



How the CESSDA Euro Question Bank integrates with different technologies

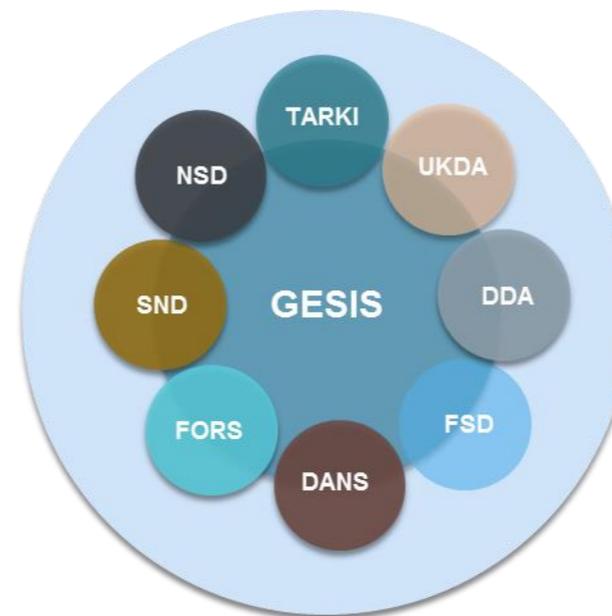
*Azadeh Hashemi, Katarina Boland, Alexander Mühlbauer, Esra Akdeniz, Wolfgang Zenk-Möltgen
GESIS – Leibniz Institute for the Social Sciences*

2018-12-04 | EDDI18 – 10th Annual European DDI User Conference

 cessda.eu  [@CESSDA_Data](https://twitter.com/CESSDA_Data)

Outline

- ◆ Objectives
- ◆ Use Cases
- ◆ Metadata Schema
- ◆ Architecture
- ◆ Technologies
- ◆ Current Prototype
- ◆ Future Development



Project Partners

Objectives

- ◆ The CESSDA Euro Question Bank (EQB) project provides a central search facility across all CESSDA's survey questions in different languages.
- ◆ Survey questions from different Service Providers are included, a big variety of surveys is covered.
- ◆ EQB contains associated information like topic classifications, basic study information, and dataset documentation in different languages.
- ◆ Allows researchers to explore existing survey items for re-use, and lead them to the datasets via the CESSDA Data Catalogue

Use Cases

- ◆ Researchers can find survey questions by using search terms in question text, answer categories, pre- and post-question text, and study title.
- ◆ Filters can be set by topic classification, study series, study year, study geography, interview language, and methodology.
- ◆ Related questions are shown, based on similarities or usage in different waves or countries of a study
- ◆ Several questions can be compared to find differences in wording or answer scales.
- ◆ Found questions can be exported for later usage in new studies.
- ◆ There are more use cases on providing documentation to the EQB, on usage statistics, and more advanced functionalities.

Metadata Schema

- ◊ EQB uses the CMM metadata model. It combines an object structure that is not too complicated with properties that are compatible to the DDI schema.
- ◊ The metadata schema is based on a selection of elements from the DDI-Lifecycle standard and allows different DDI flavors to be used as input for the EQB.
- ◊ This allows Service Providers to use their own implementation of DDI-Lifecycle or DDI-Codebook (or even other xml schemata).
- ◊ It enables researchers to explore questions on the basis of a common model, but still access the original structured metadata.

Metadata Schema

Example: Questionitem

No.	EQB (identifying) element name	UI label	Definition	Occ	lang	type: integer/ string/date	CV/ ISO Code	R
Information on Questions								
1	questions	Questions	A question in EQB can be a single question item or a question grid.	1	n.a	n.a	n.a	
1.1*	questionItem	Questionitem	A question item is a term for a single question that contains a version, ID, name, text, statement, interviewer instructions and number.	0-n	n.a	n.a	n.a	
1.1.1	questionItemVersion	Version	A version number of the question item.	0-1	no	xs:integer	no	
1.1.2	questionItemID	ID	An identifier for a question item the metadata refer-to.	1-n	no	xs:integer	no	
1.1.3	questionItemName	Name	A label of a question item the metadata refers to for display.	0-n	yes	xs:string	no	
1.1.4	questionItemText	Text	A literal question item text.	1-n	yes	xs:string	no	
1.1.5	questionItemStatement	Statement	A prequestion and/or postquestion and other statements as part of the question item text.	0-n	yes	xs:string	no	
1.1.6	questionItemInterviewerinstruction	Instruction	An instruction for the interviewer.	0-n	yes	xs:string	no	
1.1.7	questionItemNumber	Number	A number of the question item the metadata refers to within the associated instrument.	0-n	yes	xs:string	no	

