Deliverable Number: D3.11
Deliverable Title: Report on application of indexing terms in the data lifecycle
Work Package: WP3: Maximising equivalence through translation

Deliverable type: Report
Dissemination status: Public

Submitted by: CESSDA ERIC (UK Data Archive)
Authors: Lorna Balkan
Date Submitted: June 2018

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654221.
SERISS (Synergies for Europe’s Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe’s social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe’s grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey of Health Ageing and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Balkan, L (2018) Report on application of indexing terms in the data lifecycle Deliverable 3.11 of the SERISS project funded under the European Union’s Horizon 2020 research and innovation programme GA No: 654221. Available at: www.seriss.eu/resources/deliverables
## Contents

Summary .......................................................................................................................................................... 4

1 Introduction .................................................................................................................................................. 6

2 Background .................................................................................................................................................. 6

2.1 Overview .................................................................................................................................................. 6

2.2 What is indexing? ................................................................................................................................... 6

2.3 The data lifecycle ..................................................................................................................................... 7

2.4 Indexing data ........................................................................................................................................... 11

2.5 Online question banks ............................................................................................................................ 14

2.6 Summary ................................................................................................................................................ 20

3 Methodology ............................................................................................................................................... 20

3.1 Overview ................................................................................................................................................ 20

3.2 Indexing process ...................................................................................................................................... 20

3.3 Thesaurus ............................................................................................................................................... 20

3.4 Selection and training of indexers ........................................................................................................... 21

3.5 Indexer instructions ................................................................................................................................. 21

3.6 Selection of questions ............................................................................................................................... 22

3.7 Representation of questions ..................................................................................................................... 23

4 Results ....................................................................................................................................................... 25

4.1 Overview ............................................................................................................................................... 25

4.2 Difficulty in assigning terms to questions ............................................................................................. 26

4.3 Match of ELSST terms to question content .......................................................................................... 26

4.4 Suggested new terms ............................................................................................................................... 27

4.5 Indexing consistency ............................................................................................................................... 28

4.6 Conclusion .............................................................................................................................................. 28

5 Discussion and recommendations ............................................................................................................... 28

5.1 Overview ............................................................................................................................................... 28

5.2 Indexers ............................................................................................................................................... 29

5.3 Questions .............................................................................................................................................. 30

5.4 ELSST terms .......................................................................................................................................... 32

5.5 Conclusion ............................................................................................................................................ 34
Summary
The aim of this deliverable is to investigate the application of indexing terms in the data lifecycle. In particular, it investigates how and where to apply index terms to survey questions. The work describes an experiment to index survey questions drawn from three cross-national surveys: the European Social Survey (ESS), the European Values Study (EVS), and the Survey of Health, Ageing and Retirement in Europe (SHARE) using terms from the European Social Science Thesaurus (ELSST).

A secondary aim of the deliverable is to investigate how well ELSST terms cover the subject content of the selected questions from the above three surveys. The ELSST terms used in the experiment were taken from the German, Greek, and Romanian versions of the thesaurus to index survey questions in the corresponding languages.

The deliverable begins by providing a general definition of indexing, before looking at indexing survey data in particular, and especially survey questions. To do this, the data lifecycle and data creation process are discussed. It also looks at previous approaches to indexing questions and at online question banks which allow users to access survey questions.

It then describes the experiment that was carried out to assign ELSST terms to survey questions from ESS, EVS and SHARE. Results are reported and analysed. They show that while some of the difficulties in indexing questions are associated with the nature of the questions and/or ELSST terms, others lie with the indexers and the indexing technique adopted. The extent to which the ELSST terms were perceived to match the content of the question is also affected both by the coverage and nature of the ELSST terms.
as well as by the knowledge and indexing style adopted by the indexer. Recommendations are made both on how to index questions and how to promote quality assurance of ELSST terms.

Results of the question indexing task are then compared with the indexing of the same surveys at whole study-level that was reported in SERISS D.9 deliverable. They show that both techniques require extensive knowledge on the part of the indexer of the data, as well as of the thesaurus and indexing process. Recommendations are made concerning how and where to index in the data lifecycle.

The experiment thus contributed to the ongoing debate on how to index survey questions, and where and when to index in the data lifecycle. It also provided useful feedback on the quality of ELSST terms.
1 Introduction
The aim of this deliverable is to investigate the application of indexing terms in the data lifecycle. In particular it investigates how and where to apply index terms to survey questions. It describes an experiment to index survey questions drawn from three cross-national surveys: the European Social Survey (ESS), the European Values Study (EVS), and the Survey of Health, Ageing and Retirement in Europe (SHARE) using the European Social Science Thesaurus (ELSST).

A secondary aim of the deliverable is to investigate how well ELSST terms cover the subject content of the selected questions from the above three surveys. ELSST terms from the German, Greek, and Romanian versions of the thesaurus were used to index the questions in the corresponding languages.

The deliverable starts by introducing the process of indexing in general, then discusses the data lifecycle and data creation process before looking at issues surrounding the indexing of data. It then focuses on the indexing of survey questions in particular. It also looks at how users currently access survey questions via online question banks and other systems.

We then describe the experiment we carried out to assign index terms to survey questions from ESS, EVS and SHARE. Results are discussed and analysed. Results are also compared with the indexing of the ESS and EVS studies at whole study-level that was reported in the SERISS D.9 deliverable.

Recommendations are made concerning how and where to index in the data lifecycle, as well as on improvements to the thesaurus.

2 Background

2.1 Overview
This section starts by defining what indexing means in general. It then discusses the data lifecycle and, in particular, the data creation process, before looking at how to index survey data, in particular survey questions. Previous approaches to indexing survey questions are discussed, as well as online systems for searching and browsing survey questions.

2.2 What is indexing?
Indexing is the process of evaluating information and assigning index terms to it. It is used to assist users in information retrieval. This report is more specifically about subject indexing, i.e. the assignment of subject terms, but we will refer to it simply as “indexing”.

The ISO indexing standard (ISO 5963:1985 (last reviewed and confirmed in 2015)) describes a 3-step process for assigning subject terms to documents:

1. examine the document and establish its subject content
2. identify the principal concepts present in the subject
3. express these concepts in the terms of the indexing language
Establishing what the document is about (steps 1 and 2) is not always easy, since the subject content is not always stated explicitly (for an overview see Lancaster 2003, pp. 13-19) and may involve knowledge that is shared by the creator(s) and user(s) of the text (Golub et al. 2017, p.5). Moreover, indexing involves not only deciding what a document is about but also why it is likely to be of interest to a particular group of users (Lancaster 2003, p.9; ISO 5963:1985).

These indexing goals affect two major parameters in indexing:

- exhaustivity
- specificity

Exhaustivity refers to the extent to which the contents of a work are covered by the terms used in indexing, and relates to breadth of coverage of indexing. Specificity, on the other hand, relates to depth of treatment of content (Lancaster 2003, p.204). In general, it is good indexing practice to use the most specific term available to describe some feature of a document (the so-called Principle of specificity) (Lancaster 2003, and ISO 5963:1985). Lancaster gives as an example an article discussing the cultivation of oranges, which should be indexed with ORANGES rather than CITRUS FRUITS or FRUITS. If the indexing language does not include a term at the required level of specificity then a broader term can be used (or a request can be made to add the term to the vocabulary). Alternatively, a combination of existing terms can be used to cover the concept (Lancaster 2003, p.35). In some circumstances, however, more general concepts may be preferred if, for example, the subject matter is not central to the document (ISO 5963:1985).

Index terms may be selected from a controlled vocabulary, typically a thesaurus, or assigned freely. The advantage of using a controlled vocabulary is that it controls for both synonymy and ambiguity, producing more consistent and accurate indexing. Synonymy is controlled by defining two types of term: Preferred Terms (PTs) and Use For terms (UFs), which are typically synonyms or near-synonyms of the Preferred Term. Only Preferred Terms are used for indexing, while UFs are used to guide the user to a Preferred Term. Ambiguity is controlled by restricting a term to refer to one concept only, unlike everyday language.

Controlling for both synonymy and ambiguity increases the efficiency of search, since the user needs to query the database only once, using either the Preferred Term or any of its Use For terms, in order to retrieve objects about the same concept. Moreover, the data received will contain no ‘false positives’.

Terms in a controlled language are typically arranged hierarchically. Thus, if the user interface allows, users may broaden or narrow their search by choosing terms higher or lower in the hierarchy. A term in a controlled language may also be associated with other, non-hierarchically related terms, which offer the user yet more search suggestions.

2.3 The data lifecycle
The data lifecycle starts with planning research and extends beyond the project for which the data are originally created to secondary use.
The data lifecycle can be represented graphically (see Figure 1, taken from the UK Data Service’s ‘research data lifecycle’ website).

![Data Lifecycle Diagram](image)

**Figure 1 Data lifecycle, UK Data Service**

The first three stages of the lifecycle are of most relevance to the current report. The data creation process spans the research planning and data collection stages, and is discussed in Section 2.3.1 below. The processing data stage is where data are digitised, checked, validated, anonymised, described and documented, and managed and stored.

