that can be assigned to diverse items within a collection, such as datasets or publications, to describe in detail subjects and topics covered. Using a thesaurus such as ELSST, data providers (such as data archives, research centers, and libraries) can attach similar keywords to items in their collections that are different in nature but cover similar topics. This ensures that materials in the collections are described in a consistent way. In turn, this helps researchers in two ways; making catalogue searches easier by providing a set list of keywords to choose search terms from and improving the accuracy of search results by ensuring that all items covering the selected topic are found.

ELSST is developed by a team drawn from several CESSDA Service Providers, who collaborate to ensure ELSST concepts remain current and relevant to researchers across the social sciences. It is currently available in 14 languages: Czech, Danish, Dutch, English, Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Slovenian, Spanish, and Swedish. We hope to add an Icelandic translation during 2021.

Two multilingual, open-source development tools are used to manage ELSST. Both are based on the [Simple Knowledge Organization System](#) (SKOS), a data-sharing standard model used to share and link web-based, controlled, structured vocabularies. ELSST's background thesaurus management system, used for concept editing, update, and translation, is [VocBench](#), a collaborative development platform for managing thesauri and other vocabularies, developed by the University of Rome Tor Vergata. SKOS output generated from VocBench feeds the online presentation of ELSST via [Skosmos](#), a SKOS-based web browser and publishing tool developed by the National Library of Finland. Both systems offer excellent linked open data possibilities, including API facilities, enabling the potential matching of ELSST with other major international thesauri.

4 CESSDA Vocabulary Service

The [CESSDA Vocabulary Service](#) provides user-friendly browsing and downloading of controlled vocabularies as well as an editor for managing the vocabularies. Currently, there are 28 controlled vocabularies available in the service. Of these, 24 vocabularies are typologies of research methods created by the [DDI Alliance](#). The other vocabularies include CESSDA Topic Classification and vocabularies used for harmonizing the metadata in the CESSDA Data Catalogue.

Some vocabularies in the service have only the English terms and definitions, but many, particularly those included in the CESSDA Metadata Model, have various language versions, up to 11 languages. There are plans to provide the first non-European language in the near future, namely Japanese. The service has been transferred to a new software platform in 2020, with Java Spring-Boot backend and Angular 9 frontend. The Beta version 2.0.0 was released in October and is already used for managing and publishing the vocabularies. An updated API is also available.

5 CESSDA Data Catalogue

The [CESSDA Data Catalogue](#) contains the metadata of research data holdings of CESSDA Service Providers and associated data repositories. It allows the discovery of social science and health-related data from different countries, facilitating comparative, cross-national, and longitudinal

research. The catalogue currently holds 36,386 metadata records in various languages provided by 13 organizations.

Work is ongoing to develop an updated version of the catalogue, which will be released in January 2021. The amendments will include improved filters and enhanced search and display options.

6 CMM (CESSDA Metadata Model)

The CESSDA Metadata Office (MDO) maintains CESSDA's official metadata schema, the CESSDA Metadata Model (CMM), which is based on the [Data Documentation Initiative](#) standard, more specifically, the Lifecycle specification version 3.2. The first version of the CMM was produced within the CESSDA Metadata Management project. It contains elements relevant for CESSDA, the CESSDA Service Providers, and CESSDA tools.

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>. Please note that this DOI is collective, meaning that it will always link to the document's latest version. MDO has been working on a new version of the CMM, which is currently under review. As soon as it has been thoroughly reviewed, the new version will be published on Zenodo, using the same collective DOI as the older versions.

There have been several changes that help improve the consistency and understandability of the CMM. Among those changes are the inclusion of new elements (e.g., metadata access conditions), the deletion of elements (e.g., translation status for some elements), changes concerning the controlled vocabulary information (e.g., the inclusion of the controlled vocabulary Contributor role), and more. In 2021, there will not be a new version of the CMM, but preparatory work will be done for a new version for the year after.

The [CMM User Guide](#) (DOI: [10.5281/zenodo.3236193](https://doi.org/10.5281/zenodo.3236193)) accompanying the CMM is currently being edited according to the changes of the new CMM version. As soon as the editing process is done and the new version of the CMM User Guide has been reviewed, it will also be published on Zenodo.