Metadata Schema

◇ Example DDI-Lifecycle

DDI LIFECYCLE XPATHS EXAMPLES

/ddi:DDIInstance/g:ResourcePackage/dc:ControlConstructScheme/dc:Sequence with Ref to: /ddi:DDIInstance/g:ResourcePackage/dc:ControlConstructScheme/dc:Ques

/ddi:DDIInstance/g:ResourcePackage/dc:QuestionScheme/dc:QuestionItem/r:Version

/ddi:DDIInstance/g:ResourcePackage/dc:QuestionScheme/dc:QuestionItem/r:UserID with @typeOfUserID="QuestionID"

/ddi:DDIInstance/g:ResourcePackage/dc:QuestionScheme/dc:QuestionItem/dc:QuestionItemName/r:String with dc:QuestionItemName@context="title-english"

/ddi:DDIInstance/g:ResourcePackage/dc:QuestionScheme/dc:QuestionItem/dc:QuestionText/dc:LiteralText/dc:Text

/ddi:DDIInstance/g:ResourcePackage/dc:ControlConstructScheme/dc:StatementItem/dc:DisplayText/dc:LiteralText/dc:Text

/ddi:DDIInstance/g:ResourcePackage/dc:InterviewerInstructionScheme/dc:Instruction/dc:InstructionText/dc:LiteralText/dc:Text

/ddi:DDIInstance/g:ResourcePackage/dc:QuestionScheme/dc:QuestionItem/dc:QuestionItemName/r:String with dc:QuestionItemName@context="questionNumber"

Metadata Schema

- ◊ Main objects: question items and question grids
 - ◊ With text, interviewer instructions, and statements
 - ◊ With associated concepts and show cards
 - ◊ With response domains: code domain, numeric domain, text domain
 - ◊ Used in instruments that are used by studies
 - ◊ Associated to variables in a dataset that belongs to a study
- ◊ DDI-Lifecycle XPathes are defined as examples
- ◊ Different XML representations are possible

Metadata Schema

◇ Examples of different representations in DDI

```
<QuestionItem isUniversallyUnique="true" versionDate="2018-10-05T08:34:06.7025845Z" xmlns="ddi:datacollection:3_2">
  <r:URN>urn:ddi:int.esseric:f90382ec-02be-4805-b991-ad3f6f6b07e6:1</r:URN>
  <r:Agency>int.esseric</r:Agency>
  <r:ID>f90382ec-02be-4805-b991-ad3f6f6b07e6</r:ID>
  <r:Version>1</r:Version>
  <QuestionItemName>
    <r:String xml:lang="en-GB">ppltrst</r:String>
  </QuestionItemName>
  <QuestionText audienceLanguage="en-GB">
    <LiteralText>
      <Text>Using this card, generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted</Text>
    </LiteralText>
  </QuestionText>
  <CodeDomain blankIsMissingValue="false">
    <r:CodeListReference>
      <r:Agency>int.esseric</r:Agency>
      <r:ID>81c1a5ae-c257-4156-87e5-02731ff2223a</r:ID>
      <r:Version>1</r:Version>
      <r:TypeOfObject>CodeList</r:TypeOfObject>
    </r:CodeListReference>
    <r:ResponseCardinality minimumResponses="1" maximumResponses="1" />
  </CodeDomain>
  <r:ResponseCardinality minimumResponses="1" maximumResponses="1" />
  <r:ConceptReference>
    <r:Agency>int.esseric</r:Agency>
    <r:ID>657ee242-bd45-4aaf-bb55-673cf90240e8</r:ID>
    <r:Version>1</r:Version>
    <r:TypeOfObject>Concept</r:TypeOfObject>
  </r:ConceptReference>
</QuestionItem>
```

