

## Metadata

---

Metadata describe available data resources of an archive 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.

You prepare the metadata that make the 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, we use established “controlled vocabularies” of standardised items. 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 publications in Dataverse. Please do not use thousands separators, as they may cause problems when metadata are harvested by other data catalogues. For example, instead of writing “12,300 participants answered the survey”, do not use the separator and add the text “12300 participants answered the survey”.

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 \*.
- If we ask you for information that you do not have or is not applicable to you, fill the field with “n/a”.
- If you are not sure which category of a controlled vocabulary applies, choose the higher level one.

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 the footnotes.

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

<https://data.aussda.at/dataset.xhtml?persistentId=doi:10.11587/QDETRI>

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:



This document is an extract from “Heider, Veronika; Butzlaff, Iris; Bodlos, Anita; Hönegger, Lisa (2020). Self-Deposit Manual. Self-depositing data at AUSSDA” which is available [here](#) and is licensed under a [Creative Commons Attribution 4.0 International Licence](#) (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/>) 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 need more, try to choose a broader concept.

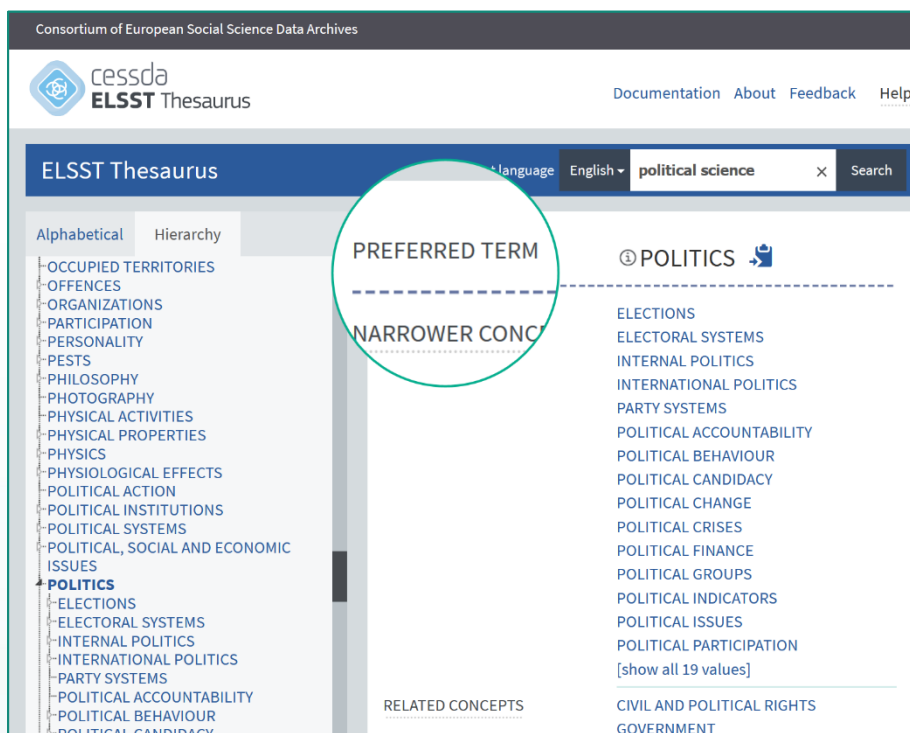


Figure 1 Controlled vocabulary for Keywords: ELSST.