7 DDI Profiles for CESSDA Tools, UML Class Model

The CESSDA Metadata Office (MDO) is currently working on DDI profiles (<https://ddialliance.org/resources/ddi-profiles>) for the metadata schemas of the [CESSDA Data Catalogue](#) (CDC) and the CESSDA Euro Question Bank (EQB).

DDI profiles are formal, machine-actionable XML (Extensible Markup Language) documents that specify additional constraints on the content of a DDI XML document, over and above those specified by the document's associated XSD (XML Schema Definition) schema.

[DDI profiles](#) for CDC, based on the DDI Codebook versions DDI2.5 and DDI 1.2.2, have been published on Zenodo in September 2020. Those profiles can be found via the DOI: [10.5281/zenodo.4050123](https://doi.org/10.5281/zenodo.4050123). The published DDI profiles are intended to be used in the context of validating DDI2.5 and DDI1.2.2 documents that are being submitted to CDC from participating CESSDA Service Providers.

DDI profiles contain the rules that metadata documents need to follow. The DDI profiles contain human-readable information, e.g., if the use of an element is mandatory, recommended, or optional, for metadata experts in general and in particular for those applying the DDI profiles. The Profiles furthermore support machine-actionable validation by the CESSDA Metadata Validator (CMV). For further information on the CMV, see below.

MDO is currently working on further DDI profiles for the CDC and also the CESSDA EQB. The additional profiles will be a DDI3.2 profile for the CDC and EQB. MDO will create DDI2.5 and DDI3.2

Profiles for EQB. The MDO is also working on a UML class model based on the CESSDA Metadata Model. The remaining profiles, as well as the UML model, will be completed in the beginning of 2021. They will then be opened for review to the CESSDA Service Providers.

8 CESSDA Metadata Validator (CMV)

The CESSDA Metadata Office has successfully created the CESSDA Metadata Validator (CMV). The CMV is an online and command-line based set of tools. It enables the validation of DDI documents against a given DDI profile. The tool checks if a DDI file is valid according to the rules captured in a DDI profile. If the metadata document (DDI file) breaks the rules, the CMV will display a list of validation errors specifying which rules have been broken. The CMV's primary goal is to improve metadata quality, so that resulting aggregations of such metadata in the CESSDA Data Catalogue (CDC) have improved utility for the research community.

The CMV is currently under review and has been tested by the CESSDA Service Providers. This testing was mostly done using the published CDC DDI2.5 and DDI 1.2.2. Profiles discussed above. However, the CMV is designed to validate any DDI file against any DDI profile. Therefore, it can be widely adopted in the future for many use cases.

9 MDO at the Online EDDI Conference 2020

From the 30th of November to the 2nd of December, the European DDI User Conference ([EDDI](#)) – 12th Annual European DDI User Conference – 2020 conference will take place. Due to the pandemic, the conference will be held online this year. This conference is the annual European meeting for users of the Data Documentation Initiative (DDI) Standard. In the framework of the EDDI Conference, users of the DDI Standard get a chance to meet, discuss questions, challenges, ideas, and development potentials concerning the standard as well as associated tools.

On Wednesday the [2nd of December 2020 \(10:20 am - 10:40 pm CET\)](#), Darren Bell and Kerrin Borschewski from the CESSDA Metadata Office (MDO) will give a presentation at the EDDI 2020, titled "Making models into reality: the DDI Profile journey at CESSDA" (see also [abstract](#)). The first part of the presentation will be an introduction to CESSDA ERIC and MDO, the current work of MDO, as well as an introduction to the CESSDA Metadata Model (CMM). In the second part, general information on the XSD of the CMM and the CMM UML model will be presented. The presentation's focus will be on the third part, which concerns the DDI profiles that MDO has created. The profiles will be introduced, including information on "lessons learned."

10 MDO at large

An [article](#) presenting the work of the CESSDA Metadata Office (MDO) 2019 has been published by the [IASSIST Quarterly](#) journal. The article is a project report on the various activities of CESSDA MDO. It covers metadata models, schema, controlled vocabularies, and thesauri, including MDO's collaboration with the DDI Alliance on multilingual translations of DDI vocabularies for CESSDA Service Providers. Furthermore, the article contains information on the communication, training, and advice provided by MDO, including DDI use across CESSDA. It illustrates the project's impact on the social science and research data management community and offers an outline of the project's future plans.