Metadata Schema

◇ Examples of different representations in DDI

```
<QuestionItem id="quei-56272875-845c-4b76-bd13-ab76a57e8c02" version="1.0.0" versionDate="2018-06-14T09:08:19.145+02:00">
  <UserID xmlns="ddi:reusable:3_1" type="dk.dda:queitopseudovariid-0.1">V462</UserID>
  <VersionResponsibility xmlns="ddi:reusable:3_1">Birgitte</VersionResponsibility>
  <VersionRationale xmlns="ddi:reusable:3_1" translatable="true" translated="false" xml:lang="da">
    Version: 1.0.0, date: 2018-06-14T09:08:19.145+02:00</VersionRationale>
  <QuestionItemName translatable="true" translated="false" xml:lang="da">Spm. 84: De offentlige
  budgetter er ...</QuestionItemName>
  <QuestionText translatable="true" translated="false" xml:lang="da">
    <LiteralText>
      <Text>Spm. 84: De offentlige budgetter er pressede. Hvilken af strategierne på dette kort,
      mener du, vil være bedst til at afhjælpe dette pres? Du må kun vælge én strategi. Vis kort
      14A</Text>
    </LiteralText>
  </QuestionText>
  <CodeDomain>
    <CodeSchemeReference xmlns="ddi:reusable:3_1">
      <ID>cods-166f963c-8970-445d-bcae-5708f1ee12d1</ID>
      <IdentifyingAgency>dk.dda</IdentifyingAgency>
      <Version>1.0.0</Version>
    </CodeSchemeReference>
  </CodeDomain>
  <ConceptReference>
    <ID xmlns="ddi:reusable:3_1">conc-5cbab215-5b9c-46b5-9cf4-8eb630bbbd15</ID>
    <IdentifyingAgency xmlns="ddi:reusable:3_1">dk.dda</IdentifyingAgency>
    <Version xmlns="ddi:reusable:3_1">1.0.0</Version>
  </ConceptReference>
</QuestionItem>
```