Data documentation is essential to ensure that data can be understood during a research project, and reused by other researchers at the end of the project. It explains how data were created or digitised, what data mean, what their content and structure are, and any manipulations that may have taken place. The UK Data Service recommends (see its website on ‘prepare and manage data’) that data documentation should be begun at the onset of a project and continued throughout the research. It also recommends that documentation for research data contain both study-level information about the research and data creation, as well as descriptions and annotations at the variable, data item or data file level.

Metadata are part of the data documentation process. They describe the content, context and provenance of datasets in a standardised and structured manner, and are used for resource discovery, providing searchable information that helps users to find data, or for online data browsing (see the UK Data Service ‘catalogue metadata’ website). During data processing, data cataloguers use information in
the documentation they receive from the data creator to create a metadata record, typically including index terms from a thesaurus.

Most social science data archives in the world use a metadata standard called the Data Documentation Initiative (DDI), which was originally designed for describing social, behavioural and economic sciences data.

2.3.1 Data creation process

In survey research, the researcher starts by formulating research questions about interrelated concepts. The next step is to define the concepts, and identify corresponding variables one wants to measure. Variables refer to different aspects or dimensions of the concept. This is followed by developing procedures or operations for selecting indicators to measure variables (a process called ‘operationalisation’) (Engel and Schutt 2014). Indicators, usually survey questions in the context of a questionnaire, indicate the values of cases on a variable.

Engel and Schutt (2014) give as an example the concept ‘income’ which can be represented by the variable ‘annual income’, and this variable can be measured with responses to the single question, or item “What was your total income from all sources in 2012?”

More complex concepts may need to be subdivided into a number of variables. Wolf et al. (2016, p.200) give an example of a more complex concept taken from the ESS survey. ‘Social legitimacy of the welfare state’ is the core theoretical concept in the ESS social welfare module. This concept is analysed and subdivided into a number of dimensions, including ‘popular perceptions of welfare state consequences’. This dimension in turn is subdivided into a number of sub-dimensions including economic, moral and social consequences. In the operationalisation stage of the dimension ‘popular perceptions of welfare state consequences’, indicators for each of the sub-dimensions of the construct are selected. They are all indicators of the form ‘to what extent do you agree or disagree that social benefits and services in [country]...”

D21... place too great strain on the economy?
D22... prevent widespread poverty?
D23... lead to a more equal society?
D25... cost business too much in taxes and charges?
D26... make it easier for people to combine work and family life?
D27... make people lazy?
D28 ... make people less willing to care for another?
D29... make people less willing to look after themselves and their family?’
This is shown graphically in Figure 2 below.

![Figure 2 Operationalisation of the perceptions of welfare state consequences](image)

It is important to note that not all variables appear explicitly in the questionnaire.

After data collection using the questionnaire, data are coded, i.e. a numerical value is assigned to responses to facilitate data processing and analysis. Some questions have coded response categories on the questionnaire, others are coded after collection during a manual or automated process (Statistics Canada 2010, p.3). Standardised coding schemes may be used, such as the International Standard Classification of Education (ISCED) (UNESCO, 2011) for coding education. The meaning of all codes must be documented (CESSDA’s Expert tour guide on data management).

Note that DDI metadata are available for all levels of a study, including question level. The Euro Question Bank (EQB) project, which aims to create a search facility across all CESSDA’s survey holdings, is also actively working on metadata for questions, in collaboration with the DDI.

2.3.2 Types of survey questions
Two main types of questions can be found in questionnaires (definitions and examples are taken from Statistics Canada, 2010, pps.62-68):

- open: Open questions are questions where response categories are not provided to the respondent. For example:
  
  *What is the most important problem facing Canada?*

- closed: Closed questions are where response categories are listed with the question.

Closed questions come in a variety of forms. Below are examples of the most common:

- two-choice questions: these have only two answers. For example:
Did you smoke cigarettes yesterday?

- Yes
- no

- multiple choice questions: multiple choice questions ask the respondent to select one response from a list of choices. For example:

  How old are you in years?
  - Under 20
  - 20-29
  - 30-39
  - 40-49
  - 50 or more

- checklist questions: checklist questions ask the respondent to pick one or more responses from the list. For example:

  During your vacation, in what types of accommodation did you stay?
  - Hotel (including tourist home)
  - Motel
  - Camping or trailer van
  - Home of friends or relatives
  - Commercial cottage or cabin
  - Other (hostels, universities, etc.)

- rating questions: respondents are asked to rate their answer. For example:

  How satisfied are you with our customer service?
  - very satisfied
  - satisfied
  - dissatisfied
  - very dissatisfied

A collection of related questions that measure an underlying concept is known as a scale. Scales may be used in preference to single-item measures whose reliability is affected by the fact that responses are markedly influenced by contextual factors such as the preceding item (Hupper et al. 2008, p.3). The SHARE survey, for example, includes the EURO-D scale to measure depression (Dewey and Prince 2005).

2.4 Indexing data

2.4.1 Traditional approaches to indexing data

At the end of a research project, data are archived, together with their associated documentation, including the questionnaire and abstract. Together, this constitutes a study.

Typically, indexing is applied to a study as a whole (which we will refer to henceforth as ‘study-level’ indexing). Study-level indexing may be more or less specific. At its most general level, it captures the key themes of the study (e.g. ‘social attitudes’) which may be referred to in the study title or abstract but which may not be specifically referred to in the data or questionnaire. At its most specific level, study-level indexing may also include index terms for variables or questions.
For example, indexing practice for survey data at the UK Data Archive (UK Data Archive 2017) is generally to index at question level, rather than variable level since indexing at specific variable level “may result in disappointment for users if they choose a study from the catalogue based on the index terms allocated, to find that only one variable out of a large study covers the concept they are looking for. Similarly, using data dictionaries (either those supplied by the depositor or those generated during data curation) is not always useful, as some variables listed may not be populated. Adding index terms to cover these variables may also mislead users if no relevant data are available.” (p.7). Nevertheless, it states that the indexer should ensure that all variables are covered where possible, including derived variables, such as derived measures of social class, socio-economic status, etc. which will be mentioned in the documentation but may not always be included in the questionnaire.

Study-level indexing may result in very long lists of index terms (the UK Data Archive assigns up to 450 terms to certain studies). Moreover, where a combination of terms has been used to represent a concept (e.g. if the terms CHILDREN and HEALTH are used to cover the concept CHILD HEALTH), it is difficult to identify the association between these terms in a long list of keywords.

As with document indexing, index terms applied to data may be drawn from thesauri and other controlled vocabularies, as is the practice at the UK Data Archive, or may be freely assigned (as is the practice at the GESIS-Leibniz-Institute for the Social Sciences). The Data Archive uses The Humanities and Social Science Electronic Thesaurus (HASSET), a monolingual thesaurus from which ELSST has been derived.

While study-level indexing is typically carried out by trained indexers, in self-archiving contexts such as the UK Data Service’s ReShare, data creators may be responsible for assigning their own index terms to their study.

2.4.2 Previous approaches to indexing questions

2.4.2.1 Introduction

As noted above, index terms are usually assigned at study level. Study-level indexing may include index terms pertaining to individual variables or questions, but it is not possible to identify which term is associated with any one variable or question. In this section we look at an alternative to study-level indexing, namely question-level indexing, i.e. where each question is indexed separately, and can thus more easily be findable for researchers. While this is a relatively new field, the increasing availability of question/variable banks, where users can search for individual questions and/or variables, is making it a more attractive option. We review what little work there has been in this field below.

2.4.2.2 Friedrich and Siegers (2016)

Friedrich and Siegers (2016) propose an approach to indexing survey questions based on work on indexing images. In image indexing, a distinction is frequently made between ofness and aboutness of images (see for example Shatford 1986, and Shatford Layne 1994). ‘Ofness’ is what a picture consists of, i.e. what is visibly depicted, while ‘aboutness’ is what the picture is about, i.e. what it signifies. Shatford Layne (1994) gives the example of a person crying which might be about sorrow.
Friedrich and Siegers (2016) suggest that the notions of ‘ofness’ and ‘aboutness’ can usefully be applied to indexing survey questions. The explicit content of a question, as expressed by its literal wording, corresponds to the ‘ofness’ (or ‘topic’) of the question, while ‘aboutness’ (or ‘attributes’) of a question correspond to the construct the primary investigator wishes to measure. They propose that questions should be indexed for both their topics and attributes. They note that in many cases a question’s attributes are not directly expressed in the question’s wording and require the indexer to examine a question’s context, for which both subject knowledge and data literacy are needed.

They distinguish between ‘subjective’ and ‘directive’ terms. Subjective terms can be “any subject area that is relevant to the social sciences” (p.635) and can be used to index the topic of a question, while directive terms can be used to index attributes of a question. Directive terms fall into four broad classes, capturing cognition, evaluation, emotion and action (see Figure 3).

