

## Metadata Schema of the CESSDA EuroQuestionBank: Documentation and Publication of Survey Questions in a European Question Bank; Version 1.0

Akdeniz, Esra; Zenk-Möltgen, Wolfgang

Veröffentlichungsversion / Published Version

Verzeichnis, Liste, Dokumentation / list

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Akdeniz, E., & Zenk-Möltgen, W. (2019). *Metadata Schema of the CESSDA EuroQuestionBank: Documentation and Publication of Survey Questions in a European Question Bank; Version 1.0*. (GESIS Papers, 2019/15). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.65593>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

gesis

Leibniz-Institut  
für Sozialwissenschaften

GESIS Papers

2019|15

## Metadata Schema of the CESSDA EuroQuestionBank

Documentation and Publication  
of Survey Questions in a European  
Question Bank

Version 1.0

*Esra Akdeniz & Wolfgang Zenk-Möltgen*



GESIS Papers 2019|14

# **Metadata Schema of the CESSDA EuroQuestionBank**

**Documentation and Publication of Survey  
Questions in a European Question Bank**

**Version 1.0**

*Esra Akdeniz & Wolfgang Zenk-Möltgen*

## **GESIS Papers**

GESIS – Leibniz-Institut für Sozialwissenschaften  
Data Archive for the Social Sciences (DAS)  
Unter Sachsenhausen 6-8  
50667 Köln

E-Mail: [esra.akdeniz@gesis.org](mailto:esra.akdeniz@gesis.org)

ISSN: 2364-3781 (Online)  
Herausgeber,  
Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften  
Unter Sachsenhausen 6-8, 50667 Köln

# 1 Introduction

---

Existing question data banks for social science survey research mainly operate at the national level. National question data banks serve their constituent research communities and only include questions from surveys that have been conducted within their own countries. This limits the pool of available questions to draw from, as well as possibilities for cross-national comparative research, an increasingly important aim of social science investigation. The CESSDA<sup>1</sup> EuroQuestion-Bank (EQB) project implements a cross-national question bank with a central search facility across all CESSDA's survey holdings to display survey questions in different languages in a user-friendly application. A wide variety of surveys, survey questions, and question-related information are provided by CESSDA Service Providers (SPs), who contribute their documentation by using DDI standardized metadata.

Following CESSDA Service Providers cooperate in this project with GESIS

- CSDA Czech Social Science Data Archive (Czechia)
- DANS Data Archiving and Networked Services (The Netherlands)
- DNA Danish National Archives (Denmark)
- FORS Swiss Centre of Expertise in the Social Sciences (Switzerland)
- FSD Finnish Social Science Data Archive (Finland)
- GESIS Leibniz Institute for the Social Sciences (Germany)
- NSD Norwegian Centre for Research Data (Norway)
- SND Swedish National Data Service (Sweden)
- SO.DA.NET/EKKE Greek research infrastructure for the Social Sciences (Greece)
- TÁRKI Social Research Institute (Hungary)
- UKDS UK Data Service (United Kingdom)

This paper outlines functionalities of the EQB application, technical aspects of the EQB architecture and introduces the EQB metadata schema.

## 1.1 The functionalities of EQB

Researchers in the social sciences using EQB will profit from the single point of entry in question discovery that pertains to their research interests. They are able to search fielded questions from several data archives by entering keywords or topics in a search field. The system browses the terms in all available question texts, answer categories, study titles, and post-/prequestion texts. In addition, the platform provides a range of features to improve the search results for the users:

- a filter for series, interview language, country, collection year and mode of collection
- a function to switch between different language versions to view other translations of the same fielded question

---

<sup>1</sup> Consortium of European Social Science Data Archives

- a function to sort result list by relevance, study title or collection year
- a detailed view of the question with question number, question text, question ID/version, pre/post-questiontext, interviewer instruction, response categories and variables
- an overview of the study the question appeared in, including title, number, version, description, CESSDA topics, date of collection, interview language, time method, mode of collection, sampling procedure, metadata provider and publisher
- links to study related documents like survey instruments, codebooks or method reports
- a link to the CESSDA Data Catalogue (CDC) to explore additional study information
- a comparison of questions and relations between questions in a study
- a function to download questions and all additional information in different formats

## 1.2 The EQB architecture

The main goal of EQB from the technical side of view is to build a stable and maintainable productive search web application with content from CESSDA SPs and other suppliers. The current EQB architecture consists of the EQB front-end and several EQB back-end components (see fig. 1).

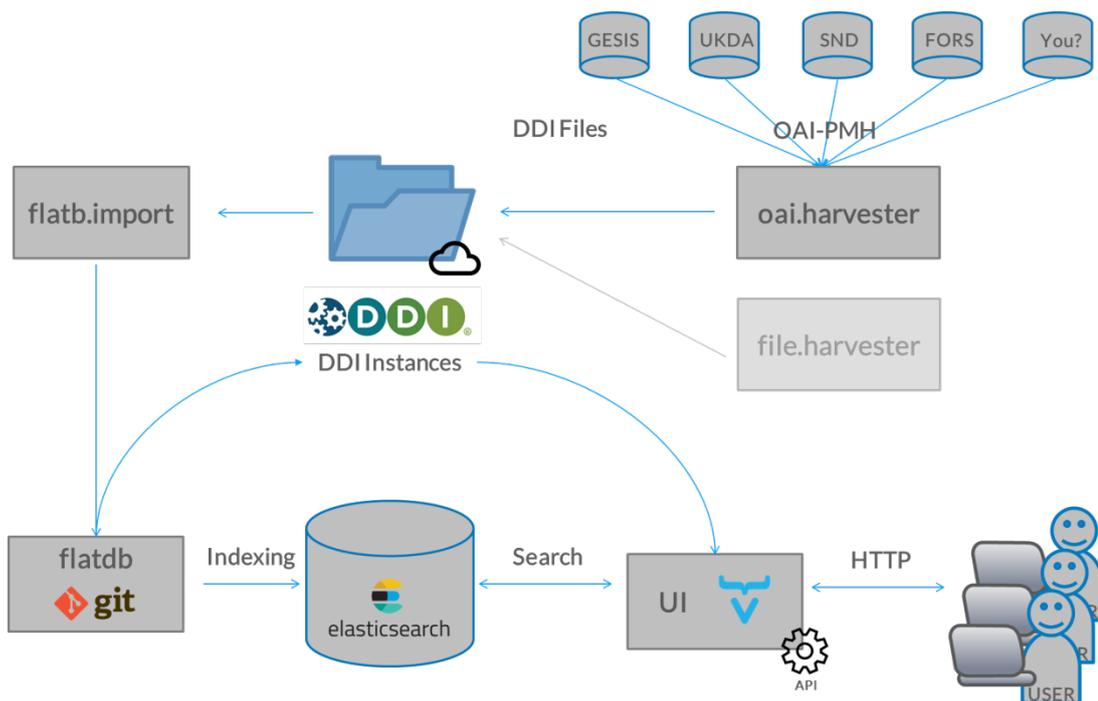


Figure 1: Modular Architecture of the EQB application<sup>2</sup>

<sup>2</sup> Azadeh Hashemi, Katarina Boland, Alexander Mühlbauer, Esra Akdeniz, Wolfgang Zenk-Möltgen. 2018. "How the CESSDA Euro Question Bank integrates with different technologies." EDDI 2018 - 10th Annual European DDI User Conference, 04.12.2018. Updated by Thomas Krämer.