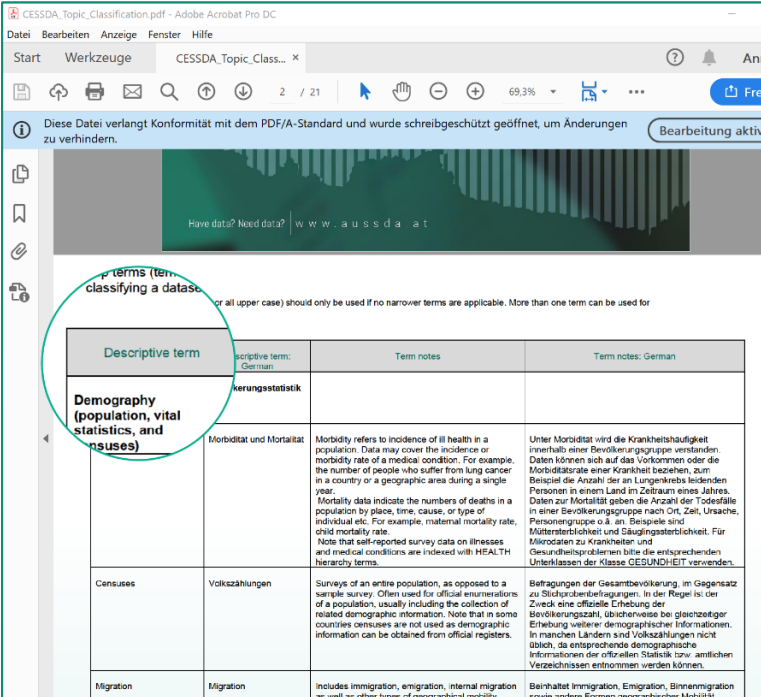
## CESSDA Topic Classification

The CESSDA Topic Classification (<http://aussda.at/cessda-topic-classification>) is not as extensive as ELSST. In the column "Descriptive term" you see broader topics in bold font and sub-topics in normal font.

Important for you:

- Please choose from the column "Descriptive term".
- In the column "Term notes" you can see how each term is described.
- Searching for a topic: Use the search function of your PDF reader to find matching topics (e.g. by starting the search with Strg./Ctrl.+F) or scroll through the document. You may need to download the document first in order to be able to use the search function.

Choose a **maximum of 7 topics**. If you need more, choose a broader topic.



Descriptive term	Descriptive term: German	Term notes	Term notes: German
<b>Demography (population, vital statistics, and censuses)</b>	Bevölkerungsstatistik		
	Morbidität und Mortalität	Morbidity refers to incidence of ill health in a population. Data may cover the incidence or morbidity rate of a medical condition. For example, the number of people who suffer from lung cancer in a country or a geographic area during a single year. Mortality data indicate the numbers of deaths in a population by place, time, cause, or type of individual etc. For example, maternal mortality rate, child mortality rate. Note that self-reported survey data on illnesses and medical conditions are indexed with HEALTH hierarchy terms.	Unter Morbidität wird die Krankheitshäufigkeit innerhalb einer Bevölkerungsgruppe verstanden. Daten können sich auf das Vorkommen oder die Morbiditätsrate einer Krankheit beziehen, zum Beispiel die Anzahl der an Lungenkrebs leidenden Personen in einem Land im Zeitraum eines Jahres. Daten zur Mortalität geben die Anzahl der Todesfälle in einer Bevölkerungsgruppe nach Ort, Zeit, Ursache, Personengruppe o.ä. an. Beispiele sind: Müttersterblichkeit und Säuglingssterblichkeit. Für Mikrosdaten zu Krankheiten und Gesundheitsproblemen bitte die entsprechenden Unterklassen der Klasse GESUNDHEIT verwenden.
	Censuses	Surveys of an entire population, as opposed to a sample survey. Often used for official enumerations of a population, usually including the collection of related demographic information. Note that in some countries censuses are not used as demographic information can be obtained from official registers.	Befragungen der Gesamtbevölkerung, im Gegensatz zu Stichprobenbefragungen. In der Regel ist der Zweck eine offizielle Erhebung der Bevölkerungszahl, üblicherweise bei gleichzeitiger Erhebung weiterer demographischer Informationen. In manchen Ländern sind Volkszählungen nicht üblich, da entsprechende demographische Informationen der offiziellen Statistik bzw. amtlichen Verzeichnissen entnommen werden können.
	Migration	Includes immigration, emigration, internal migration as well as other types of geographical mobility.	Beinhaltet Immigration, Emigration, Binnenmigration sowie andere Formen geographischer Mobilität.

Figure 2 Controlled vocabulary for Topics: CESSDA Topic Classification.

## DDI – Data Documentation Initiative

For some metadata fields, we use terms from the DDI controlled vocabularies. Each of these terms has a website including a table with three columns. The terms that you can copy can be found in the column in the middle with the name “**Descriptive Term of the Code**”. In the column on the right, you can see a description of each term.