<table>
<thead>
<tr>
<th>Cognition</th>
<th>Evaluation</th>
<th>Emotion</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCEPTION</td>
<td>ATTITUDE</td>
<td>MOOD</td>
<td>BEHAVIOUR</td>
</tr>
<tr>
<td>KNOWLEDGE</td>
<td>PREFERENCE</td>
<td>FEAR</td>
<td>USE/UTILIZATION</td>
</tr>
<tr>
<td>AWARENESS</td>
<td>JUDGMENT</td>
<td>ANGER</td>
<td>CHOICE</td>
</tr>
<tr>
<td>INTEREST</td>
<td>PREJUDICE</td>
<td>HAPPINESS</td>
<td>INTERACTION</td>
</tr>
<tr>
<td>BELIEF</td>
<td>SATISFACTION</td>
<td>HATE</td>
<td>COMMUNICATION</td>
</tr>
</tbody>
</table>

*Figure 3 Examples of directive terms from four attribute classes*

They propose that directive terms associated with topic terms can be suggested as facets at retrieval time so that users can narrow their search.

They illustrate with the following examples from questions about corruption in Europe in the European Barometer 76 survey. The first statement:

*There is corruption in the national public institutions in Germany*

can be indexed with the subject terms CORRUPTION and PUBLIC INSTITUTIONS. Additionally, it can be assigned the directive term PERCEPTION, because it is asking how the respondent perceives corruption.

The second statement:

*Are you personally affected by corruption in your daily activities?*

would be indexed with the subject terms CORRUPTION and EVERYDAY LIFE. This time, however, since the respondent is being asked if they have personal experience of corruption, the statement can be assigned the directive term EXPERIENCE.

2.4.2.3 Friedrich and Kempf (2014)

Friedrich and Kempf (2014) describe a conceptual model for indexing survey data that captures the different levels of a study as far as subject content is concerned. Starting with the most general, these levels are (a) the study as a whole (b) the individual variables and (c) the questions. Indexing at study level captures what the study is about; indexing at variable level captures all the individual phenomena that have been studied; and indexing at question level captures what questions have been asked and answered.
They propose a layered or ‘cascading’ model of indexing where index terms are assigned according to the different levels of the study (i.e. whole study, variable and question), and where lower levels inherit the index terms from the next highest level.

<table>
<thead>
<tr>
<th>Analytical levels with examples of information found in additional study material</th>
<th>Assigned and implied terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study level</td>
<td>[...] We studied xenophobia in a population of 12-25 year old Europeans. [...]</td>
</tr>
<tr>
<td>Variable level</td>
<td>[...] We applied a measurement of ethnocentricity. [...]</td>
</tr>
<tr>
<td>Question level</td>
<td>[...] It would be good if every year more foreign young people came to our country to study. [...]</td>
</tr>
</tbody>
</table>

Figure 4 Cascading indexing model

Friedrich and Kempf intend this model to show how questions cannot be indexed in isolation from their wider context if you want to capture more than just their literal meaning.

2.5 Online question banks

2.5.1 Overview

As noted above, question/variable banks are becoming increasingly popular, either integrated into an online data catalogue or as standalone systems. Below we present some examples from member archives of the Consortium of European Social Science Data Archives (CESSDA) and other archives. Links to other question banks can be found on the UK Data Service website at [https://www.ukdataservice.ac.uk/get-data/other-providers/question-banks/uk](https://www.ukdataservice.ac.uk/get-data/other-providers/question-banks/uk) and [https://www.ukdataservice.ac.uk/get-data/other-providers/question-banks/non-ukcross-national](https://www.ukdataservice.ac.uk/get-data/other-providers/question-banks/non-ukcross-national)

2.5.2 Nesstar

Nesstar is a long-standing tool developed by the Norwegian Centre for Research Data (NSD) that allows users to search, browse and analyse data at variable level.

Both ESS and EVS data are available to view and download via Nesstar applications: ESS via the ESS website, and EVS via the Zacat catalogue at GESIS.

Nesstar has also been implemented by many CESSDA and CESSDA-related archives, including the following:

- ADP Nesstar: [http://nesstar2.adp.fdv.uni-lj.si/webview/](http://nesstar2.adp.fdv.uni-lj.si/webview/)
Nesstar supports search within both variables and questions. Below is a view of the ESS Nesstar application. Variables are shown together with their question text and any pre- and post-question text, and the response categories. Users can also create simple online cross-tabulations, produce graphs, and download subsets of variables in a variety of formats.

![ESS Nesstar application](image)

Figure 5 ESS Nesstar application

### 2.5.3 Variable and Question bank, UK Data Service

The ‘Variable and Question bank’ of the UK Data Service allows users to search and browse variables and questions from the UK Data Service’s survey datasets. There is the option to search only question text/responses, or those that belong to a set of harmonised concepts and questions (see Office for National Statistics 2011). A special feature is the ‘question intent’ filter which distinguishes between behavioural and attitudinal questions. This is similar to the ‘directive’ terms of Friedrich and Siegers (2016). Currently, this feature is manually assigned.
The results page has links to the previous and next variables, as well as to Nesstar and the related survey catalogue record in the UK Data Service data catalogue, Discover.

2.5.4 Aila data catalogue, FSD

The Aila data catalogue at the Finnish Social Science Data Archive (FSD) offers a ‘questions and variables’ search option as part of its advanced search. Search, in either Finnish or English, can be narrowed to:

- question text (including variable label, literal question, pre-question text, and post-question text)
- category label (i.e. the response options)

Results of a search are shown with variable label, question text, frequencies, and summary statistics, and include a link to the previous and following variable, as well as to the study description. For example:
2.5.5 CLOSER Discovery

CLOSER Discovery is the search engine of the CLOSER project. It enables researchers to search and browse questionnaires and data from eight leading UK longitudinal studies. Users can narrow their search to questions or variables.

Additionally, users can search or browse by topic. Each question and variable is mapped to a topic drawn from a controlled vocabulary (for more details see the CLOSER confluence page on topics). The project is currently investigating whether it is viable to map items to only one topic but notes that assigning too many topics per item “would become unwieldy from a user point of view” (CLOSER topics page). CLOSER is also investigating how scales linked to instruments that use them can be added to the search engine (see the CLOSER confluence page on scales).
Searching for “newspapers” within questions, the results page shows the question text, with the variable name and label, the place of the question within the questionnaire, the preceding and following questions, and has link to the complete questionnaire from which the question is taken and the associated study. See Figure 9:
While some question banks have been integrated into the main data catalogue (e.g. Aila) others (e.g. UK Data Service Variables and Question bank) offer separate user interfaces. However, most question banks provide links to the main study description. Some, like CLOSER, offer a direct link to the questionnaire from which the questions are taken.

All question banks provide the question together with its metadata, including variable, variable label, pre-question text, and question responses. They also provide a link to the next and previous variable or question(s).

With the exception of CLOSER, none of the question banks we found uses a controlled vocabulary for indexing their questions/variables.
2.6 Summary
In this section we looked at what it means to index, and the special challenges posed by indexing data in general, and survey questions in particular. The next section discusses the work that was carried out in the present indexing task.

3 Methodology

3.1 Overview
This section describes the main decisions regarding the methodology for the question indexing exercise. It covers the choice of thesaurus; the selection and training of indexers; indexer instructions; and the choice and representation of questions.

3.2 Indexing process
The indexing was carried out in three stages:

1. Indexers assigned index terms to questions using the instructions in Section 3.5 below.
2. The reviewer gave each indexer 1:1 feedback on their indexing.
3. Indexers revised their indexing on the basis of the feedback.

The second step was introduced because the indexers were unfamiliar with ELSST. The author of this report, who has excellent knowledge of ELSST, acted as reviewer and suggested terms to indexers that had been proposed by the other two indexers. The indexers were free to reject or accept these suggestions as they liked.

3.3 Thesaurus
The version of ELSST used for the experiment was the published version of 6 September 2016.

In ELSST, there are two types of term: Preferred Terms and Use For terms. Preferred Terms describe the concept in question, and are used for indexing. A Use For term (UF), or non-preferred term, is not used for indexing, but guides the user to a Preferred Term instead. Use For terms are typically synonyms or near synonyms, or antonyms. In some cases, a Use For term may be narrower in scope than the Preferred Term. For example, PARENTAL BENEFITS in ELSST has the UFs MATERNITY ALLOWANCES and PATERNITY BENEFITS.

