

## Metadata Guide

---

In a data catalogue metadata provide information about data resources to facilitate searching for and cataloguing of data. Metadata offer a systematic overview of the data resources and contain e.g. information on the author(s) or keywords on the data. Open and unrestricted access to metadata is essential for effective data use and re-use. To this end, metadata in the AUSSDA Dataverse are always published under the public domain dedication (CC0 1.0 Universal) and may thus be freely and openly accessed and used by the public.

This guide will help you to make your (meta)data **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable (FAIR). By following the FAIR principles, you make research data attractive to other researchers. The description of the deposited data follows international standards set by the Data Documentation Initiative (DDI)<sup>1</sup> and the Consortium of European Social Science Data Archives (CESSDA ERIC)<sup>2</sup>. For several metadata entries, you can select from pre-defined lists of items or words. These lists are referred to as “controlled vocabularies”. This standardisation allows for comparability of the metadata in the entire AUSSDA data holdings and also for comparability within the international CESSDA data catalogue. Therefore, using the controlled vocabularies is mandatory for all datasets in the AUSSDA Dataverse as it is very helpful to re-users of your data.

Here are some tips:

- In Table 1, you can find more information on all metadata fields.
- **Required** fields that you need to fill are marked with a red asterisk \*.
- Applicable to all free text fields: Please do not use any thousands separators (like commas, periods, blank spaces), as they cause problems when metadata are harvested by other data catalogues. For example, instead of writing “12,300 participants answered the survey”, omit the separator and write “12300 participants answered the survey” instead.

### Metadata table with examples

Table 1 gives an overview over the different metadata fields. The table also presents some examples and contains a column that serves as a checklist. We provide links to the related vocabularies in the text and in the footnotes.

You can find an example of a dataset in the AUSSDA Dataverse here:

<https://doi.org/10.11587/8SFV2L>.

Some of the links in the table below are outgoing links to controlled vocabularies that we use to follow CESSDA ERIC standards. Here are some tips on how to navigate these pages:



“Heider, Veronika; Hirsch, Lisa (2024). Metadata Guide. AUSSDA: Version 2.” is licensed under a [Creative Commons Attribution 4.0 International Licence](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0).

<sup>1</sup> <https://ddialliance.org/>

<sup>2</sup> <https://www.cessda.eu/>

## ELSST – The European Language Social Science Thesaurus

The European Social Science Thesaurus (<https://thesauri.cessda.eu/elsst/en/3>) is a very extensive resource. In the upper, right corner you can type in a word you would like to use as a keyword or you can navigate on the left side by either scrolling alphabetically or in the hierarchical view.

Important for you:

- Please choose from *preferred terms*, *broader or narrower or related concepts* (clickable terms). Other terms, like *entry terms*, should not be used to index your dataset. In the *example* in Figure 1 the search word was “political science”. As this is an entry term, the user has been forwarded to the *preferred term* “Politics”.
- If you cannot find a keyword in ELSST, add the term to your abstract text to make it findable in Dataverse.

Choose a **maximum of 14 keywords**. If you have more keywords, choose broader concepts to group keywords.

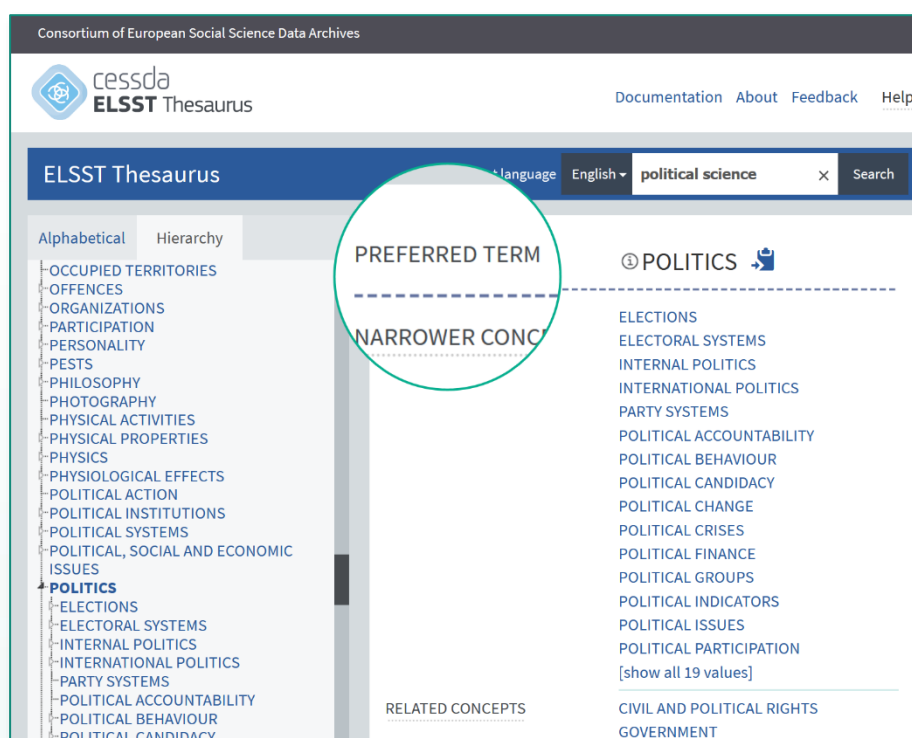


Figure 1 Controlled vocabulary for Keywords: ELSST.