The EQB front-end (UI) has been implemented with Vaadin (<https://vaadin.com/>), a user interface technology using server-side Java code. Beforehand, mock-ups have been created to demonstrate the functionality and design of the application based on the requirement analysis. Developers also made sure to be in-line with the "User Experience Guidelines for CESSDA ERIC USER INTERFACES"<sup>3</sup>.

The EQB back-end has been developed by different technologies such as Docker, Elasticsearch, and a relational database. Docker helps to create containers with virtual operating systems to run EQB in a virtualization environment. The integration of Docker into EQB simplifies the deployment of the application by handling all components of the architecture as separate applications. It also ensures that the containers have all the necessary packages and are easy to be transported and installed on different virtual servers. Elasticsearch is used to establish the search index. The storage of the various metadata elements is done by DDIFlatDB using a relational database system.

A harvester component has been implemented, to query end-points via `oai.harvester` or `file.harvester` for each SP. To index and ingest DDI files according to the EQB metadata model, CESSDA SPs contribute their standardized metadata in DDI format. They provide DDI files and the XPathes (in DDI-Lifecycle or DDI-Codebook) for every property of the EQB metadata schema. So the EQB back-end is based on DDI as a metadata standard to ensure accurate metadata exchange and to enhance metadata interoperability between EQB and SPs. Currently, metadata files are provided in the formats DDI 3.2, DDI 3.1, DDI 2.5, and DDI 1.2.

The developers use the database DDI-FlatDB<sup>4</sup> to handle the different versions of DDI, different grouping, and unequal interpretation of the elements (so-called dialectDefinitions). Example XML files are being imported into DDI-FlatDB specific to the originating SP and pushed into the flatDB relational database. The flatDB uses different dialectDefinitions for each metadata element and DDI format, and it provides a standardized way to query questions and other objects by the indexing component. This component is the prerequisite for indexing the correct metadata elements in the search index.

With Elasticsearch, the definition of the search index to be used by the search engine in EQB is being configured. The search index allows retrieving the metadata elements by the components of the EQB front-end in a rapid way. Users have access to the User Interface of the CESSDA EuroQuestionBank via HTTP.

EQB provides basic integration with other CESSDA projects such as the CESSDA Data Catalogue (CDC, former PaSC, Product and Service Catalogue) and the CESSDA Metadata Management (CMM) projects. All components of the EQB front-end and EQB back-end need to be available in a stable and maintainable version to achieve a user-friendly search application and to attain the competitiveness and sustainability of an international question bank with advanced functionalities.

---

<sup>3</sup> <https://drive.google.com/open?id=1agS-NEIVSU4q8oEeijkgon5K70S4QNLG>

<sup>4</sup> Hopt, Oliver, Claus-Peter Klas, Wolfgang Zenk-Möltgen, and Alexander Mühlbauer. 2017. "Efficient and Flexible DDI handling for the development of multiple Applications." IASSIST 2017 - Data in the Middle: The common language of research, 24.05.2017.

## 2 EQB metadata schema

---

The EQB metadata schema is an essential condition to access and display metadata accurately from all SPs. It shows in a standardized way, the overall structure of the EQB model. The schema includes all relevant metadata elements with a specific category of information. The schema uses the CMM metadata model, covers requirements for EQB, and has compatibility with the DDI-Lifecycle<sup>5</sup> metadata standard.

### 2.1 Mandatory properties

Some of the metadata properties are mandatory elements. SPs must include these required properties to meet the metadata requirements of EQB. These are:

'questionItem' or 'questionGrid' with 'questionID' and 'questionText'
'studyTitle'
'SPURL' linking to the study on the Service Provider's website
'publicationYear' of the study
'referenceToMetadataProvider' who is the CESSDA Service Provider that provides the metadata

For more information on the mandatory elements, please see table 1.

### 2.2 Recommended properties

EQB contains recommended properties that are technically optional properties, but strongly recommended to enhance that metadata can be found in EQB. These are:

'referenceOfQuestionItemToResponseDomain'
'gridDimension'
'referenceOfQuestionGridToResponseDomain'
'responseDomainTypes'
'referenceOfAnswerCodeToAnswerCategory'
'variableType'
'studyPID'

---

<sup>5</sup> <https://ddialliance.org/Specification/DDI-Lifecycle/3.2/>

'CDCURL'
'referenceStudyToInstrument'
'timeMethod'
'country'
'samplingProcedure'
'modeOfCollection'
'dataCollectionDate'
'dataCollectionStartdate'
'topicClassification'

## 2.3 Controlled vocabularies

In addition to implementing a metadata standard or schema, EQB uses controlled vocabularies. The controlled vocabularies provide a consistent way to describe the data. The following controlled vocabularies will aid the EQB users in searching and finding the correct data: 'timeMethod', 'samplingProcedure', 'modeOfCollection', 'topicClassification' (see appendix for more details).

## 2.4 ISO Codes

For some metadata fields ISO Codes are in use: 'instrumentLanguage', 'country'

### 2.4.1 Multilingual support

EQB supports multilingual metadata. The metadata can be provided in any language compliant with ISO 639-1<sup>6</sup>.

### 2.4.2 Dates and times

EQB uses ISO 8601 to describe 'dataCollectionDate' and 'publicationYear'. The code is an internationally accepted way to represent dates and times using numbers:

- Date: the complete calendar date with the structure YYYY-MM-DD e.g. 2018-03-24
- Month and year: the date with the structure YYYY-MM e.g. 2018-03
- Year: the date with the structure YYYY e.g. 2018

---

<sup>6</sup> <https://www.iso.org/iso-639-language-codes.html>

## 2.5 The EQB metadata properties in detail

Table 1 contains a detailed description of the EQB metadata schema. In the first column are the sequence numbers (*sequence no.*). The numbers describe the hierarchical order of the elements and show which child element (e.g. 1.1 question item and 1.2 question grid) belongs to its upper parent element (e.g. 1. questions). Some upper/container elements are written in bold letters. They were included to structure the schema and do not contain any actual metadata content. The *property* names (written in small letters without blanks) are together with the sequence numbers the technically identifying metadata names in the model (a unique key). The occurrence (*occ*) indicates cardinality/quantity constraints for the properties as follows:

0-n = optional and repeatable
0-1 = optional, but not repeatable
1-n = required and repeatable
1 = required, but not repeatable

Service providers can also include the recommended properties (*R*) to add more information and to provide a better user experience.

Table 1: EQB metadata schema

No.	Property	Occ	Lang	Type	Definition	Remark
Information on question(s)						
<b>1</b>	<b>question</b>	1-n	n.a.	n.a.	A question in EQB can be a single question item or a question grid.	
<b>1.1<sup>7</sup></b>	<b>questionItem</b>	0-n	n.a.	n.a.	A question item is a single question that contains a version, an ID, a name, a text, a statement, an interviewer instruction and a number.	
1.1.1	questionItemVersion	0-1	no	xs:integer	A version number of the question item.	do not publish all versions, only the latest
1.1.2	questionItemID	1-n	no	xs:integer	An identifier for a question item the metadata refers to.	
1.1.3	questionItemName	0-n	yes	xs:string	A label of a question item the metadata refers to.	
1.1.4	questionItemText	1-n	yes	xs:string	A literal question item text.	
1.1.5	questionItemStatement	0-n	yes	xs:string	A prequestion and/or postquestion and other statements as part of the question item text.	

<sup>7</sup> \* A selection must be made: either a QuestionItem or an QuestionGrid must be specified. So the \* means "Select either one of the following".