The rationale for treating a term that is narrower in scope to the term in question as a Use For term rather than a Preferred Term in its own right is explained in ISO 25964-1 (2011), the latest international guidelines on thesaurus construction (see p.48):

“This technique is sometimes used to reduce the number of preferred terms in a thesaurus. It limits the specificity of indexing and searching that can be achieved, and is beneficial when the collection to be indexed has very little information on the subject area in question. The presence of the non-preferred terms provides extra entry points to assist user access. Subsumption of narrower concepts in a broader one should not, however, be used when there is a wealth of information on the subject in question and greater specificity is needed to pinpoint concepts with accuracy.”
Preferred Terms are arranged in hierarchies, represented by Broader Terms (BTs) and Narrower Terms (NTs), where the scope of the Narrower Term falls within the scope of the Broader Term. They may also be associated with other Related Terms (RTs), to which they are related, but not hierarchically. For example, NURSES has the following relationships:

NURSES
   BT HEALTH PROFESSIONALS
   NT MIDWIVES
   RT NURSING CARE
   RT OCCUPATIONS

Additionally, a Preferred Term may be associated with a number of note fields. Scope notes describe what the term means in the context of the thesaurus; Scope note source describes the source of the scope note; Use notes help indexers to know when to apply a term, and to distinguish it from closely related terms; History notes describe the history of the term. An example from ELSST:

MARRIAGE:
   **Scope note**: A LEGALLY RECOGNIZED RELATIONSHIP BETWEEN ADULTS THAT CARRIES CERTAIN LEGAL RIGHTS AND OBLIGATIONS
   **Scope note source**: UK DATA ARCHIVE
   **Use note**: DO NOT USE FOR MEASURES OF MARRIAGE RATES OR FOR THE CEREMONY OF MARRIAGE.
   **History note**: THE UF "SAME-SEX MARRIAGE" WAS ADDED JULY 2015. PREVIOUSLY "SAME-SEX MARRIAGE" WAS A UF OF "CIVIL PARTNERSHIPS".

3.4 Selection and training of indexers
Three indexers were recruited specifically for this task, one for each language: German, Greek and Romanian. They were all native speakers of the languages of the questions they were asked to index and had an excellent knowledge of English. The German indexer was a linguist and teacher, and the Greek and Romanian indexers were both post-graduate social scientists currently employed at the UK Data Archive. None of them had prior experience of indexing, but were given training in the use of ELSST and in how to index, using the indexing procedures described in the following section.

3.5 Indexer instructions
Indexers were instructed to find the online version of the questionnaire they were asked to index, and begin by reading the relevant documentation. They were then given a list of questions to index, and asked to cut and paste them into an Excel template (see Appendix 3 for the EVS question template). For each survey they were asked to:

- assign ELSST terms to the selected questions
- comment on how well the terms matched the content of the question (see below)
- suggest any terms missing from ELSST
- rate how difficult it was to assign index terms on a scale 1-5, where 1 is very easy and 5 is very difficult
- comment on why it was difficult to assign index terms where relevant
The following degrees of equivalence between the meaning expressed by the questions, and the meaning expressed by ELSST terms, are distinguished:

a. **Exact** equivalence: there is no difference in the meaning expressed by the question and by the ELSST term(s)
b. **Inexact** equivalence: there is a small difference in the meaning expressed by the question and by the ELSST term(s)
c. **Partial** equivalence: there is a partial overlap in the meaning expressed by the question and by the ELSST term(s) because either:
   i. the meaning expressed by the question is **narrower** than that expressed by the ELSST term(s)
   ii. the meaning expressed by the question is **broader** than the content of the ELSST term(s)
   iii. the meaning expressed by the question **partially overlaps** with the meaning expressed by the ELSST term(s)
   iv. **Non-equivalence**: there is no overlap in the meaning expressed by the question and that expressed by the ELSST term(s)

d. **Don’t know**

Full instructions for German are shown in Appendix 1. The instructions for the other languages differed only in the language of the examples and the links to the relevant questionnaires.

Indexers were asked to follow the indexing rules of specificity and exhaustivity (a copy of these guidelines can be found in Appendix 2). Note that they were encouraged to use the most specific terms they could find.

### 3.6 Selection of questions

#### 3.6.1 Overview

Questions were selected for indexing from the following two surveys in German, Greek and Romanian:

- European Social Survey (ESS)
- European Values Study (EVS)

Additionally, German and Greek questions were indexed from:

- The Survey of Health, Ageing and Retirement in Europe (SHARE)

(SHARE is not available in Romanian.)

For all surveys, the language versions of the questionnaires used were: German (Germany); Greek (Greece) and Romanian (Romania). The choice of languages was dictated principally by the personnel available to undertake the indexing work (see Section 3.4).

#### 3.6.2 European Social Survey (ESS)

A total of 62 questions were indexed in German, Greek and Romanian.
The latest version of ESS that was available for each language was ESS 7 (2014) for German, ESS 5 (2010) for Greek, and ESS 4 (2008) for Romanian. Core questions where the English source question had not changed across Versions 5, 6 and 7 were selected for indexing.

The selected questions covered the following subject areas:

- television viewing and social trust
- politics
- subjective wellbeing, social exclusion, religion, perceived discrimination, national and ethnic identity, immigration

The question numbering in the examples below is taken from ESS 7.

### 3.6.3 European Values Study (EVS)

A total of 121 questions were indexed in German, Greek and Romanian.

Questions were selected from EVS 4 (2008). Only those that had appeared in all 4 waves of the study were selected, since they were regarded as more important. They covered the following subject areas:

- perceptions of life
- work
- family
- politics and society
- religion and morale

### 3.6.4 The Survey of Health, Ageing and Retirement in Europe (SHARE)

A total of 51 questions were indexed in German and Greek (this survey is not available in Romanian).

The version of the survey chosen for indexing was SHARE Version 6.

The questionnaires were drawn from the CAPI main questionnaire. The modules that were chosen for indexing had appeared in waves 1, 2, 4, 5 and 6, and were as follows:

- PH: Physical health (36 questions)
- BR: Behavioral risks (15 questions)

### 3.7 Representation of questions

As mentioned above, indexers were asked to cut and paste questions from the online questionnaires. They were thus able to see the preceding and following questions, which were sometimes necessary to understand them. (See Section 5.3.4 for examples of questions that rely on the previous question to be understood.)

The ESS German and Greek questionnaires, but not the Romanian questionnaire, are introduced by a list of topics, and their associated question numbers. The topic list from the source questionnaire for ESS 7 is shown in Figure 10.
Figure 10 Topics page of ESS 7 source questionnaire

Different topic sections are introduced with some explanatory text. For example, the politics section is introduced by *Now we want to ask a few questions about politics and government.*

In the EVS, questions are shown alongside their variables, again with introductory text for each topic. See for example Figure 11:

![Image of EVS 2009 British questionnaire](image)

*Figure 11 Q1, EVS 2009 British questionnaire*

Show cards are also included in the questionnaire.

www.seriss.eu  GA No 654221  24
SHARE questionnaires are in machine readable format. Show cards are not included. Questions are shown with their variables and variable names (in English). Topics are introduced with some introductory text – see Figure 12 from the English source questionnaire.

```
IF (BR IN Test) OR (ALL IN Test)

BLOCK

IF MN01_Longitudinal = 0

Tx_FL_BR001

BR01_EverSmokedDaily
The following questions are about smoking and drinking alcoholic beverages. Have you ever smoked cigarettes, cigars, cigarillos or a pipe daily for a period of at least one year?
EVER SMOKED DAILY
1. Yes
5. No

IF BR01_EverSmokedDaily = 3

Tx_FL_BR002

BR02_StillSmoking
[The following questions are about smoking and drinking alcoholic beverages.]
Do you smoke at the present time?
SMOKE AT THE PRESENT TIME
1. Yes
5. No

Tx_FL_BR003

BR03_HowManyYearsSmoked
```

Figure 12 Two questions from the SHARE 6 generic main questionnaire

4 Results

4.1 Overview

As mentioned above, a total of 62 questions in ESS, and 121 questions in EVS were indexed in each of the three languages: German, Greek and Romanian. A further 51 questions in SHARE were indexed in German and Greek only.

For each survey, and for each indexer, we report:

- the average level of difficulty of assigning ELSST terms to a question, as measured on the 1-5 scale, where 1 is the easiest and 5 the most difficult, as rated by the indexer
- the number of exact, inexact, partial or non-equivalent matches of ELSST terms to the question content, as rated by the indexer

We summarise the indexers’ comments on the indexing task, and list each indexer’s new term suggestions for each survey.

We also look at how consistent indexing was across indexers. We do this by measuring, for each survey:

- the average number of index terms assigned to a question per indexer
• the number of questions where at least one term was assigned by all three indexers to the same question
• the number of questions where at least one term was assigned by at least two indexers

4.2 Difficulty in assigning terms to questions

On average, indexers rated the difficulty of assigning terms to questions as follows, where 1 is easy and 5 is very hard:

<table>
<thead>
<tr>
<th>Language</th>
<th>ESS</th>
<th>EVS</th>
<th>SHARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>2.3</td>
<td>2.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Greek</td>
<td>2.1</td>
<td>1.8</td>
<td>2.3</td>
</tr>
<tr>
<td>Romanian</td>
<td>2.1</td>
<td>2.0</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Figure 13 Average difficulty rating assigned to questions

The German and Greek indexers found SHARE more difficult to index than the other two surveys. Both the Greek and Romanian indexers found the EVS easier to index than the ESS, although the German indexer found the opposite.

Not all indexers found the same questions equally difficult to index. However, in each survey, all three indexers rated the difficulty of indexing some questions between 3-5. They are as follows (the ESS numbering is from ESS 7):

• ESS: B11, B12, B13, B14, B15, B17, B30, B31
• EVS: Q29C, Q56, Q68E
• SHARE: PH061, PH072, PH076, PH071, PH089, PH049, PH050, PH059, PH659, BR016, BR033

Indexers reported various reasons for difficulty in assigning terms. Missing terms were the chief reason (See Section 4.4 below). Another reason reported by the Greek indexer for some ESS questions (e.g. ESS B30 and B31) was the preponderance of verbs (index terms are mainly noun phrases). The Romanian indexer commented that some ESS questions contained too much detail to index properly. For example ESS B29, B30, and C4.