Full citation information of the article: Förster, A., Borschewski, K., Bolton, S., & Jääskeläinen, T. (2020). The matter of meta in research data management: Introducing the CESSDA Metadata Office Project. *IASSIST Quarterly*, 44(3), <https://doi.org/10.29173/iq970>.

March 2020

1 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's aims are to promote the results of social science research and to support national and international research and cooperation. The Consortium was established in 1976. It started out as an umbrella organization of seven social science data archives. As a leading research infrastructure for the social sciences CESSDA was granted the status of a European Research Infrastructure Consortium (ERIC) by the European Commission in 2017.

The CESSDA members work to improve data access for researchers. For this, CESSDA provides large-scale, integrated and sustainable data services to the social sciences and supports national and international research and cooperation. It brings together social science data archives across Europe.

2 CESSDA Metadata Office (MDO)

The CESSDA Metadata Office (MDO) 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 (MDO) project. The MDO project 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 updates the metadata materials it is responsible for (such as the [CESSDA Metadata Model](#)) on a regular basis. Those materials are based on the [Data Documentation Initiative](#) (DDI) metadata standard.

3 ELSST (European Language Social Science Thesaurus)

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

ELSST was originally based on the monolingual Humanities and Social Science Electronic Thesaurus (HASSET), developed by the UK Data Archive at the University of Essex. It has been enhanced and translated in the course of following projects. From 2019, ELSST has been managed within the MDO

alongside CESSDA's other key metadata assets, the [CESSDA Metadata Model \(CMM\)](#) and the [CESSDA Vocabulary Service \(CVS\)](#) (for further information see below). The ELSST is managed by a dedicated team within the MDO project, in close cooperation with expert translators.

The latest version of the multilingual ELSST was released in September 2019. The new release contains two major 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 (including the source language English), with translation of Preferred Terms 98-100% complete in each target language: Czech, Danish, Dutch, Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Slovenian, Spanish and Swedish. Content development and structural work in 2019 have concentrated on the reduction of complex polyhierarchies, and revision of 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](#).

4 CVS (CESSDA Vocabulary Service)

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 member 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, as well as 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.

5 CMM (CESSDA Metadata Model)

MDO maintains CESSDA's official metadata schema, the CESSDA Metadata Model (CMM). 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 comprehensible to users. The CMM is based on the DDI Lifecycle metadata standard, because it is currently the most complete standard for the social sciences and promotes interoperability.

The CMM is updated on an annual basis. The first versions of the CESSDA Metadata Model (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>. Please note that

this DOI is a collective DOI, meaning that it will always link to the latest version of the document. Currently, the CMM is available in both Excel and PDF format. The CMM User Guide was revised at the end of 2019 and will also be uploaded on Zenodo as soon as it is officially approved.

6 DDI Profiles for CESSDA Tools, UML Class Model and CESSDA Metadata Validator (CMV)

The CESSDA Metadata Office is currently working on DDI profiles (<https://ddialliance.org/resources/ddi-profiles>) for the metadata schemas of the CESSDA Data Catalogue (CDC) and the CESSDA Euro Question Bank (EQB).

The 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 MDO 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. To further assist the production of quality metadata by the SPs the MDO is currently working on the CESSDA Metadata Validator (CMV). This tool will help the CESSDA SPs to validate their metadata against the DDI profiles of the CESSDA tools.

7 MDO at large

In 2019, members of the MDO team contributed to the following conferences and workshops outside of CESSDA, presenting the work of the MDO project:

- [IASSIST 2019](#), Sydney
- [CLOSER workshop](#) 'Understanding metadata management in longitudinal studies', London ([Program](#))
- [Federal Statistical Office Germany colloquium](#) 'Potentials and applications of georeferenced data', Wiesbaden
- Sciences Po workshop 'Documenting longitudinal data', Paris
- [EDDI 2019 conference](#), Tampere

An article presenting the work of MDO 2019 has been accepted for publication by [the IASSIST Quarterly](#) journal and will be published in 2020.

The CESSDA Metadata Office MDO will continue its promotion inside and outside of CESSDA. Presentations on latest MDO work are intended for the [IASSIST conference](#) 2020 in Gothenburg and for the [EDDI conference](#) (the European DDI conference) 2020 in Paris.