No.	Property	Occ	Lang	Type	Definition	Remark
1.1.6	questionItemInterviewerInstruction	0-n	yes	xs:string	An instruction for the interviewer.	
1.1.7	questionItemNumber	0-n	yes	xs:string	A number of the question item the metadata refers to within the associated instrument.	e.g. "QuestionNo2"/ "QNo2"/"FrageNo2"/"F2"
1.1.8	referenceOfQuestionItemToResponseDomain	0-1	no	n.a.	A reference of a specific question item to an existing response domain type (e.g. CodeDomain, NumericDomain or TextDomain) using the DDI reference structure (Agency, ID, Version).	R, See item 2.0 for details
1.1.9	referenceOfQuestionItemToConcept	0-1	no	n.a.	A reference of a specific question item to an existing concept or keyword that is related to the question using the DDI reference structure (Agency, ID, Version).	
<b>1.1.10</b>	<b>showcard</b>	0-1	n.a.	n.a.	A card, a piece of paper, or an electronic screen containing answer categories to a question, from which the respondent chooses the answer to the survey question.	
1.1.10.1	showcardFileName	1	yes	xs:string	A file name of the showcard.	
1.1.10.2	showcardURL	0-1	no	xs:string	A link to the showcard.	
<b>1.2*</b>	<b>questiongrid</b>	0-n	n.a.	n.a.	A question grid is a term for a matrix or ranking scale question that contains an ID, a version, a number, a name, a text, a statement and an interviewer instruction.	
1.2.1	questionGridVersion	0-1	no	xs:integer	A version number of the question grid.	
1.2.2	questionGridID	1-n	no	xs:integer	An identifier for a question grid the metadata refers to.	
1.2.3	questionGridName	0-n	yes	xs:string	A label of a question grid the metadata refers to.	
1.2.4	questionGridText	1-n	yes	xs:string	A literal question grid text.	
1.2.5	questionGridStatement	0-n	yes	xs:string	A prequestion and/or postquestion and other statements as part of the question grid text.	
1.2.6	questionGridInterviewerInstruction	0-n	yes	xs:string	An instruction for the interviewer.	
1.2.7	questionGridNumber	0-n	yes	xs:string	A number of the question grid the metadata refers to within the associated instrument.	
1.2.8	referenceOfQuestionGridToConcept	0-n	no	n.a.	A reference of a specific question grid to a concept or keyword that is related to the question.	
<b>1.2.9</b>	<b>gridDimension</b>	0-n	no	n.a.	A grid holds a subquestion number and a subquestion text.	R

No.	Property	Occ	Lang	Type	Definition	Remark
1.2.9.1	subQuestionNumber	0-n	yes	xs:integer	A value for the subquestion of a question grid.	
1.2.9.2	subQuestionText	1-n	yes	xs:string	A statement/or question which is part of the grid and comes after preceding a question grid text.	
<b>Information on responseDomain(s)</b>						
<b>1.2.9.3</b>	<b>responseDomain</b>	1-n	n.a.	n.a.	A domain used for capturing the response to a question.	R, One of the following should be selected/one for each GridDimension
1.2.9.3.1	referenceOfQuestion-GridToResponseDomain	0-1	no	n.a.	A response domain used by question grid item included by reference. Can be replaced by several domains e.g. CodeDomain, NumericDomain or TextDomain.	Most common response domains: CodeDomain, NumericDomain, TextDomain and ScaleDomain. See item 2.0 for details
1.2.9.3.2	outParameter	0-1	no	xs:integer	An out parameter that can be used by variable to connect it with.	
1.2.9.3.3	gridAttachment	1	no	xs:integer	A grid attachment defines the grid dimension and the subquestion for the response value.	
<b>1.2.10</b>	<b>showcard</b>	0-1	n.a.	n.a.	A card, a piece of paper, or an electronic screen containing answer categories to a question, from which the respondent chooses the answer to the survey question.	
1.2.10.1	showcardFileName	1	yes	xs:string	A file name of the showcard.	
1.2.10.2	showcardURL	0-1	no	xs:string	A link to the showcard.	
<b>2</b>	<b>responseDomainTypes</b>	0-n	n.a.	n.a.	A domain used for capturing the response to a question.	R
<b>2.1</b>	<b>codeDomain</b>	0-1	no	n.a.	A response domain is capturing a coded response (where both codes and their related category value are displayed) for a question item or variable.	
<b>2.1.1</b>	<b>Codelist</b>	1	no	n.a.	A structure used to associate a list of code values to specified categories.	
2.1.1.1	codelistID	1	no	xs:string	An identifier of the codelist the metadata refer to.	can be either kind of ID (e.g. UserID), does not need to be a DDI ID

No.	Property	Occ	Lang	Type	Definition	Remark
2.1.1.2	codelistName	0-1	yes	xs:string	A name of the codelist.	
2.1.1.3	codelistLabel	0-1	yes	xs:string	A label of the codelist.	
2.1.1.4	codelistDescription	0-1	yes	xs:string	A description of the codelist.	
2.1.2	codelistResponseCardinality	0-1	no	xs:integer	A number that indicates the minimum and the maximum number of possible answers within the codelist.	multiple choice answer/single choice answer
2.1.3	answerCode	1-n	no	xs:string	A structure that links a unique value of code to a specified category.	
2.1.4	answerCategory	1-n	yes	xs:string	A description of a particular category or response.	
2.1.4.1	answerMissingValue	0-1	no	n.a.	A value for missing data.	If the category is describing a classification of "missing data" set the value of isMissing to "true".
2.1.4.2	referenceOfAnswerCodeToAnswerCategory	0-n	no	n.a.	A reference of an answer code to an answer category using the DDI reference structure (Agency, ID, and Version).	R
<b>2.2</b>	<b>numericDomain</b>	0-1	n.a.	n.a.	A response domain capturing a numeric response for a question item.	included in line or by reference
2.2.1	numericDomainName	0-1	yes	xs:string	A name of a numeric domain.	Only if included by reference.
2.2.2	numericDomainLabel	0-1	yes	xs:string	A label of a numeric domain.	
2.2.3	numericDomainDescription	0-1	yes	xs:string	A description of a numeric domain.	
2.2.4	numericDomainRange	0-1	no	xs:decimal	A decimal that structures a numeric range. Low and high values are designated.	e.g. Please select the year of birth? (1900-2018)
<b>2.3</b>	<b>textDomain</b>	0-1	n.a.	n.a.	A response domain capturing a textual response.	
2.3.1	textDomainName	0-1	yes	xs:string	A name of a text domain.	use only if included by reference.
2.3.2	textDomainLabel	0-1	yes	xs:string	A label of a text domain.	
2.3.3	textDomainDescription	0-1	yes	xs:string	A description of a text domain.	
2.3.4	textDomainLength	0-1	no	xs:integer	A number indicating the length of the text.	e.g. What is your name? (1-50)
<b>Information on concept(s)</b>						
<b>3</b>	<b>concept</b>	0-n	n.a.	n.a.	A concept or keyword is describing the content of a question.	
3.1	conceptID	1	no	xs:string	An identifier of the concept.	
3.2	conceptName	0-1	yes	xs:string	A name of the concept.	