4.3 Match of ELSST terms to question content

The types of match reported by each indexer for each survey are shown in tables 13-15 below.

Most matches in all surveys were rated either exact or partial. Partial matches were generally because the ELSST term was broader than the concept expressed in the question.

Both the German and Greek indexers found more exact matches in EVS than ESS, although the opposite was the case for the Romanian indexer. The highest number of exact matches (69%) was reported for the EVS by the Greek indexer, while the lowest number of exact matches (14%) was reported by the German indexer for SHARE.
ESS:

<table>
<thead>
<tr>
<th>Match type %</th>
<th>German</th>
<th>Greek</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact</td>
<td>40.5%</td>
<td>53%</td>
<td>68%</td>
</tr>
<tr>
<td>inexact</td>
<td>1.5%</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>partial</td>
<td>55%</td>
<td>42%</td>
<td>29%</td>
</tr>
<tr>
<td>non-equivalent</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>don’t know</td>
<td>3%</td>
<td>0%</td>
<td>3%</td>
</tr>
</tbody>
</table>

*Figure 14 Match between content of questions and ELSST terms in ESS questions*

EVS:

<table>
<thead>
<tr>
<th>Match type %</th>
<th>German</th>
<th>Greek</th>
<th>Romanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact</td>
<td>51%</td>
<td>69%</td>
<td>42%</td>
</tr>
<tr>
<td>inexact</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>partial</td>
<td>47%</td>
<td>29%</td>
<td>55%</td>
</tr>
<tr>
<td>non-equivalent</td>
<td>0%</td>
<td>0%</td>
<td>1.5%</td>
</tr>
<tr>
<td>don’t know</td>
<td>2%</td>
<td>2%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

*Figure 15 Match between content of questions and ELSST terms in EVS questions*

SHARE:

<table>
<thead>
<tr>
<th>Match type</th>
<th>German</th>
<th>Greek</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>exact</td>
<td>14%</td>
<td>43%</td>
</tr>
<tr>
<td>inexact</td>
<td>0</td>
<td>4%</td>
</tr>
<tr>
<td>partial</td>
<td>86%</td>
<td>49%</td>
</tr>
<tr>
<td>non-equivalent</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>don’t know</td>
<td>0</td>
<td>4%</td>
</tr>
</tbody>
</table>

*Figure 16 Match between content of questions and ELSST terms in SHARE questions*

4.4 Suggested new terms

As mentioned above, missing terms were the most common reason indexers gave for difficulty in assigning terms to questions. A list of new terms suggested by indexers for each survey is given in Appendices 4-6. Terms marked with a ‘?’ are where only the English equivalent was suggested in that language; terms shown in red are where the term existed in ELSST as a Use For term of an existing Preferred Term. This meant that they could not be used for indexing in their own right. The indexers were thus forced to use the broader Preferred Terms instead.

For all surveys, both the Greek and German indexers reported that there was a particular lack of terms for time phrases (for example ‘nowadays’, for the last 10 years, ‘how many years’, ‘used to..’).

Other areas where the German indexer reported a lack of terms include emotions and specific behaviour/moral aspects (for example ‘fairness’, ‘helpfulness’, ‘selfishness’, ‘exploitation’, ‘sense of..’).
Some of the suggested new terms (e.g. HEALTH STATUS, CHRONIC DISEASES, EATING HABITS, and ACTIVITIES OF DAILY LIVING) have now been added to ELSST.

4.5 Indexing consistency
In the ESS, indexers assigned between 1 and 4 terms to each question; in the EVS, indexers assigned between 1 and 6 terms to each question; and in SHARE, indexers assigned between 1 and 5 terms:

<table>
<thead>
<tr>
<th></th>
<th>ESS</th>
<th>EVS</th>
<th>SHARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>1.8</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Greek</td>
<td>2</td>
<td>2.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Romanian</td>
<td>1.6</td>
<td>1.8</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Figure 17 Average number of ELSST terms assigned to a question

On average, the least number of terms were assigned by the Romanian indexer to the ESS questions, while the most number of terms were assigned by the German indexer to the EVS and SHARE questions.

The number of questions where two of the indexers assigned at least one term was approximately the same for ESS and EVS but markedly lower for SHARE:

<table>
<thead>
<tr>
<th>Survey</th>
<th>percentage of questions where all 3 indexers assign at least 1 term in common</th>
<th>percentage of questions where at least 2 indexers assign at least one term in common</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESS</td>
<td>53%</td>
<td>87%</td>
</tr>
<tr>
<td>EVS</td>
<td>59%</td>
<td>82%</td>
</tr>
<tr>
<td>SHARE</td>
<td>n/a</td>
<td>71%</td>
</tr>
</tbody>
</table>

Figure 18 Percentage of questions assigned at least one term in common by (a) three indexers (b) two indexers

Unsurprisingly, there were fewer cases for EVS and ESS where all three indexers assigned at least one term in common.

4.6 Conclusion
Overall, there were marked differences in how difficult each indexer found the indexing task as a whole, and the individual surveys. Indexers also differed in how often they felt there was an exact match between the ELSST terms assigned and the question content. This figure varied between 14-69%. There was also relatively little consistency between indexers, with only 53-59% of questions sharing an index term in common by all three indexers for the ESS and EVS surveys.

5 Discussion and recommendations

5.1 Overview
In the previous section we reported indexers’ own views on how difficult it was to assign index terms to the questions and how well the index terms assigned cover the subject content of the questions. Here we look in more detail at factors that have a bearing on the above and on the lack of consistency in indexing between indexers. These factors are:
We look at each one in turn and make recommendations. Note that the question numbers for the ESS survey are from ESS7.

5.2 Indexers

As noted above, none of the indexers had experience of indexing, or of using a thesaurus. Lack of in-depth knowledge of the thesaurus meant that indexers did not always find the most specific term, as with the following clusters, where the broadest term is highlighted:

- TELEVISION; TELEVISION VIEWING;
- VOTING; VOTING BEHAVIOUR; VOTING INTENTIONS
- LANGUAGES; LANGUAGES USED AT HOME

In the review stage, when alternative terms were suggested to them, they were generally accepted.

**Recommendation:** To promote the retrieval of appropriate terms, it is important that Scope notes, Use notes and Related Terms are provided for ELSST terms.

Unlike the German indexer, the Greek and Romanian indexer did have experience of working with social science data. This possibly explains why, on average, the Greek and Romanian indexer found it easier to assign index terms than the German indexer.

**Recommendation:** It is recommended that indexers have a background in data and the social sciences.

Indexers also varied in how they interpreted the indexing rules. While most of the time the indexers focused on the literal content of a question, in some instances they chose to index the construct behind the question. This corresponds to the ‘aboutness’ of the question (see Friedrich and Siegers 2016).

Examples include:

**ESS B30:**

English original: *How about people of a different race or ethnic group from most [British people/people in the UK]?*

Greek translation: *Ποια η άποψή σας για τα άτομα από διαφορετική φυλή ή εθνική ομάδα από τα περισσότερα άτομα της χώρας μας?

English gloss of Greek translation: *What do you think about people from a different race or ethnic group than most of our country?*

(This question follows a question on what respondents think about people of the same race or ethnic groups as the majority of Greeks coming to live in Greece.) The Greek indexer commented that the question is clearly about RACIAL DISCRIMINATION, although it does not explicitly contain this term.)
ESS B27:

English original: *Using this card, please say to what extent you agree or disagree with each of the following statements...* Gay men and lesbians should be free to live their own life as they wish.

German translation: *Sagen Sie mir, wie sehr Sie jeder der folgenden Aussagen zustimmen oder wie sehr Sie diese ablehnen.... Schwule und Lesben sollten ihr Leben so führen dürfen, wie sie es wollen.*

English gloss of German translation: *Tell me how much you agree with each of the following statements or how much you reject them.... Gays and lesbians should be allowed to lead their lives as they wish.*

The German indexer assigned the term MORAL CONCEPTS, on account of the fact that its Use For term is MORAL VALUES in German, because she felt that the question “steers towards moral values of the interviewee.”

Indexers also differed in the degree of specificity they chose to index at. This was noticeable, for example, in the SHARE questions about eating habits (BR026- BR029). While the German indexer looked to find terms for individual foodstuffs (legumes, fish, meat, etc.), the Greek indexer chose to assign the more generic term FOOD AND NUTRITION (although he would have preferred to have its Use For term EATING HABITS as an option) and commented that he felt this level of indexing would be sufficient.

**Recommendation**: To promote consistency in indexing, clearer indexing guidelines need to be drawn up on how to index questions. This will require further research into how researchers wish to use and look for questions, in particular, if they wish to search for a question by its underlying construct/scale as well as or instead of its literal content. Note also that the level of specificity indexers are able to index at is dictated by the thesaurus itself. It is up to the ELSST management committee to decide how specific the thesaurus should be, based on the indexing requirements of its members.