Metadata Schema

Examples of different representations in DDI

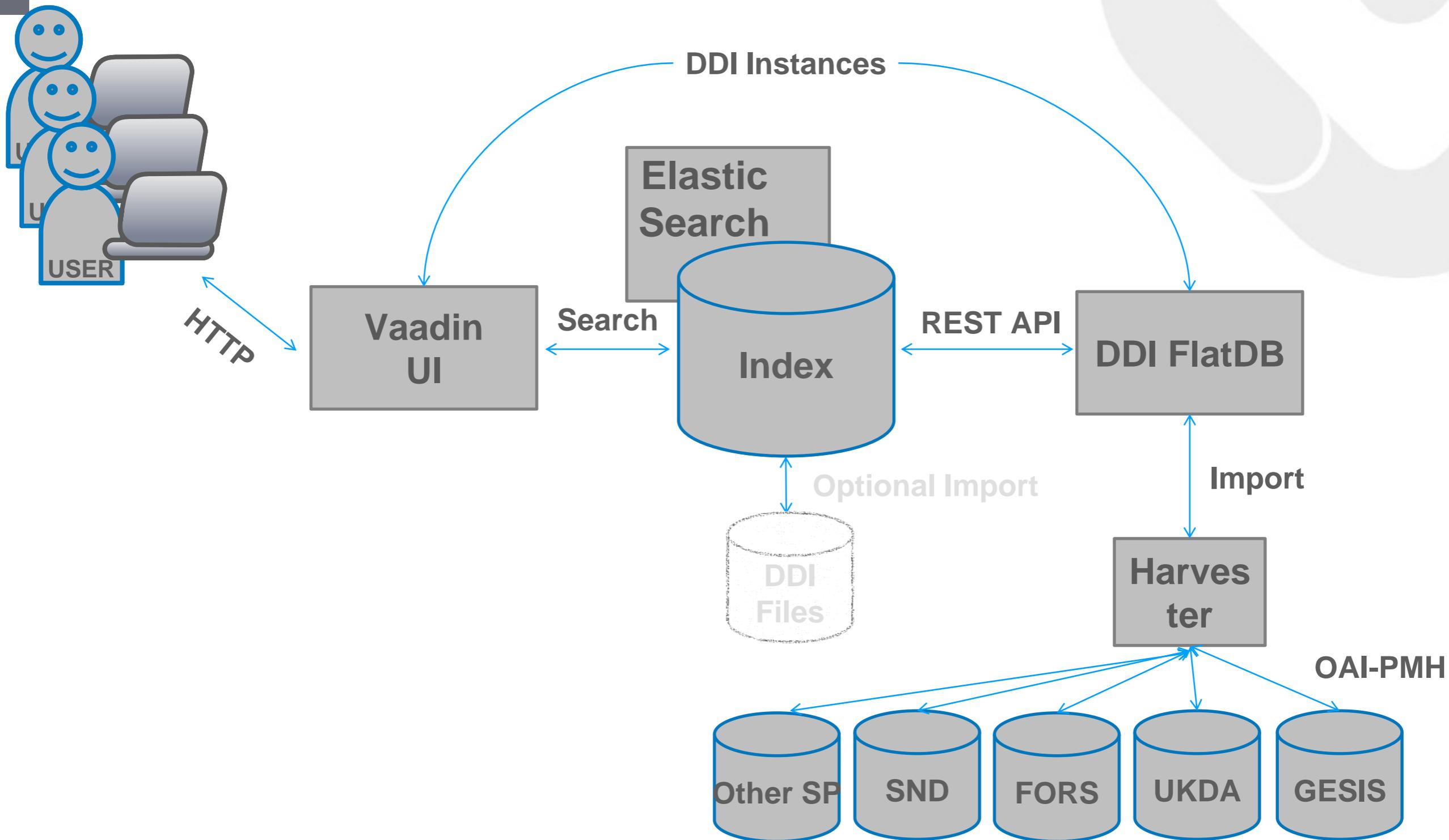
```
<var ID="K17" name="K17" files="F1" dcml="0" intrvl="discrete">
  <labl level="variable">[k17] Onko puolisosi ansiotyössä?</labl>
  <qstn>
    <preQTxt>Ansiotyöksi katsotaan työskentely palkansaajana, ammatinharjoittajana, yrittäjänä tai oman perheen
    yrityksessä vähintään tunti viikossa. Jos puolisosi on tilapäisesti pois ansiotyöstä äitiys-, isyys- tai
    muun loman, palkattoman vapaan tai muun samankaltaisen syyn vuoksi, vastaa hänen tavanomaisen
    työtilanteensa mukaan.</preQTxt>
    <qstnLit>Onko puolisosi ansiotyössä? Valitse vain yksi vaihtoehto.</qstnLit>
    <postQTxt></postQTxt>
    <ivuInstr></ivuInstr>
  </qstn>
  <valrng>
    <range min="1" max="3" />
  </valrng>
  <sumStat type="vald">830</sumStat>
  <sumStat type="min">1</sumStat>
  <sumStat type="max">3</sumStat>
  <sumStat type="mean">1.41</sumStat>
  <sumStat type="stdev">.537</sumStat>
  <catgry>
    <catValu>1</catValu>
    <labl level="category">On tällä hetkellä ansiotyössä &gt; Siirry kysymykseen 18.</labl>
    <catStat>508</catStat>
  </catgry>
  <catgry>
    <catValu>2</catValu>
    <labl level="category">Ei ole tällä hetkellä ansiotyössä mutta on ollut aiemmin &gt; Siirry kysymykseen 19.
    </labl>
    <catStat>303</catStat>
  </catgry>
  <catgry>
    <catValu>3</catValu>
    <labl level="category">Ei ole koskaan ollut ansiotyössä &gt; Siirry kysymykseen 22.</labl>
    <catStat>19</catStat>
  </catgry>
</var>
```



Architecture

- ◊ The EQB system architecture consists of
 - ◊ the EQB-Frontend, implemented with the Vaadin user interface technology, a Java web application framework,
 - ◊ the EQB-Backend, using Elasticsearch for the search index, a Lucene based full-text search engine,
 - ◊ the FlatDB relational database on MySQL to store DDI snippets, a GESIS development to tackle the issue of different usage of the DDI standard,
 - ◊ and the harvester component, used to collect metadata from the different endpoints of Service Providers.

EQB Architecture



Technologies

- ◊ EQB uses Docker, an operating system virtualization tool to manage, centralize, and deploy heterogeneous applications and services.
- ◊ All services are encapsulated as Docker Images and are deployed as Docker Containers.
- ◊ All services are orchestrated by executable configuration.
- ◊ EQB uses the CESSDA Technical Framework as foundation.
- ◊ EQB uses the work of the GESIS internal ExploreData project.