No.	Property	Occ	Lang	Type	Definition	Remark
3.3	conceptLabel	0-1	yes	xs:string	A label of the concept.	
3.4	conceptDescription	0-1	yes	xs:string	A description of the concept.	
Information on dataset(s)						
<b>4</b>	<b>dataset</b>	0-n	n.a.	n.a.	A collection of data, where every column of the statistical data matrix represents a particular variable and each row corresponds to a given member of the data set in question.	
4.1	datasetPID	0-n	no	xs:string	A number to capture information - type and PID Value - on latest PID(s) of the dataset. Content = PID Value	depends on the institute; some assign study ID, some dataset ID -> if dataset changes, metadata have to be updated (Only makes sense to have the latest version for EQB)
4.2	datasetTitle	1-n	yes	xs:string	A file name of the data set.	
4.3	referenceOfDatasetTo- Study	1	no	n.a.	A reference from dataset to a specific study using the DDI reference structure (Agency, ID, and Version).	
<b>4.4</b>	<b>variable</b>	0-n	n.a.	n.a.	A variable represents a measurable quantity (quantitative variable) or takes on values that are names of labels (qualitative variable).	
4.4.1	variableName	1-n	yes	xs:string	A name of a variable within a dataset.	
4.4.2	variableLabel	0-n	yes	xs:string	A label of a variable for display.	
4.4.3	variableDescription	0-n	yes	xs:string	An additional description of the variable or the file content.	
4.4.4	variableType	0-1	no	xs:string	A variable can have different types according to the ways they can be studied, measured, and presented.	e.g. categorical, continuous, nominal, ordinal; R
4.4.5	variableRepresentation	0-1	no	n.a.	A reference of a variable to a codelist using the DDI reference structure (Agency, ID, Version).	
4.4.6	referenceOfVariable- ToQuestion	0-1	no	n.a.	A reference of a variable to a question using the DDI reference structure (Agency, ID, and Version).	
Information on instrument(s)						
<b>5</b>	<b>instrument</b>	0-n	n.a.	n.a.	A tool to obtain data from respondents. For most social and	

No.	Property	Occ	Lang	Type	Definition	Remark
					behavioral surveys, the instrument involves a questionnaire with questions and response options.	
5.1	instrumentName	1	yes	xs:string	A name of the instrument.	
5.2	instrumentLanguage	1	yes	xs:string	A language for the instrument that is being used in the study using the ISO language code. It is a requirement to know in which language a study conducted the interviews. When a questionnaire is being prepared, a language has to be chosen to know if a study uses an instrument in German or English.	ISO CODE See: "EQB MODEL IN TIME AND SPACE ETC." <a href="https://drive.google.com/file/d/1x_XdB497LGN_G_4vIQWD3BM_BwTj4u3_ul/view?usp=sharing">https://drive.google.com/file/d/1x_XdB497LGN_G_4vIQWD3BM_BwTj4u3_ul/view?usp=sharing</a>
5.3	referenceOfInstrumentToQuestion	1-n	n.a.	n.a.	A reference of the instrument to a question using the DDI reference structure (ID, Version, and Agency).	
<b>Information on study</b>						
<b>6</b>	<b>studyInformation</b>	1	n.a.	n.a.	Detailed information on a study that is related to a question.	
6.1	studyPID	0-n	yes	xs:string	A value of a persistent identifier for the study.	R
6.2	studyNumber	1-n	no	xs:string	A unique study number of the archive.	If a SP do not have a studyNumber a number can be generated automatically.
6.3	studyVersionNumber	0-1	no	xs:string	A number of the version (free text) that the Service Provider uses to identify this version of the study.	
6.4	studyVersiondate	0-1	no	xs:date	A date of the version that the Service Provider uses to identify this version of the study.	
6.5	studyTitle	1-n	yes	xs:string	A title or name for the study.	Mandatory to send title in one language version, but optional to send other language versions.
6.6	studyDescription	0-1	yes	xs:string	An abstract that describes the content(s) of the study.	
6.7	SPURL	1-n	yes	xs:string	A URL or URN (a reference to a web resource that specifies its location) linking to the study on the SP website.	Add @xml:lang attribute to the URL, because the URL of

No.	Property	Occ	Lang	Type	Definition	Remark
						different language version may be different.
6.8	CDCURL	0-n	yes	xs:string	A URL or URN (a reference to a web resource that specifies its location) linking to the study to a CDC page.	R
6.9	publicationYear	1	no	xs:date	A year of publication (by a publisher) of the study.	ISO 8601 Format: YYYY
6.10	publisherName	0-n	yes	xs:string	A name of the institution publishing the metadata, i.e. CESDDA Service Provider providing the metadata information or the name of the actual publisher of the metadata, who is not the SPs. Only on study level.	See difference to referenceOf-MetadataProvider
6.11	referenceStudyToPrincipalInvestigatorPerson	0-n	no	n.a.	A reference of a study to a person who is a principal investigator using the DDI reference structure (ID, Version, Agency). Principal investigators are the main researchers involved in producing the data.	
6.12	referenceStudyToPrincipalInvestigatorInstitution	0-n	no	n.a.	A reference of a study to an institution that is a principal investigator using the DDI reference structure (ID, Version, and Agency). Principal investigators are the main institutions involved in producing the data.	Please reference the different principal investigators - if there is more than one - in priority order within the XML. The order in which they are referenced will decide upon the order they appear in.
6.13	referenceStudyToInstrument	0-n	no	n.a.	A reference of a study to an instrument using the DDI reference structure (ID, Version, and Agency).	R
6.14	timeMethod	0-n	no	xs:string	A type of time dimension of the data collection.	CV of DDI: Time Method: <a href="http://www.ddi-alliance.org/Specification/DDI-CV/TimeMethod_1.2.html">http://www.ddi-alliance.org/Specification/DDI-CV/TimeMethod_1.2.html</a>
6.14.1	timeMethodName	1	no	xs:string	A name of the vocabulary used for time method.	R The name of the CV is

No.	Property	Occ	Lang	Type	Definition	Remark
						"TimeMethod".
6.14.2	timeMethodAgency	0-1	no	xs:string	A name of the agency maintaining the vocabulary for time method.	The name of the agency is "DDI Alliance".
6.14.3	timeMethodURI	1	no	n.a.	An identifier of the vocabulary for time method.	
6.15	country	0-n	no/yes	xs:string	A country in which the study took place.	EQB Filter Format: ISO Code The 2 letter codes of ISO 3166 are not multilingual (standards), hence 'no' for multilingual and 'yes' for the content of the element itself. attribute @abbr = "no" >Norway<, so the content of the element can be multilingual, hence yes for multilingual and no for ISO Code. R
6.16	samplingProcedure	0-n	no	xs:string	A type of sampling procedure used for data collection. Controlled vocabulary includes a typology of sampling methods.	DDI Sampling Procedure <a href="http://www.ddi-alliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html">http://www.ddi-alliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html</a>
6.16.1	samplingProcedure-Name	1	no	xs:string	A name of the used vocabulary for sampling procedure.	The name of the CV is "SamplingProcedure".
6.16.2	samplingProcedureAgency	0-1	no	xs:string	A name of the agency maintaining the vocabulary for sampling procedure.	The name of the agency is "DDI Alliance".
6.16.3	samplingProcedureURI	1	no	n.a.	An identifier of the vocabulary for sampling procedure.	
6.17	modeOfCollection	0-n	no	xs:string	A type of data collection - the procedure, technique, or mode of inquiry used to attain the data.	EQB Filter CV of DDI: Mode Of Collection: <a href="http://www.ddi-alliance.org/Specification">http://www.ddi-alliance.org/Specification</a>