<sup>3</sup> CESSDA and Service Providers (2023) The European Language Social Science Thesaurus (ELSST), <https://elsst.cessda.eu>. DOI: 10.5281/zenodo.8364268

## CESSDA Controlled Vocabularies

For most metadata fields, we use terms from controlled vocabularies that can be found at <https://vocabularies.cessda.eu/>. Each of these vocabularies has its own website including a table with three columns. The terms we use are in the middle column, named “Code descriptive term (en)”. In the column on the right, you can see a description of each term (“Code definition (en)”).

Try to find the most narrow term that fits your project (for example “Face-to-face interview: Paper-and-pencil (PAPI)” in Figure 2, and move to broader terms if you cannot find a fitting one: Choose the broader term “Face-to-face interview” if neither of the narrower terms “Face-to-face interview: Computer-assisted (CAPI/CAMI)” or “Face-to-face interview: Paper-and-pencil (PAPI)” fit.

CVs search > ModeOfCollection v.



**DDI Alliance Controlled Vocabulary for Mode Of Collection**

CV name: Mode Of Collection

CV short name: ModeOfCollection

CV definition: The procedure, technique, or mode of inquiry used to attain the data.

CV notes: This vocabulary was first published by the DDI Alliance. Please see: <https://ddialliance.org/controlled-vocabularies/all>.

Language: English (en) Version 4.0.3

Details	Versions	Identity and general	Usage	Licence and citation	Export/Download
Code value	<b>Code descriptive term (en)</b>		Code definition(en)		
▼ Interview	<b>Interview</b>		A pre-planned communication between two (or more) people - the interviewer(s) and the i		
▼ Interview.FaceToFace	face-to-face interview		Data collection method in which a live interviewer conducts a personal interview, presentin		
Interview.FaceToFace.CAPIorCAMI	Face-to-face interview: Computer-assisted (CAPI/CAMI)	Computer-assisted personal interviewing (CAPI), or computer-assisted mobile interviewing the same device. The administration of the interview is managed by a specifically designed			
Interview.FaceToFace.PAPI	Face-to-face interview: Paper-and-pencil (PAPI)	Paper-and-pencil interviewing (PAPI). The interviewer uses a traditional paper questionnaire			
▼ Interview.Telephone	Telephone interview		Interview administered on the telephone. Use this broader term if not CATI, or if not known		
Interview.Telephone.CATI	Telephone interview: Computer-assisted (CATI)	Computer-assisted telephone interviewing (CATI). The interviewer asks questions as directe			
Interview.Email	E-mail interview		Interviews conducted via e-mail, usually consisting of several e-mail messages that allow th		
Interview.WebBased	Web-based interview		An interview conducted via the Internet. For example, interviews conducted within online fo		

Figure 2 Example of a controlled vocabulary.

Metadata item	Description	Example	Yeah, I got this!
Study title * <sup>4</sup>	Full title of the dataset.	Your title (SUF edition)	
Author *	This is part of how the data set will be cited, so the order of person(s)/author(s) is relevant. Add every person/author's family name, given name or the name of the organization responsible for this dataset as well as the organization with which the person/author is affiliated. In case you have one, you can also provide us with your ORCID (which is a persistent digital identifier (PID) that distinguishes researchers from each other and ensures that your research work is recognized).	Random, Jane (University of Vienna) ORCID 123456	
Abstract *	An abstract describing the purpose, nature, and scope of your dataset.		
Keywords * <sup>5</sup>	This field contains English keyword terms that describe important aspects of the dataset. We use <a href="https://elsst.CESSDA.eu/id/">ELSST</a> . You can select up to 14 keywords to describe your dataset. If you need more keywords, try to choose broader concepts. If you cannot find a keyword in ELSST, make sure that your abstract contains this term to make it findable in Dataverse.	Trust in government Elections Politics	
Topic Classification * <sup>6</sup>	Topics from the Topic Classification indicate the broad important topic(s) and subjects that the data cover. We use the <a href="https://vocabularies.CESSDA.eu/urn/urn:ddi:int:CESSDA:cv:TopicClassification:4.2.2?lang=en">CESSDA Topic Classification</a> . You can use up to 7 topics to describe your dataset.	Public health Censuses Education	
Related Publication	Related publication(s) connected to the dataset, e.g. journal articles, monographs, chapters in books (with doi/ISBN/ISSN/Isid or other persistent id).	Author's Last name, F. M. (Year published). Article title. <i>Periodical Title, Volume(Issue)</i> , pp.-pp. doi: 10.01234/ABC4E. ( <a href="https://doi.org/10.01234/ABC4E">https://doi.org/10.01234/ABC4E</a> )	