Current Prototype

Language (0) EN ▼

Search ... 🔍 [Reset](#)

Question Text Answer categories Study Title Pre-and post question texts [Search help](#)

FILTER & BROWSE [Questions \(98154\)](#)

Sort by Relevance ▼

CESDA Topic-Classificatio

Series

Study Year

Geography

Interview language

Methodology

[Reset all filters](#)

<input type="checkbox"/> Compare <input type="checkbox"/> Basket	Q.B26: Which provider(s) does your household use for its Internet service? Response Categories ZA4416_SU "Eurobarometer 64.4 (Nov 2005- Jan 2006)"	Detailed Question View Language <input checked="" type="radio"/> EN <input type="radio"/> DE <input type="radio"/> FI
<input type="checkbox"/> Compare <input type="checkbox"/> Basket	Q.B26: Which provider(s) does your household use for its Internet service? Response Categories ZA4416_SU "Eurobarometer 64.4 (Nov 2005- Jan 2006)"	Detailed Question View Language <input checked="" type="radio"/> EN <input type="radio"/> DE <input type="radio"/> FI
<input type="checkbox"/> Compare <input type="checkbox"/> Basket	Q.B14: Please tell me which of the following features would make you personally switch to an advanced mobile phone service. Response Categories ZA4416_SU "Eurobarometer 64.4 (Nov 2005- Jan 2006)"	Detailed Question View Language

Page 1 of 9816 10 per page ▼

Current Prototype

democracy

◇ Search

Question Text Answer categories Study Title Pre-and post question texts

[Search help](#)

◇ Filter

FILTER & BROWSE

CESSDA Topic-Classificatio

Series

Study Year

Geography

Interview language

Methodology

[Reset all filters](#)

Current Prototype

Filter

Filter according to: CESSDA Topic-Classification +

When sorting the studies into the CESSDA Topic Classification, studies were organized into the respective subcategories of a main category, if possible. In case a study could be assigned to all subcategories of a specific main category or to none, then the study was put into the main category only. To select these "main category only" studies, please unset the check mark from the subcategories.

The subcategories are also not disjunctive, which means that a study can be sorted into different subcategories.

Select: CESSDA Topic-Classification

Search ...

- ▶ Labour and employment (168)
- ▼ Demography and population (56)
 - Fertility (1)
 - Migration (42)
 - Censuses (13)
- ▶ Education (40)
- History (5)
- ▶ Society and culture (660)
- ▶ Law, crime and legal systems (37)
- ▶ Health (82)
- ▶ Trade, industry and markets (8)

[Reset selection](#) [Apply](#) [Cancel](#)

FILTER & BROWSE

CESSDA Topic-Classification ON

× Migration

Series ON

× International..

Collection year

Geography ON

× Finland

× Germany

× Italy

Language

Mode of collection

[Reset all filters](#)

Current Prototype

- ◆ Results
- ◆ Compare
- ◆ Links to details
- ◆ Languages

Questions (523)

- Compare
- Basket

Q11:V18Q11:V18: Split A: On the whole, are you very satisfied, fairly satisfied,

[Response Categories](#)

[ZA2251_SU "Central and Eastern Eurobarometer 2 \(Current Affairs and the Media\)"](#)

Sort by

Relevance

not very satisfied or not satisfied at all with the way democracy

[Detailed Question View](#)

Language

EN DE FI

Current Prototype

Questions

 [Variable description.pdf](#)

 **Variable qa12a:** "DEMOCRACY SATISFACTION - COUNTRY"

Question wording:

On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the way democracy works in (OUR COUNTRY)?

Interview instruction:

SHOW CARD WITH SCALE – ONE ANSWER PER COLUMN - READ OUT

Response category:

Value	Value label
1	Very satisfied
2	Fairly satisfied
3	Not very satisfied
4	Not at all satisfied
5	DK

Notes:

1. Last trend: EB80.1, QA18A&B

todo: display question with related variable information

Future Development

- ◆ Objectives for CESSDA EQB 2019
 - ◆ Launch a stable und user-friendly search service
 - ◆ Expand coverage to more Service Providers
 - ◆ Support more metadata format flavours
 - ◆ Conduct training and release documentation

Thank you!

Contact: Wolfgang.Zenk-Moeltgen@geis.org

This presentation is offered under the CC-BY 4.0 license. This license does not apply to the following copyrighted material used in this presentation: the logo of GESIS, the logo of CESSDA, the slideshow layout of CESSDA.