No.	Property	Occ	Lang	Type	Definition	Remark
						<a href="#">ication/DDI-CV/ModeOfCollection_3.0.html</a>
6.17.1	modeOfCollectionName	1	no	xs:string	A name of the used vocabulary for mode of data collection.	The name of the CV is "ModeOfCollection".
6.17.2	modeOfCollectionAgency	0-1	no	xs:string	A name of the agency maintaining the vocabulary for mode of data collection.	The name of the agency is "DDI Alliance".
6.17.3	modeOfCollectionURI	1	no	n.a.	An identifier of the vocabulary for mode of data collection.	
<b>6.18</b>	<b>dataCollectionDate</b>	0-n	no	xs:date	A year the data was collected.	EQB Filter R
<b>6.18.1</b>	<b>dataCollectionStartDate</b>	0-n	no	xs:date	A start date of the data collection (if there are multiple start dates, use the first date).	ISO 8601 R
6.18.1.1	date	0-1	no	xs:date	A complete calendar date with the structure YYYY-MM-DD e.g. 2018-03-24.	ISO 8601
6.18.1.2	monthyear	0-1	no	xs:date	A date with the structure YYYY-MM, e.g. 2018-03.	ISO 8601
6.18.1.3	year	1	no	xs:date	A date with the structure YYYY, e.g. 2018.	ISO 8601
<b>6.18.2</b>	<b>dataCollectionEndDate</b>	0-n	no	xs:date	An end date of data collection (if there are multiple end dates, use the last).	ISO 8601
6.18.2.1	date	0-1	no	xs:date	A complete calendar date with the structure YYYY-MM-DD, e.g. 2019-03-24.	ISO 8601
6.18.2.2	monthyear	0-1	no	xs:date	A date with the structure YYYY-MM, e.g. 2019-03	ISO 8601
6.18.2.3	year	1	no	xs:date	A date with the structure YYYY, e.g. 2019.	ISO 8601
<b>6.19</b>	<b>topicClassification</b>	0-n	yes	xs:string	An association of a study with a classification to identify the general topics of a study. As a classification, the use of the CESSDA Topic Classification is mandatory.	R
6.19.1	topicClassificationName	1	no	xs:integer	A name of the used classification system [remark: since the use of the CESSDA Topic Classification is mandatory, the 'Classification-SystemName' is by default 'CESSDA Topic Classification'].	The name of the CV is "CESSDA Topic Classification".
6.19.2	topicClassificationAgency	0-1	no	xs:string	A name of the agency maintaining the classification. [remark:	The name of the agency is

No.	Property	Occ	Lang	Type	Definition	Remark
					since the use of the CESSDA Topic Classification is mandatory, the default is 'CESSDA'].	"CESSDA".
6.19.3	topicClassificationURI	1	no	xs:string	A persistent identifier of the classification system.	
6.20	referenceStudyToDocument	0-n	no	n.a.	A reference of a study to a document using the DDI reference structure (ID, Version, and Agency).	
<b>Information on study group(s)</b>						
<b>7</b>	<b>studyGroupInformation</b>	0-n	n.a.	n.a.	Information on the study group the study appears in.	
7.1	studyGroupID	0-n	no	xs:integer	A number to capture information - type and PID Value - on PID(s) of the study group. Content = PID Value	
7.2	studyGroupTitle	1-n	yes	xs:string	A title or name for the study group.	
7.3	url	1-n	no	xs:string	An URL or URN (a reference to a web resource that specifies its location) linking to the study group.	
7.4	referenceOfStudyGroupToStudy	1-n	no	n.a.	A reference of a study group to a study.	
7.5	referenceToDocument	0-n	no	n.a.	A reference of a study group to a document, e.g. questionnaire.	
<b>Information on person(s)</b>						
<b>8</b>	<b>persons</b>	0-n	n.a.	n.a.	Any kind of person that is related to the study.	
8.1	<b>person</b>	1	no	xs:string	A first, middle and last name of a person.	
8.1.1	firstName	1	no	xs:string	A name of a person.	
8.1.2	middleName	0-1	no	xs:string	A middle name of a person.	
8.1.3	lastName	1	no	xs:string	A last name of a person.	
8.2	ReferenceToAffiliationName	0-n	yes	xs:string	A reference to the institution a person is affiliated to.	
<b>Information on institution(s)</b>						
<b>9</b>	<b>institution</b>	0-n	n.a.	n.a.	Any kind of institution that is related to the study.	
9.1	institutionName	1	yes	xs:string	A name of the organization or institution	Since you cannot document the order of an institution in DDI, put the institution names in the same order you would like to have them displayed (for example put

No.	Property	Occ	Lang	Type	Definition	Remark
						the main institution first).
9.2	institutionAbbreviation	0-1	no	xs:string	An abbreviation of the institution's name.	
Information on meta-metadata						
10	referenceToMetadataProvider	1	no	n.a.	The CESSDA Service Provider that provides the metadata. Difference between publisherName and referenceToMetadataProvider: publisherName is the name of the institution who published the dataset, e.g. Infratest, but the metadata will be distributed by GESIS, who is the metadata provider. So the referenceToMetadataProvider is the name of the owner (technical information).	R
Information on document(s)						
<b>11</b>	<b>Document</b>	0-n	n.a.	n.a.	All documents that a SP might want to add in addition (e.g. Codebook).	
11.1	documentID	0-1	no	xs:string	An identifier for a document.	
11.2	documentTitle	1	yes	xs:string	A title of a document.	
11.3	documentFormat	1	no	xs:string	A textual description of the technical format of a document (e.g. XML).	
11.4	documentURL	0-1	no	xs:string	An URL to a document.	

## 3 Appendix

### 3.1 TimeMethod<sup>8</sup>

#### 3.1.1 Usage

Description: Describes the time dimension of the data collection.

**DDI 3.2:** A brief textual description or classification of the type of the time methodology used. Supports the use of an external controlled vocabulary.

Module Name	Element Name
datacollection	TypeOfTimeMethod

**DDI 2.5:** The time method or time dimension of the data collection. The "method" attribute is included to permit the development of a controlled vocabulary for this element. For forward-compatibility, DDI 3 XHTML tags may be used in this element.

Element Number in DDI 2.1	Element/Attribute Name
2.3.1.1	timeMeth@method

#### 3.1.2 Table: Codelist for TimeMethod

Value of Code	Descriptive Term of the Code	Definition of the Code
Longitudinal	Longitudinal	Data collected repeatedly over time to allow studying change in a population. At least some of the questions or modules are repeated over waves. Use the broad term when none of the sub-terms is suitable.
Longitudinal.CohortEventBased	Longitudinal: Cohort/Event-based	Data collected over time from the same cohort of respondents. The individuals in the cohort are connected in some way or have shared some significant experience within a given period. In some cases, the samples may differ between waves but are drawn from the same cohort. Examples: birth year, disease (clinical trials), common problem (intervention studies), education, employment, family formation, participation in an event.
Longitudinal.TrendRepeatedCrossSection	Longitudinal: Trend/Repeated cross-section	Data collected from different samples or different groups of people from the same population at several points in time, using at least partly the same set of questions/variables. Conclusions are drawn for the population. Examples: European Social Survey (ESS), national longitudinal crime surveys.