<sup>4</sup> Asterisks indicate required fields \*

<sup>5</sup> <https://elsst.CESSDA.eu/id/>

<sup>6</sup> <https://vocabularies.CESSDA.eu/urn/urn:ddi:int:CESSDA:cv:TopicClassification:4.2.2?lang=en>

Metadata item	Description	Example	Yeah, I got this!
Language *	The language(s) of your dataset(s).	German, English	
Grant information: Grant agency and grant number	This is the name of the agency and the grant or contact number that provided funding/sponsored the effort.	H2020: FA123456	
Date of collection: Start *	This field describes the date when the data collection started in the format YYYY-MM-DD. If you are collecting data in multiple waves, we recommend that you add the very first start date and the last end date of your data collection period.	2023-10-22	
Date of collection: End *	This is the date when the data collection ended in the format YYYY-MM-DD.	2024-03-30	
Kind of Data * <sup>7</sup>	Describes the kind of data included in your data file. Possible options are: Numeric, Text, Still image, Geospatial, Audio, Video, Software, Interactive resource, 3D, other.  See descriptions here: <a href="#">Controlled vocabulary Kind of Data</a>	Numeric; Text	
Type of Data Sources * <sup>8</sup>	This field contains information on the type of your data sources. These could be: Population group (for surveys), Communication: Public (for press releases), Research data, Registers/records/accounts, Other, etc. Please note that "Research data" should only be chosen if your data relies on pre-existing research data (e.g. you merge or combine research data).  The full list can be found here: <a href="#">Controlled vocabulary Data Sources.</a>	Population group	
Geographic Coverage *	This field contains information on the geographic coverage of your data. For example: AT (for Austria); Vienna, AT; Luftenberg, Oberösterreich, AT.	Graz, Austria	

<sup>7</sup> <https://vocabularies.cessda.eu/vocabulary/GeneralDataFormat?v=2.0.3&lang=en>

<sup>8</sup> <https://vocabularies.cessda.eu/vocabulary/DataSourceType?v=1.0.2&lang=en>

Metadata item	Description	Example	Yeah, I got this!
Unit of Analysis Type * <sup>9</sup>	Describes the basic unit of analysis or unit of observation in your dataset. Choose from terms such as Individual, Family, Household, Media unit: Text, Media unit: Still image, Organization/Institution, Other, etc.  See descriptions here: <a href="#">Controlled vocabulary Unit of Analysis</a>	Individual	
Universe	Description of the population covered by your dataset; the group of people or other elements that are the object of the study and to which the study results refer to. Age, nationality, and residence commonly help to delineate a given universe, but any number of other factors may be used, such as age limits, sex, marital status, race, ethnic group, nationality, income, veteran status, criminal convictions, and more. The universe may consist of elements other than persons, such as housing units, court cases, deaths, countries, and so on. This term is also known as the universe of interest, population of interest, and target population.	Austrian resident population from age 14	
Time Method * <sup>10</sup>	Describes the time method or time dimension of the data collection, such as Longitudinal, Longitudinal: Trend/Repeated cross-section, Longitudinal: Panel, Time series, Cross-section, Other.  The full list can be found here: <a href="#">Controlled vocabulary Time Method</a>	Longitudinal: Panel	
Sampling Procedure * <sup>11</sup>	This field contains the type of sample and sample design used to select the survey respondents to represent the population. Choose from terms such as Total universe/Complete enumeration, Probability, Probability: Stratified, Non-probability, Non-probability: Quota, Mixed probability and non-probability, Other.  The full list can be found here: <a href="#">Controlled vocabulary Sampling Procedure</a>	Non-probability: Quota	
Method of Data Collection * <sup>12</sup>	Describes the method used to collect your data, such as Telephone interview: CATI, Self-administered questionnaire: Web-based (CAWI), Focus group,	Self-administered questionnaire: Web-based (CAWI)	

<sup>9</sup> <https://vocabularies.cessda.eu/vocabulary/AnalysisUnit?v=2.1.3&lang=en>

<sup>10</sup> <https://vocabularies.cessda.eu/vocabulary/TimeMethod?v=1.2.0&lang=en>

<sup>11</sup> <https://vocabularies.cessda.eu/vocabulary/SamplingProcedure?v=1.1.0&lang=en>

<sup>12</sup> <https://vocabularies.cessda.eu/vocabulary/ModeOfCollection?v=4.0.0&lang=en>

Metadata item	Description	Example	Yeah, I got this!
	Content coding, Transcription, Compilation/Synthesis, Automated data extraction: Web scraping, Other. The full list can be found here: <a href="#">Controlled vocabulary Method of Data Collection</a>		
Type of Research Instrument * <sup>13</sup>	Describes the type of data collection instrument used to collect your data, such as Questionnaire, Structured questionnaire, Interview scheme and/or themes, Data collection guidelines, Programming script, Other. The full list can be found here: <a href="#">Controlled vocabulary Type of Research Instrument</a>	Structured questionnaire	

**Table 1 Metadata fields, examples and links**

<sup>13</sup> <https://vocabularies.cessda.eu/vocabulary/TypeOfInstrument?v=1.1.0&lang=en>