5.3 Questions

We noted in the previous section that the Greek indexer found some questions difficult to index if they contained a large amount of verbs (thesaurus terms are usually expressed as noun phrases). Other characteristics of questions that contributed to difficulty in assigning index terms to them, and/or accounted for differences in the terms assigned by different indexers, include the following:

- ambiguity in either the source or target questions
- difference in the wording of question text across languages
- context dependency of questions

5.3.1 Ambiguity

On the whole, questions were clear and unambiguous. However, occasionally, some ambiguity was found. An example is Q42E from the German version of the EVS survey:
English original: *Here is a list of things which some people think make for a successful marriage. Please tell me, for each one, whether you think it is very important, rather important or not very important for a successful marriage?... Good housing*

German translation: *Hier auf dieser Liste steht Verschiedenes, was manche für eine gute Ehe für wichtig halten. Unabhängig davon, ob Sie verheiratet sind oder nicht: Sagen Sie mir bitte für jeden Punkt, ob Sie meinen, dass dieser für eine gute Ehe sehr wichtig, ziemlich wichtig oder nicht besonders wichtig ist? Gute Wohnverhältnisse*

The problem is that ‘Wohnverhältnisse’ is ambiguous in German, since it may refer to living conditions in general not just the (physical) condition of housing.

5.3.2 Different wording in question text across languages

There was a tendency for indexers to choose terms that reflected the wording of the question. Thus where the wording differed between languages, this led to indexers choosing different index terms. Some examples are the following:

**ESS A2:**

English original: *And again on an average weekday, how much of your time watching television is spent watching news or programmes about politics and current affairs?*

Romanian translation: *Într-o zi lucrătoare a săptămanii (de luni până vineri) cât timp vă petreceţi urmărend la TV știri sau programe pe teme politice și problemele pe care le rezolvă guvernul?*

The Romanian translation refers to “programmes on political issues and issues that the government is resolving” which led the Romanian translator to index it with the indexing term PROBLEM SOLVING. The German and Greek translations were closer to the English source, however, so neither assigned this term.

**ESS B24:**

English original: *Using this card what do you think overall about the state of education in Britain/the UK nowadays?*

German translation: *Sagen Sie mir bitte, wie Sie - alles in allem - den derzeitigen Zustand des Bildungssystems in Deutschland einschätzen.*

Greek translation: *χρησιμοποιώντας αυτή την κάρτα, μπορείτε να μας πείτε τι πιστεύετε γενικά για την παρούσα κατάσταση της παιδείας στην Ελλάδα;*

Romanian translation: *Acum, utilizând o scală de la 0 la 10, unde 0 înseamnă foarte proastă și 10 foarte bună, cum evaluați situația sistemului de învățământ din România?*

Both the German and Romanian translation refer to ‘educational system’, which prompted the indexers to assign the ELSST term EDUCATIONAL SYSTEMS. However, the Greek translation is closer to the English source and refers to ‘the state of education’, so the Greek indexer chose the term EDUCATION instead.
Not all differences in wording, however, resulted in different indexing choices.

**5.3.4 Context dependency**

Different degrees of context dependency occur in questions.

Some questions rely on the previous question or questions to be understood. Examples include the following:

ESS B30:

   English original: *How about people of a different race or ethnic group from most [British people/people in the UK]?*

The English original cannot be understood without its previous question:

ESS B29:

   English original: *Now, using this card, to what extent do you think [Britain / the UK] should allow people of the same race or ethnic group as most of [Britain / the UK]’s people to come and live here?*

This is also the case for the German and Greek translations, but not for Romanian, where B30 makes sense on its own:

   Romanian translation: În ce măsură România ar trebui să permită ca persoane de rasa sau etnie diferită sa vina și să locuiască aici?

   English gloss of Romanian translation: *To what extent should Romania allow people of different race or ethnicity to come and reside here?*

Since indexers were asked to consider the previous question when indexing, this type of context-dependency did not affect indexing, but it could pose a problem for automatic indexing, for example.

Context dependency also includes the wider context of a question, e.g. the topic to which the question belongs.

**Recommendation:** The above examples show that a question’s context (previous questions, topic headings, etc.) needs to be taken into account when indexing questions. If indexing questions is to include the construct associated with them, fuller documentation (e.g. variable descriptions) will also be needed.

**5.4 ELSST terms**

In the previous section we noted that indexers attributed much of the difficulty of indexing to missing terms. Note that what counts as ‘missing’ is subjective, and depends on the level of specificity the indexer chooses to index at.

We discussed how some of the ‘missing’ terms existed in ELSST as Use For terms, not Preferred Terms. In these cases, the indexer was forced to use the Preferred Term instead, which was broader in meaning.
For example, the Greek indexer assigned FOOD AND NUTRITION to SHARE questions BR026- BR029, since its Use For term EATING HABITS was not available to assign instead.

Lists of new term suggestions for ELSST are given in Appendices 4-6. Some of these terms describe emotions, which is one of the classes of ‘directive terms’ discussed by Friedrich and Siegers (2016).

Other problems relating to ELSST terms include:

- translation issues
- missing UFs in the source or target terms
- lack of distinction between closely related terms

5.4.1 Translation issues

Some differences in assigning terms could be traced to differences in the term’s translations. Examples include the following:

ESS B18A:

English original: Is there a particular political party you feel closer to than all the other parties?

POLITICAL ALLEGIANCE was chosen by German and Romanian indexers only. The Greek indexer chose the term PARTY IDENTIFICATION instead, since the Greek translation of the ELSST term POLITICAL ALLEGIANCE (ΠΟΛΙΤΙΚΗ ΥΠΑΚΟΗ) is used when MPs of a party follow the policy guidelines of the party.

EVS Q5aA:

English original: Please look carefully at the following list of voluntary organisations and activities and say ...a) which, if any, do you belong to?

... Social welfare services for elderly,

The Romanian rejected the ELSST term MEMBERSHIP (AFILIERE) as an indexing option, since it is used for political membership only. She suggested the term APARTENENTA instead. The English term MEMBERSHIP in ELSST has no such restriction.

EVS Q5aD:

English original: Please look carefully at the following list of voluntary organisations and activities and say ...a) which, if any, do you belong to?

... Trade unions

The Greek indexer chose the index term ΣΥΝΔΙΚΑΛΙΣΤΙΚΕΣ ΟΡΓΑΝΩΣΕΙΣ, which is the translation of the ELSST term EMPLOYERS’ ORGANIZATIONS instead of ΜΕΛΟΣ ΕΠΑΓΓΕΛΜΑΤΙΚΗΣ ΟΡΓΑΝΩΣΗΣ, the ELSST translation of TRADE UNION MEMBERSHIP, since he found it a better translation for TRADE UNION MEMBERSHIP in Greek.

5.4.2 Missing UFs in the source or target terms
Where a Use For term that represents a narrower concept to its Preferred Term is present in one language and not the other, this may cause inconsistency in indexing. For example, WORKMATES is a UF of WORKPLACE RELATIONS in English, and has been translated into German and Romanian, but not Greek.

Conversely, LEBENZUFRIEDENHEIT (SATISFACTION WITH LIFE) is a Use For term for ZUFRIDENHEIT (SATISFACTION) in German, but has no equivalent in the source language. This meant that the term ZUFRIDENHEIT was felt to be a better match in German for questions about satisfaction with life than in the other languages.

**Recommendation:** Non-synonymous Use For terms should never be added to the target language without also being added to the source language. And all non-synonymous Use For terms in the source language should be translated.

### 5.4.3 Closely related terms

The existence of many closely related and partially overlapping terms accounted for much of the lack of consistency between indexers. Examples include the following:

- **Example 1:** POLITICAL INFLUENCE/ POLITICAL PARTICIPATION/ POLITICAL ACTION (and SOCIAL PROTEST); LOBBYING (see ESS B12; EVS Q55)
- **Example 2:** VOTING/VOTING BEHAVIOUR/VOTING INTENTION (see ESS B9)
- **Example 3:** RELIGION, RELIGIOUS AFFILIATION RELIGIOUS BELIEFS (see ESS C9-11)
- **Example 4:** RELIGIOUS GROUPS; RELIGIOUS COMMUNITIES; RELIGIOUS AFFILIATIONS; RELIGIOUS SOCIETIES; RELIGIOUS INSTITUTIONS (see EVS Q5aB)
- **Example 5:** JUSTICE, ADMINISTRATION OF JUSTICE, LEGAL SYSTEMS (see ESS B3)
- **Example 6:** MENTALLY ILL, MENTAL DISORDERS (see EVS Q6G)
- **Example 7:** PARTENTAL ROLE, PARENTAL RESPONSIBILITY (see EVS Q43, 50, 51)
- **Example 8:** DISEASES, ILL HEALTH (see SHARE PH061)
- **Example 9:** DRINKING BEHAVIOUR; ALCOHOL USE (see SHARE BR039)
- **Example 10:** HEALTH, WELL-BEING (HEALTH) (see SHARE PH003)

**Recommendation:** In order to guide indexers to using the most appropriate term in a cluster, Use Notes should be added to the terms.