<sup>8</sup> [https://ddialliance.org/Specification/DDI-CV/TimeMethod\\_1.2.html](https://ddialliance.org/Specification/DDI-CV/TimeMethod_1.2.html)

<b>Value of Code</b>	<b>Descriptive Term of the Code</b>	<b>Definition of the Code</b>
Longitudinal.Panel	Longitudinal: Panel	Data collected over time from, or about, the same sample of respondents. Differs from cohort/event-based data in that the selection of respondents is not based on their being connected in some way or having shared some significant experience.
Longitudinal.Panel.Continuous	Longitudinal: Panel: Continuous	Data collected from a panel of respondents on a regular basis.
Longitudinal.Panel.Interval	Longitudinal: Panel: Interval	Data collected from a panel of respondents only when information is needed.
TimeSeries	Time series	Data collected repeatedly over time to study change in observations. These are typically "objective" measurements of phenomena that can be observed externally, as opposed to attitudes/opinions or feelings. Examples may include economic/financial indicators, natural/meteorological phenomena, vital statistics, etc.
TimeSeries.Continuous	Time series: Continuous	Measurements are taken at every instant in time. Examples: lie detectors, electrocardiograms, etc.
TimeSeries.Discrete	Time series: Discrete	Measurements are taken at (usually regularly) spaced intervals. Examples: macroeconomics (weekly share prices, monthly profits, sales); meteorology (hourly temperature); measurements of individuals (blood pressure, weight, height); sociology (crime figures, employment figures), etc.
CrossSection	Cross-section	Data collected by observing subjects within the study period, without regard to changes over time. May include more than one collection event. Analysis of cross-sectional data often consists in comparing the differences and similarities among subjects.
CrossSectionAd-HocFollowUp	Cross-section ad-hoc follow-up	Data collected at one point in time to complete information collected in a previous cross-sectional study; the decision to collect follow-up data was not included in the original study design.
Other	Other	Use if the time method is known, but not found in the list.

## 3.2 SamplingProcedure<sup>9</sup>

### 3.2.1 Usage

Description: A typology of sampling methods.

**DDI 3.2:** A brief textual description or classification of the type of sampling procedure used. Supports the use of an external controlled vocabulary.

Module Name	Element Name
datacollection	TypeOfSamplingProcedure

**DDI 2.5:** The 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.

Element Number in DDI 2.1	Element/Attribute Name
2.3.1.4	sampProc

### 3.2.2 Table: Codelist for SamplingProcedure

Value of Code	Descriptive Term of the Code	Definition of the Code
TotalUniverse-CompleteEnumeration	Total universe/Complete enumeration	All units (individuals, households, organizations, etc.) of a target population are included in the data collection. For example, if the target population is defined as the members of a trade union, all union members are invited to participate in the study. Also called "census" if the entire population of a regional unit (e.g. a country) is selected.
Probability	Probability	All units (individuals, households, organizations, etc.) of a target population have a non-zero probability of being included in the sample and this probability can be accurately determined. Use this broader term if a more specific type of probability sampling is not known or is difficult to identify.
Probability.SimpleRandom	Probability: Simple random	All units of a target population have an equal probability of being included in the sample. Typically, the entire population is listed in a "sample frame", and units are then chosen from this frame using a random selection method.
Probability.SystematicRandom	Probability: Systematic random	A fixed selection interval is determined by dividing the population size by the desired sample size. A starting point is then randomly drawn from the sample frame, which normally covers the entire target population.  From this starting point, units for the sample are chosen based on the selection interval. Also known as interval sampling. For example, a company survey seeks a sample of 1,000 employees out of 10,000 total. Beginning with a random starting number, every 10th name

<sup>9</sup> [https://ddialliance.org/Specification/DDI-CV/SamplingProcedure\\_1.1.html](https://ddialliance.org/Specification/DDI-CV/SamplingProcedure_1.1.html)

Value of Code	Descriptive Term of the Code	Definition of the Code
		from the employee list of the company will be invited to participate in the study.
Probability.Stratified	Probability: Stratified	The target population is subdivided into separate and mutually exclusive segments (strata) that cover the entire population. Independent random samples are then drawn from each segment. For example, in a national public opinion survey the entire population is divided into two regional strata: East and West. After this, sampling units are drawn from within each region using simple or systematic random sampling. Use this broader term if the specific type of stratified sampling is not known or difficult to identify.
Probability.Stratified.Proportional	Probability: Stratified: Proportional	The target population is subdivided into separate and mutually exclusive segments (strata) that cover the entire population. In proportional stratified sampling the number of elements chosen from each stratum is proportional to the population size of the stratum when viewed against the entire population. For example, a country is divided into two regional strata that comprise 80 percent (West) and 20 percent (East) of the total population. For a sample of 1,000 people, 800 (i.e., 80 percent) would be drawn from the West and 200 (i.e., 20 percent) from the East to accurately represent their proportion in the total population.
Probability.Stratified.Disproportional	Probability: Stratified: Disproportional	The target population is subdivided into separate and mutually exclusive segments (strata) that cover the entire population. In disproportional sampling the number of units chosen from each stratum is not proportional to the population size of the stratum when viewed against the entire population. The number of sampled units from each stratum can be equal, optimal, or can reflect the purpose of the study, like oversampling of different subgroups of the population. For example, a country is divided into two regional strata that comprise 80 percent (West) and 20 percent (East) of the country's population. If equal representation of the two regions is needed in a study, half the sample may be drawn from the West and half from the East, so that each region is represented by 50 percent of the sample. If a more detailed analysis of the population from the East is needed, 40 percent of the units may be drawn from the West and 60 percent from the East, so that the East is over-represented.
Probability.Cluster	Probability: Cluster	The target population is divided into naturally occurring segments (clusters) and a probability sample of the clusters is selected. Data are then collected from all units within each selected cluster. Sampling is often clustered by geography, or time period. Use this broader term if a more specific type of cluster sampling is not known or is difficult to identify.
Probability.Cluster.SimpleRandom	Probability: Cluster: Simple random	The target population is divided into naturally occurring segments (clusters) and a simple random sample of the clusters is selected. Data are then collected from all units within each selected cluster. For example, for a sample of students in a city, a number of schools would be chosen using the random selection method, and then all of the students from every sampled school would be included.
Probability.Cluster.Stratified	Probability: Cluster:	The target population is divided into naturally occurring segments (clusters); next, these are divided into mutually exclusive strata and a

Value of Code	Descriptive Term of the Code	Definition of the Code
dRandom	Stratified random	random sample of clusters is selected from each stratum. Data are then collected from all units within each selected cluster. For example, for a sample of students in a city, schools would be divided into two strata by school type (private vs. public); schools would be then randomly selected from each stratum, and all of the students from every sampled school would be included.
Probability.Multistage	Probability: Multistage	Sampling is carried out in stages using smaller and smaller units at each stage, and all stages involve a probability selection. The type of probability sampling procedure may be different at each stage. For example, for a sample of students in a city, schools are randomly selected in the first stage. A random sample of classes within each selected school is drawn in the second stage. Students are then randomly selected from each of these classes in the third stage.
Nonprobability	Non-probability	The selection of units (individuals, households, organizations, etc.) from the target population is not based on random selection. It is not possible to determine the probability of each element to be sampled. Use this broader term if the specific type of non-probability is not known, difficult to identify, or if multiple non-probability methods are being employed.
Nonprobability.Availability	Non-probability: Availability	The sample selection is based on the units' accessibility/relative ease of access. They may be easy to approach, or may themselves choose to participate in the study (self-selection). Researchers may have particular target groups in mind but they do not control the sample selection mechanism. For example, students leaving a particular building on campus may be approached, or individuals may volunteer to participate in response to invitations that do not target them specifically, but a larger group to which they may belong. Also called "convenience" or "opportunity" sampling.
Nonprobability.Purposive	Non-probability: Purposive	Sample units are specifically identified, selected and contacted for the information they can provide on the researched topic. Selection is based on different characteristics of the independent and/or dependent variables under study, and relies on the researchers' judgement. The study authors, or persons authorized by them have control over the sample selection mechanism and the universe is defined in terms of the selection criteria. Also called "judgement" sampling. For example, a medical researcher may intentionally select individuals who are similar in most respects, except on the outcome of the research topic, which can be a specific disease. Some types of purposive sampling are typical/deviant case, homogeneous/maximum variation, expert, or critical case sampling.
Nonprobability.Quota	Non-probability: Quota	The target population is subdivided into separate and mutually exclusive segments according to some predefined quotation criteria. The distribution of the quotation criteria (gender/age/ethnicity ratio, or other characteristics, like religion, education, etc.) is intended to reflect the real structure of the target population or the structure of the desired study population. Non-probability samples are then drawn from each segment until a specific number of units has been reached. For example, if the target population consists of 45 percent females and 55 percent males, a proportional quota sample will have the same gender percentages, while in a non-proportional quota