DDI Controlled Vocabulary for Kind of Data Format		
<b>Description</b>		
Describes the physical format(s) of the data documented in the logical product(s) of a study unit.		
<b>Details</b>		
<b>Short Name:</b>	KindOfDataFormat	
<b>Long Name:</b>	Kind of Data Format	
<b>Version:</b>	1.0	
<b>Version Notes:</b>		
<b>Version Changes:</b>		
<b>Canonical URI:</b>	urn:ddi-cv:KindOfDataFormat	
<b>Canonical URI of this version:</b>	urn:ddi-cv:KindOfDataFormat:1.0	
<b>Location URI:</b>	<a href="http://www.ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0_Genericode1.0_DDI-CV/Profile1.0.xml">http://www.ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0_Genericode1.0_DDI-CV/Profile1.0.xml</a>	
<b>Alternate format location URI:</b>	<a href="http://www.ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0.html">http://www.ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0.html</a>	
<b>Alternate format location URI:</b>	<a href="http://www.ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0_InputSheet_Excel2003.xls">http://www.ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0_InputSheet_Excel2003.xls</a>	
<b>Agency Name:</b>	DDI Alliance	
<b>Code List</b>		
Value of Code	Descriptive Term of the Code	Definition of the Code
Numeric	Numeric	Data consisting largely of values expressed as digits from 0 to 9 and, optionally, signs for negative values, decimal points, or letters only when intended to represent numbers (for example, A-F or a-f in hexadecimal).
Text	Text	Data consisting largely of text, including letters, numbers, and special characters or symbols used in writing for punctuation, abbreviation, etc. For example, interview transcriptions, narratives or essays written by study participants, newspaper articles, etc.
StillImage	Still image	Static images, such as graphs, drawings, photographs, diagnostic/medical images like X-rays, etc.

Figure 3 Example of a DDI controlled vocabulary: Kind of Data.

Metadata item	Description	Example for survey data	Example for content analysis data	Yeah, I got this!
Study title * <sup>3</sup>	Full title by which the dataset is known.	Your title (SUF edition)	Your title (SUF edition)	☒
Author *	Dataverse will use the same order of the authors in the metadata for creating the recommended citation. 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 are required. Identifiers are optional, for example, ORCID <sup>4</sup> is a persistent digital identifier (PID) that distinguishes researchers from each other and ensures that research is recognised.	Random, Jane (University of Vienna) ORCID 123456	Random, Jane (University of Vienna) ORCID 123456	☒
Abstract *	An abstract describing the purpose, nature, and scope of the dataset. The text is required here, the date field is optional.			☒
Keywords * <sup>5</sup>	Key terms that describe important aspects of the dataset as available in the controlled vocabulary <a href="https://thesauri.CESSDA.eu/elsst/en/">ELSST (European Language Social Science Thesaurus)</a> . You can use up to 14 keywords from ELSST to describe your dataset. If you need more keywords, try to choose broader concepts.	Trust in government Elections Politics	Electoral issues	☒
Topic Classification * <sup>6</sup>	The classification field indicates the broad important topic(s) and subjects that the data cover. You can use up to 7 classifiers from the CESSDA topic classification (controlled vocabulary). If you need more, try to choose broader terms. Here is the list of topics: <a href="http://aussda.at/cessda-topic-classification">http://aussda.at/cessda-topic-classification</a>	Public health Censuses Education	Media	
Related Publication	The information of related publications consists of multiple fields (citation, unique identifiers and URLs) and is optional.	The data are related to the paper "Random, Jane. (YYYY). Impact of X on	The data are related to the paper "Random, Jane. (YYYY). Why content	

<sup>3</sup> Asterisks indicate required fields \*

<sup>4</sup> <https://orcid.org/>

<sup>5</sup> <https://thesauri.CESSDA.eu/elsst/en/>

<sup>6</sup> <http://aussda.at/cessda-topic-classification>

Metadata item	Description	Example for survey data	Example for content analysis data	Yeah, I got this!
		Electoral Behavior". doi: 10.01234/ABC4E. (https://doi.org/10.01234/ABC4E)	analysis is great. doi: 10.01234/ABC4E. (https://doi.org/10.01234/ABC4E).	
Language *	Language(s) of the dataset	German, English	German	
Grant information: Grant agency * and grant number *	The agency that provided funding and the grant number of the project. If this field does not apply to you, please enter "n/a".	Federal Ministry of Science, Research and Economy: HRSM - XYZ4	H2020: FA123456	
Date of collection: Start *	Date when the data collection started in the format (YYYY-MM-DD). If you have several waves for a dataset, we recommend that you add the very first start date and the last end date of your data collection period. You can add more information about the different dates later as additional metadata information.	2019-10-22	2019-10-22	
Date of collection: End *	Date when the data collection ended in the format (YYYY-MM-DD).	2020-03-30	2020-03-30	
Kind of Data * <sup>7</sup>	Choose from list: Numeric, Text, Still image, Geospatial, Audio, Video, Software, Interactive resource, 3D, Other. See descriptions here: <a href="#">Controlled vocabulary Kind of Data (DDI)</a> For all DDI fields: <b>Choose a category that is predominant</b> , for <b>example</b> , if you have one string variable and five numerical variables – choose "Numeric" or if the string is very important choose both "Numeric" and "Text".	Numeric	Text; Other	