### 5.5 Conclusion

We have shown how different factors influence how difficult indexers find it to index questions, and affect their assessment of how well ELSST terms cover the subject content of the questions. These factors include the properties of the question text, and/or the availability or nature of the index terms.
as well as the skills and knowledge of the indexers themselves. We also made recommendations where appropriate. However, just how survey questions should be indexed remains an open question.

We noted that the indexing experiments also shed light on some quality issues relating to the thesaurus.

In the next section we compare the results of the indexing experiment with the results of indexing some of the same studies at study-level.

6 Comparison of indexing with ELSST at study- versus question-level for SERISS cross-national studies

6.1 Overview
In this part, we compare results of indexing ESS and EVS surveys at whole study level with indexing at question level.

6.2 Study-level indexing
Balkan (2017) conducted a comparison of how ESS and EVS had been indexed with ELSST (or HASSET) at study level by various CESSDA archives. We reported:

- the number of terms assigned by each archive
- the degree of overlap of terms assigned by each archive

Two rounds of ESS, round 1 and round 5, were indexed by EKKE, Greece and SND, Sweden, in Greek and Swedish respectively. Results showed that, for Round 1, EKKE assigned 7 terms, while SND assigned 16. Unsurprisingly the EKKE terms were more general than the SND terms (e.g. POLITICS versus POLITICAL INTEREST and POLITICAL PARTICIPATION). For ESS Round 5, while the number of terms assigned were more similar (EKKE assigns 13 terms, while SND assigns 11), there was less thematic overlap.

We also compared the indexing terms assigned by the UK Data Archive to the EVS Wave 4 Great Britain, 2009-2010 with the terms assigned to the European Values Study 2009: Finnish Data by the FSD. (The data were language-specific and thus not directly comparable.) Here the number of terms assigned differed markedly: the UK Data Archive assigned 119 HASSET terms, 111 of which are in ELSST, while FSD assigned 21. As expected, the UK Data Archive index terms tended to be more specific. For example, rather than ETHICS, or VALUES, the UK Data Archive assigns the terms MORAL BEHAVIOUR, MORAL CONCEPTS and MORAL VALUES, and instead of FAMILIES, the UK Data Archive assigns FAMILY LIFE and FAMILY ROLES.

Differences in indexing results were attributed to different indexing styles and, in the case of the EVS, differences in the data indexed. Unlike the present study, there was not enough evidence to draw any conclusion about differences being due to the quality of the ELSST source terms or their translations.

6.3 Discussion of results
Apart from the fact that the indexing described in the current report is at question level, while Balkan (2017) describe study-level indexing, other important differences between the two types of indexing include:

www.seriss.eu  GA No 654221  35
the data indexed:
  o the current report compared indexing of ESS Round 7, while Balkan (2017) compared indexing of ESS Rounds 1 and 5. In both reports, however, the same wave of EVS was indexed (wave 4 (2008)).
  o the number of questions indexed in each round/wave also differed: the current report compared only a subset of questions from the ESS and the EVS surveys, while Balkan (2017) compared the entire round.
  o the language of questionnaires: in the current study, questionnaires in 3 languages (German, Greek and Romanian) were indexed, while in Balkan (2017) Greek and Swedish questionnaires were indexed.

ELSST terms:
  o in the current study, ELSST terms were assigned in 3 languages (German, Greek and Romanian), while in Balkan (2017) Greek and Swedish index terms were used. The choice of languages in the current study was motivated by the availability of personnel to undertake the indexing work, as explained in Section 3.6.1, but is not significant as far as the comparison between this study and the work reported in Balkan (2017) is concerned.
  o The index terms in Balkan (2017) were also from an earlier version of ELSST.

the indexers:
  o the indexing in the present study was carried out by non-expert indexers, while that in Balkan (2017) was performed by trained indexers.
  o indexing guidelines: in the current report, each indexer followed the same indexing instructions, while each archive in Balkan (2017) indexed according to their own indexing guidelines.

Despite these differences, indexing at question or study level share some characteristics.

First, just as indexing at study level requires access to a study’s documentation, so too does question-level indexing, if it goes beyond the literal text of the question.

Secondly, study-level indexing may be quite general, or as specific as variable or question level. Indexing studies to question level, as carried out by the UK Data Archive for example, may mean assigning approximately as many index terms as that required to index individual questions.

Thirdly, indexing whole studies or individual questions requires skill on the part of the indexer, and knowledge not only about the data, and the subject area, but also of thesauri and indexing. This has a bearing on the question of where in the data lifecycle it is best to index questions. Although it might appear that the questionnaire creator would be the best choice, at the data creation stage, since they have intimate knowledge of the data, it is the data cataloguer who possesses the requisite indexing skills and in-depth knowledge of the thesaurus. Ideally, however, the data creator and data cataloguer could collaborate on the indexing process. For example, if the data creator employed the same controlled vocabulary to mark up questions during the survey creation process, these keywords could be used by the data cataloguer as the basis for question indexing. Any queries regarding keywords could then be discussed by both.
One major advantage of question-level indexing over study-level indexing is the level of precision it allows in searching. Increasingly, users are demanding easier ways to access data at question and variable levels.

As Friedrich and Kempf (2014) suggest, however, indexing questions should be seen as complementary to indexing whole studies, since they cater for the needs of different user types. A cascade model of indexing offers a promising approach for combining the different levels of indexing.

As far as the thesaurus developer is concerned, question-level indexing offers an advantage over study-level information in that it provides useful feedback on how thesaurus terms have been interpreted and translated across languages, and identifies gaps in coverage, thus contributing to quality control.

7 Conclusion and future steps

This report set out to answer the question of where in the data lifecycle it is best to index data. We tackled this question by looking specifically at how and where to index questions. We described an experiment to index questions from the ESS, EVS, and SHARE surveys, using native speakers of German, Greek and Romanian to index their respective language version of the studies using index terms from the ELSST thesaurus. We then reported and analysed the results to establish where the difficulties of the task lie, and reasons for inexact match between ELSST terms and the content of the questions, and for the lack of consistency between indexers. Differences were found to be due to various factors, including differences in the way the questions had been translated, and in how the terms had been interpreted and/or translated, as well as to differences associated with the indexers themselves.

Based on the results, we made some recommendations, but concluded that the best method of how questions should be indexed remains an open question. Ultimately it should be driven by user needs. More research is required on this and on how to retrieve questions in search engines. Future collaboration with projects such as CLOSER and EQB would be beneficial.

We then compared the results of the question indexing task to the previously reported work on comparing the index terms that had been assigned to the same surveys at study level. We saw that both study-level and question-level indexing require excellent knowledge of the data and documentation as well as knowledge of indexing techniques and the thesaurus used for indexing.

We concluded that while data creators know their data better than the data cataloguer, the data cataloguer has better indexing skills and knowledge of the thesaurus. We recommended that question indexing could benefit from greater collaboration between both parties. This collaboration could include, for example, data creators and data cataloguers using the same controlled language (potentially ELSST) for assigning keywords. The feasibility of such an approach could be explored in a follow-up project. The relationship between coding and indexing could also be further explored.

We also observed that since study-level indexing at a detailed level duplicates the work involved in assigning terms to questions, the two processes could potentially be combined, as in the case of

www.seriss.eu
GA No 654221
‘cascade’ indexing model described by Friedrich and Kempf (2014). The benefits of providing both types of indexing include greater flexibility for stakeholders in finding data at any level.

Finally, we also found that the question-level indexing provided a useful quality measure of ELSST terms and their translations.

References
Aila data catalogue, FSD. Available at https://services.fsd.uta.fi/catalogue/search/advanced?lang=en&query_lang=en&query_document=variable


CLOSER project. Available at https://wiki.ucl.ac.uk/display/CLOS/CLOSER

CLOSER Discovery, CLOSER project. Available at http://discovery.closer.ac.uk/

CLOSER Scales, CLOSER project. Available at https://wiki.ucl.ac.uk/display/CLOS/Scales

CLOSER Topics, CLOSER project. Available at https://wiki.ucl.ac.uk/display/CLOS/Topics

Data Documentation Initiative (DDI): Available at https://www.ddialliance.org/


Engel, R.J. and Schutt, R.K (2014): Fundamentals of social work research, Sage publications, University of Pittsburgh, USA. (Chapter 4: conceptualization and measurement is available online at: https://us.sagepub.com/sites/default/files/upm-binaries/61666_Chapter_4.pdf)

Euro Question Bank (EQB) project. Available at: https://www.cessda.eu/Projects/Work-Plans/Work-Plan-2017#eqb

www.seriss.eu


UK Data Archive (2017): Data collection indexing procedures, UK Data Archive, October 2017, UK Data Archive internal publication, University of Essex, Colchester.