Value of Code	Descriptive Term of the Code	Definition of the Code
		sample the percentages will be different, based on some study-related consideration (for instance, the need to oversample for certain under-represented segments of the population).
Nonprobability-RespondentAssisted	Non-probability: Respondent-assisted	Sample units are identified from a target population with the assistance of units already selected (adapted from "Public Health Research Methods", ed. Greg Guest, Emily E. Namey, 2014). A typical case is snowball sampling, in which the researcher identifies a group of units that matches a particular criterion of eligibility. The latter are asked to recruit other members of the same population that fulfil the same criterion of eligibility (sampling of specific populations like migrants, etc.).
MixedProbability-Nonprobability	Mixed probability and non-probability	Sample design that combines probability and non-probability sampling within the same sampling process. Different types of sampling may be used at different stages of creating the sample. For example, for a sample of minority students in a city, schools are randomly selected in the first stage. Then, a quota sample of students is selected within each school in the second stage. If separate samples are drawn from the same target population using different sampling methods, the type of sampling procedure used for each sample should be classified separately.
Other	Other	Use if the sampling procedure is known, but not found in the list.

### 3.3 ModeOfCollection<sup>10</sup>

#### 3.3.1 Usage

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

**DDI 3.2:** A brief textual description or classification of the mode of data collection. Supports the use of an external controlled vocabulary.

Module Name	Element Name
datacollection	TypeOfModeOfCollection

**DDI 2.5:** The method used to collect the data; instrumentation characteristics. XHTML formatting may be used in the txt element for forward-compatibility with DDI-L.

Element Number in DDI 2.1	Element/Attribute Name
2.3.1.6	collMode

<sup>10</sup> [https://ddialliance.org/Specification/DDI-CV/ModeOfCollection\\_3.0.html](https://ddialliance.org/Specification/DDI-CV/ModeOfCollection_3.0.html)

### 3.3.2 Table: Codelist for ModeOfCollection

Value of Code	Descriptive Term of the Code	Definition of the Code
Interview	Interview	A pre-planned communication between two (or more) people - the interviewer(s) and the interviewee(s) - in which information is obtained by the interviewer(s) from the interviewee(s). If group interaction is part of the method, use "Focus group".
Interview.FaceToFace	Face-to-face interview	Data collection method in which a live interviewer conducts a personal interview, presenting questions and entering the responses. Use this broader term if not CAPI or PAPI, or if not known whether CAPI/PAPI or not.
Interview.FaceToFace.CAPIorCAMI	Face-to-face interview: Computer-assisted (CAPI/CAMI)	Computer-assisted personal interviewing (CAPI), or computer-assisted mobile interviewing (CAMI). Data collection method in which the interviewer reads questions to the respondents from the screen of a computer, laptop, or a mobile device like tablet or smartphone, and enters the answers in the same device. The administration of the interview is managed by a specifically designed program/application.
Interview.FaceToFace.PAPI	Face-to-face interview: Paper-and-pencil (PAPI)	Paper-and-pencil interviewing (PAPI). The interviewer uses a traditional paper questionnaire to read the questions and enter the answers.
Interview.Telephone	Telephone interview	Interview administered on the telephone. Use this broader term if not CATI, or if not known whether CATI or not.
Interview.Telephone.CATI	Telephone interview: Computer-assisted (CATI)	Computer-assisted telephone interviewing (CATI). The interviewer asks questions as directed by a computer, responses are keyed directly into the computer and the administration of the interview is managed by a specifically designed program.
Interview.Email	E-mail interview	Interviews conducted via e-mail, usually consisting of several e-mail messages that allow the discussion to continue beyond the first set of questions and answers, or the first e-mail exchange.
Interview.WebBased	Web-based interview	An interview conducted via the Internet. For example, interviews conducted within online forums or using web-based audio-visual technology that enables the interviewer(s) and interviewee(s) to communicate in real time.
SelfAdministered-Questionnaire	Self-administered questionnaire	Data collection method in which the respondent reads or listens to the questions, and enters the responses by him/herself; no live interviewer is present, or participates in the questionnaire administration. If possible, use a narrower term. Use this broader term if the method is not described by any of the narrower terms - for example, for PDF and diskette questionnaires.
SelfAdministered-Questionnaire.Email	Self-administered questionnaire: E-mail	Self-administered survey in which questions are presented to the respondent in the text body of an e-mail or as an attachment to an e-mail, but not as a link to a web-based questionnaire. Responses are also sent back via e-mail, in the e-mail body or as an attachment.
SelfAdministered-Questionnaire.Paper	Self-administered	Self-administered survey using a traditional paper questionnaire delivered and/or collected by mail (postal services), by fax, or in

Value of Code	Descriptive Term of the Code	Definition of the Code
	questionnaire: Paper	person by either interviewer, or respondent.
SelfAdministered- Questionnaire.SMSorMMS	Self-administered questionnaire: Messaging (SMS/MMS)	Self-administered survey in which the respondents receive the questions incorporated in SMS (text messages) or MMS (messages including multimedia content) and send their replies in the same format.
SelfAdministered- Questionnaire.CAWI	Self-administered questionnaire: Web-based (CAWI)	Computer-assisted web interviewing (CAWI). Data are collected using a web questionnaire, produced with a program for creating web surveys. The program can customize the flow of the questionnaire based on the answers provided, and can allow for the questionnaire to contain pictures, audio and video clips, links to different web pages, etc.
SelfAdministered- Questionnaire.CASI	Self-administered questionnaire: Computer-assisted (CASI)	Computer-assisted self-interview (CASI). Respondents enter the responses into a computer (desktop, laptop, Palm/PDA, tablet, etc.) by themselves. The administration of the questionnaire is managed by a specifically designed program/application but there is no real-time data transfer as in CAWI, the answers are stored on the device used for the interview. The questionnaire may be fixed form or interactive. Includes VCASI (Video computer-assisted self-interviewing), ACASI (Audio computer-assisted self-interviewing) and TACASI (Telephone audio computer-assisted self-interviewing).
FocusGroup	Focus group	A group discussion on a particular topic, organized for research purposes. The individuals are selected with relevance to the topic, and interaction among the participants is used as part of the method.
Focus- Group.FaceToFace	Face-to-face focus group	The focus group participants meet in person to conduct the discussion.
Focus- Group.Telephone	Telephone focus group	The focus group discussion is conducted over the telephone.
FocusGroup.Online	Online focus group	The focus group discussion is conducted over the Internet in an interactive manner.
SelfAdminis- teredWritingsAndDi- aries	Self-administered writings and/or diaries	Narratives, stories, diaries, and written texts created by the research subject.
SelfAdminis- teredWritingsAndDi- aries.Email	Self-administered writings and/or diaries: E-mail	Narratives, stories, diaries, and written texts submitted via e-mail messages.
SelfAdminis- teredWritingsAndDi- aries.Paper	Self-administered writings and/or diaries: Paper	Narratives, stories, diaries, and written texts created and collected in paper form.
SelfAdminis- teredWritingsAndDi-	Self-administered	Narratives, stories, diaries, and written texts gathered from Internet sources, e.g. websites, blogs, discussion forums.