<sup>7</sup> [https://ddialliance.org/Specification/DDI-CV/KindOfDataFormat\\_1.0.html](https://ddialliance.org/Specification/DDI-CV/KindOfDataFormat_1.0.html)

Metadata item	Description	Example for survey data	Example for content analysis data	Yeah, I got this!
Type of Data Sources * <sup>8</sup>	A typology of data source that has been used to generate the data. The total list can be found here: <a href="#">Controlled vocabulary Data Sources (DDI)</a> . 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). „Population group" is most likely the right choice if you survey individuals.  <b>Examples:</b> Population group, Communication: Public, Research data, Registers/records/accounts, Other, ...	Population group	Communication: Public	
Geographic Coverage *	Information on the geographic coverage of the data.	Austria	Vienna, Austria	
Unit of Analysis Type * <sup>9</sup>	Basic unit of analysis or observation that this dataset describes. Choose from list: Individual, Organization, Family, Family: Household family, Household, Housing Unit; Event/Process, Geographic Unit, Time Unit, Text Unit, Group, Object, Other  See descriptions here: <a href="#">Controlled vocabulary Unit of Analysis (DDI)</a>	Individual	Text Unit	
Universe *	Description of the population covered by the data in the file; the group of people or other elements that are the object of the study and to which the study results refer. 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	Any news articles in the respective media outlets.	
Time Method * <sup>10</sup>	The time method or time dimension of the data collection, the total list can be found here: <a href="#">Controlled vocabulary Time Method (DDI)</a>	Longitudinal: Panel	Other	

<sup>8</sup> [http://www.ddialliance.org/Specification/DDI-CV/DataSourceType\\_1.0.html](http://www.ddialliance.org/Specification/DDI-CV/DataSourceType_1.0.html)

<sup>9</sup> [http://www.ddialliance.org/Specification/DDI-CV/AnalysisUnit\\_1.0.html](http://www.ddialliance.org/Specification/DDI-CV/AnalysisUnit_1.0.html)

<sup>10</sup> [http://www.ddialliance.org/Specification/DDI-CV/TimeMethod\\_1.2.html](http://www.ddialliance.org/Specification/DDI-CV/TimeMethod_1.2.html)

Metadata item	Description	Example for survey data	Example for content analysis data	Yeah, I got this!
	<b>Examples:</b> Longitudinal, Longitudinal: Trend/Repeated cross-section, Longitudinal: Panel, Time series, Cross-section, Other			
Sampling Procedure * <sup>11</sup>	Type of sample and sample design used to select the survey respondents to represent the population, the total list can be found here: <a href="#">Controlled vocabulary Sampling Procedure (DDI)</a> <b>Examples:</b> Total universe/Complete enumeration, Probability, Probability: Stratified, Non-probability, Non-probability: Quota, Mixed probability and non-probability, Other	Non-probability: Quota	Probability: Stratified	
Method of Data Collection * <sup>12</sup>	Method used to collect the data, the total list can be found here: <a href="#">Controlled vocabulary Collection Mode (DDI)</a> <b>Examples:</b> Telephone interview: CATI, Self-administered questionnaire: Web-based, Focus group, Content coding, Transcription, Compilation/Synthesis, Other	Self-administered questionnaire: Web-based	Content coding	
Type of Research Instrument * <sup>13</sup>	Type of data collection instrument used, the total list can be found here: <a href="#">Controlled vocabulary Type of Research Instrument (DDI)</a> <b>Examples:</b> Questionnaire, Structured questionnaire, Interview scheme and/or themes, Data collection guidelines, Programming script, Other	Structured questionnaire	Data collection guidelines	

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

<sup>11</sup> [http://www.ddialliance.org/Specification/DDI-CV/SamplingProcedure\\_1.1.html](http://www.ddialliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html)

<sup>12</sup> [http://www.ddialliance.org/Specification/DDI-CV/ModeOfCollection\\_3.0.html](http://www.ddialliance.org/Specification/DDI-CV/ModeOfCollection_3.0.html)

<sup>13</sup> [https://www.ddialliance.org/Specification/DDI-CV/TypeOfInstrument\\_1.1.html](https://www.ddialliance.org/Specification/DDI-CV/TypeOfInstrument_1.1.html)