Value of Code	Descriptive Term of the Code	Definition of the Code
aries.WebBased	writings and/or diaries: Web-based	
Observation	Observation	Research method that involves collecting data as they occur (for example, observing behaviors, events, development of condition or disease, etc.), without attempting to manipulate any of the independent variables.
Observation.Field	Field observation	Observation that is conducted in a natural environment.
Observation.Field.Participant	Participant field observation	Type of field observation in which the researcher interacts with the subjects and often plays a role in the social situation under observation.
Observation.Field.Nonparticipant	Non-participant field observation	Observation that is conducted in a natural, non-controlled setting without any interaction between the researcher and his/her subjects.
Observation.Laboratory	Laboratory observation	Observation that is conducted in a controlled, artificially created setting. For example, observing children's play in a laboratory playroom.
Observation.Laboratory.Participant	Participant laboratory observation	Type of laboratory observation in which the researcher interacts with the subjects and often plays a role in the social situation under observation. For example, observing children's play in a laboratory playroom with the researcher taking part in the play.
Observation.Laboratory.Nonparticipant	Non-participant laboratory observation	Type of laboratory observation that is conducted without any interaction between the researcher and his/her subjects.
Observation.ComputerBased	Computer-based observation	Type of observation in which data regarding computer usage are being collected by software that can be built into the computer program itself or can be a separate program. Information may be collected about the number of users, the ways in which users interact with the program(s), how much time they spend on a page, how they use specific sections of applications, how they navigate from page to page or from one application to another, etc.
Experiment	Experiment	Research method involving the manipulation of some or all of the independent variables included in the hypotheses.
Experiment.Laboratory	Laboratory experiment	An experiment conducted in a controlled, artificially created physical setting, in which a researcher manipulates one or several independent variables and measures its/their effect on the dependent variable.
Experiment.FieldIntervention	Field/Intervention experiment	An experiment conducted in a natural, uncontrolled setting, in which the researcher manipulates one or several independent variables. Intervention/clinical studies are one example of field experiments.

Value of Code	Descriptive Term of the Code	Definition of the Code
Experiment.WebBased	Web-based experiment	An experiment conducted in the virtual setting of the World Wide Web, in which experimental materials are programmed to implement artificial situations or events to be investigated in a distributed environment. (Reips, U.-D. (2002). Theory and techniques of Web experimenting. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), <i>Online Social Sciences</i> . Seattle: Hogrefe & Huber. Available at: <a href="http://www.websm.org/uploadi/editor/Reips_2002_Theory_and_techniques.pdf">http://www.websm.org/uploadi/editor/Reips_2002_Theory_and_techniques.pdf</a> [07 June 2013])
Recording	Recording	Registering by mechanical or electronic means, in a form that allows the information to be retrieved and/or reproduced. For example, images or sounds on disc or magnetic tape.
ContentCoding	Content coding	As a mode of secondary data collection, content coding applies coding techniques to transform qualitative data (textual, video, audio or still-image) originally produced for other purposes into quantitative data (expressed in unit-by-variable matrices) in accordance with pre-defined categorization schemes. For example, coded party manifesto data like the "European Parliament Election Study 2009, Manifesto Study" (doi:10.4232/1.10204)".
Transcription	Transcription	Capturing information in writing from a different source, or from a different medium, alphabet, or form of notation, like scientific formulae, or musical notes. For transcribed interviews or observations, it is recommended to document the primary mode of collection, using one of the interview or observation terms.
CompilationSynthesis	Compilation/Synthesis	Collecting and assembling data from multiple, often heterogeneous sources that have one or more reference points in common, and at least one of the sources was originally produced for other purposes. The data are incorporated in a new entity. For example, providing data on the number of universities in the last 150 years using a variety of available sources (e.g. finance documents, official statistics, university registers), combining survey data with information about geographical areas from official statistics (e.g. population density, doctors per capita, etc.), or using RSS to collect blog posts or tweets, etc.
Summary	Summary	Presentation of information in a condensed form, by reducing it to its main points. For example, abstracts of interviews or reports that are published and used as data rather than the full-length interviews or reports.
Aggregation	Aggregation	Statistics that relate to broad classes, groups, or categories. The data are averaged, totaled, or otherwise derived from individual-level data, and it is no longer possible to distinguish the characteristics of individuals within those classes, groups, or categories. For example, the number and age group of the unemployed in specific geographic regions, or national level statistics on the occurrence of specific offences, originally derived from the statistics of individual police districts.
Simulation	Simulation	Modeling or imitative representation of real-world processes, events, or systems, often using computer programs. For example, a program modeling household consumption responses to indirect tax changes; or a dataset on hypothetical patients and their

<b>Value of Code</b>	<b>Descriptive Term of the Code</b>	<b>Definition of the Code</b>
		drug exposure, background conditions, and known adverse events.
MeasurementsAnd-Tests	Measurements and tests	Assessing specific properties (or characteristics) of beings, things, phenomena, (and/ or processes) by applying pre-established standards and/or specialized instruments or techniques.
MeasurementsAnd-Tests.Educational	Educational measurements and tests	Assessment of knowledge, skills, aptitude, or educational achievement by means of specialized measures or tests.
MeasurementsAnd-Tests.Physical	Physical measurements and tests	Assessment of physical properties of living beings, objects, materials, or natural phenomena. For example, findings from hands-on medical examination (e.g., palpation or auscultation), clinical measurements and lab tests like blood analysis, blood pressure, heart rate, body weight and height, as well as general measurements like time, distance, mass, temperature, force, power, speed, GPS data on physical movement and other physical parameters or variables, like geospatial data.
MeasurementsAnd-Tests.Psychological	Psychological measurements and tests	Assessment of personality traits or psychological/behavioral responses by means of specialized measures or tests. For example, objective tests like self-report measures with a restricted response format, or projective methods allowing free responses, including word association, sentence or story completion, vignettes, cartoon test, thematic apperception tests, role play, drawing tests, inkblot tests, choice ordering exercises, etc.
Other	Other	Use if the mode of data collection is known, but not found in the list.

## Acknowledgments

---

This work was carried out within the framework of the EQB projects funded by CESSDA.

We gratefully acknowledge the following CESSDA EQB Service Providers and colleagues for their valuable suggestions and discussions:

- Alexander Mühlbauer (GESIS)
- Katarina Boland (GESIS)
- Azadeh Hashemi (GESIS)
- Carla Anselment (GESIS)
- Claus-Peter Klas (GESIS)
- Thomas Krämer (GESIS)
- Martin Friedrichs (GESIS)
- Taina Jääskeläinen (FSD)
- Toni Sissala (FSD)
- Benjamin Beuster (NSD)
- Iris Alfredsson (SND)
- Stefan Jakobsson (SND)
- Johan Fihn (SND)
- Jeannine CT Beeken (UKDS)
- Veerle Van Den Eynden (UKDS)
- Brian Kleiner (FORS)
- Emilie Morgan De Paula (FORS)
- Anne Sofie Fink Kjeldgaard (DNA)
- Lea Sztuk Haahr (DNA)
- Peter Hegedus (TARKI)
- Tomáš Čížek (CSDA)
- Apostolos Linardis (SODANET/EKKE)
- Dimitra Kondyli (SODANET/EKKE)