UK Data Service: Catalogue metadata, website. Available at https://www.ukdataservice.ac.uk/manage-data/document/metadata

UK Data Service: Prepare and manage data, website. Available at https://www.ukdataservice.ac.uk/manage-data

UK Data Service: Research data lifecycle, website. Available at https://www.ukdataservice.ac.uk/manage-data/lifecycle

UK Data Service: ReShare service. Available at https://reshare.ukdataservice.ac.uk/

UK Data Service: Variable and question bank. Available at https://discover.ukdataservice.ac.uk/variables


Appendix 1 Indexing instructions (German example)

Task description

Aim: the aim of the task is to index selected questions from 3 social science surveys:

- European Social Survey (ESS)
- Survey of Health, Ageing and Retirement in Europe (SHARE)
- European Values Study (EVS), using ELSST terms.

You will be given initial instructions on how to carry out the indexing task.

At the end of the indexing process, you will be required to write a short report to include an evaluation of:

- how well the ELSST terms capture the semantic content of the questions you indexed
- the difficulty of indexing the questions

Method

For each study:

1. Go to the online version of the questionnaire (see links below):
2. Read the relevant documentation

For each highlighted question in the paper version of the questionnaire:

2. Copy the question ID, variable name (if any), and question text into the Excel template for the relevant survey
3. Select ELSST index terms for the question and enter them into the spreadsheet. For example:

   Question text:  Abgesehen von Gottesdiensten, wie oft beten Sie – wenn überhaupt?
   ELSST term:  RELIGIONSAUSÜBUNG

4. Rate the degree of equivalence between the meaning expressed by the questions, and the meaning expressed by your selected ELSST term(s). Choose one of the following:
   a.  **Exact** equivalence: there is no difference in the meaning expressed by the question and by the ELSST term(s)
   b.  **Inexact** equivalence: there is a small difference in the meaning expressed by the question and by the ELSST term(s)
c. **Partial** equivalence: there is a partial overlap in the meaning expressed by the question and by the ELSST term(s) because either:
   v. the meaning expressed by the question is **narrower** than that expressed by the ELSST term(s)
   vi. the meaning expressed by the question is **broader** than the content of the ELSST term(s)
   vii. the meaning expressed by the question **partially overlaps** with the meaning expressed by the ELSST term(s)
   viii. **Non-equivalence**: there is no overlap in the meaning expressed by the question and that expressed by the ELSST term(s)

d. **Don’t know**

5. For all answers above, except ‘Exact equivalence’, explain why
6. Suggest any new ELSST terms that might be required to index the question exactly, e.g. BETEN in the example above
7. Rate how easy it was to index the question on a scale 1-5, where 1 is very easy and 5 is very difficult
8. If you rate the difficulty of indexing the question 3 or above, explain why

**Links to survey questionnaires**

**ESS:**

- go to [http://www.europeansocialsurvey.org/data/country_index.html](http://www.europeansocialsurvey.org/data/country_index.html)
- select ‘Germany’
- click on [ESS7_main_and_supplementary_questionnaire_DE.pdf](http://www.europeansocialsurvey.org/data/country_index.html)

**SHARE:**

- Go to [http://www.share-project.org/data-documentation/questionnaires.html](http://www.share-project.org/data-documentation/questionnaires.html)
- Click on Wave 6, then select ‘Germany’ ‘CAPI main questionnaire’

**EVS:**

Go to [https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4800&db=e&doi=10.4232/1.12458](https://dbk.gesis.org/dbksearch/sdesc2.asp?no=4800&db=e&doi=10.4232/1.12458)

- Click on ‘Data & Documents’
- Click on ‘Questionnaires’
- Click on ZA4753_q_de.pdf Germany (Questionnaire) 570
Appendix 2 Indexing guide

The following rules of indexing should be followed.

1. Use only Preferred Terms for indexing
2. Specificity

Choose the most specific terms available when assigning terms to questions. Thus, a question about goats should be indexed with the preferred term "GOATS" rather than with the broader terms "LIVESTOCK" and "DOMESTIC ANIMALS".

As a general rule, a question should not be indexed with both broader and narrower terms from the same hierarchy. For example, a question about LIVESTOCK should not be indexed with the broader term DOMESTIC ANIMALS, or the narrower term GOATS.

If you cannot find a term specific enough, use a more general term instead.

3. Exhaustivity

You may need several terms to fully express the content of a question. For example, to index a question about children’s health, you might need two terms: CHILDREN and HEALTH.
## Appendix 3 EVS template headings

<table>
<thead>
<tr>
<th>Question ID</th>
<th>Variable name</th>
<th>Question text</th>
<th>Selected ELSST terms</th>
<th>Degree of equivalence between question and selected ELSST term(s): exact, inexact, partial, non-equivalence, don't know</th>
<th>Reason for no exact equivalence</th>
<th>New term suggestions</th>
<th>Level of difficulty of assigning ELSST term(s) to question (1-5, where 1 is very easy, and 5 is very difficult)</th>
</tr>
</thead>
</table>
## Appendix 4 New term suggestions for ELSST from ESS

<table>
<thead>
<tr>
<th>German term</th>
<th>Greek term</th>
<th>Romanian term</th>
<th>English equivalent</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>? ZI LUCRATOARE</td>
<td>WORKING DAYS</td>
<td>A1-A2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>? FAIRNESS</td>
<td>A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? EXPLOITATION</td>
<td>A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? HELPFULNESS</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? EGOISM</td>
<td>SELFISHNESS</td>
<td>A5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>? INTERES PROPRIU</td>
<td>SELF INTEREST</td>
<td>A5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertrauen in die 'Exekutive'</td>
<td>TRUST IN THE EXECUTIVE</td>
<td>B4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>? TIME FRAME</td>
<td>B11-17; C5, C11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΕΞΕΛΙΞΗ (ΟΙΚΟΝΟΜΙΑ-ΚΟΙΝΩΝΙΑ)</td>
<td>DEVELOPMENT (SOCIO-ECONOMIC)</td>
<td>B11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A intra in contact</td>
<td>CONTACTING SOMEONE</td>
<td>B11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTIUNE CIVICA</td>
<td>CIVIC ACTION</td>
<td>B12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTIVITATE</td>
<td>BEING ACTIVE IN ORGANISATION/ASSOCIATION</td>
<td>B13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PETITIE</td>
<td>PETITIONS</td>
<td>B15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>? BOYCOTT</td>
<td>B17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? EXERCISE OF OFFICE</td>
<td>B22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORIENTARE POLITICA</td>
<td>POLITICAL ORIENTATION</td>
<td>B19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>? ETHNIC MAJORITY</td>
<td>B29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? LIMIT(ATION) QUANTITY</td>
<td>B29-31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? EUROPEAN DEVELOPING/POOR COUNTRIES</td>
<td>B30A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? NON-EUROPEAN DEVELOPING/POOR COUNTRIES</td>
<td>B31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? IMPACT</td>
<td>B32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? INFLUENCE</td>
<td>B32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>? BALANCE/IMBALANCE</td>
<td>B32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>ENRICHMENT</td>
<td>B33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>POSITIVE/NEGATIVE IMPACT</td>
<td>B33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>SENSE OF SECURITY/PROTECTION</td>
<td>C6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>TIMES OF DAY/DARKNESS</td>
<td>C6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΚΑΤΑΣΤΑΣΗ ΥΓΕΙΑΣ</td>
<td>STARE DE SĂNĂTATE</td>
<td>HEALTH STATUS</td>
<td>C7</td>
<td></td>
</tr>
<tr>
<td>Rugăciune</td>
<td>PRAYER</td>
<td>C15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix 5 New term suggestions for ELSST from EVS

<table>
<thead>
<tr>
<th>Key</th>
<th>Original language suggestion not available</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TERM SUGGESTION in red</strong></td>
<td>Exists in ELSST as Use For term</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th>Term Suggestion in red</th>
<th>English equivalent</th>
<th>Variable number</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>German term</td>
<td><strong>APARTEN ENTA</strong></td>
<td><strong>MEMBERSHIP</strong></td>
<td>V10</td>
<td></td>
</tr>
<tr>
<td>Greek term</td>
<td><strong>πολιτικές παρατάξεις</strong></td>
<td><strong>POLITICAL GROUPS</strong></td>
<td>V14</td>
<td></td>
</tr>
<tr>
<td>Romanian term</td>
<td><strong>ανθρώπινα δικαιώματα</strong></td>
<td><strong>HUMAN RIGHTS</strong></td>
<td>V16</td>
<td>instead of current translation: ΔΙΚΑΙΩΜΑΤΑ ΤΟΥ ΑΝΘΡΩΠΟΥ</td>
</tr>
<tr>
<td><strong>VORSTRAFEN</strong></td>
<td><strong>PREVIOUS CONVICTION S</strong></td>
<td>V46</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>?</strong></td>
<td><strong>LIKE/DISLIKE</strong></td>
<td>V47</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>?</strong></td>
<td><strong>Άτομα άλλης φυλής</strong></td>
<td><strong>PEOPLE OF ANOTHER RACE</strong></td>
<td>V47</td>
<td></td>
</tr>
<tr>
<td><strong>RASĂ</strong></td>
<td><strong>RACE</strong></td>
<td>V47</td>
<td>It's a UF of ETHNIC GROUPS in English, not Romanian</td>
<td></td>
</tr>
<tr>
<td><strong>Λεβεντιάστες</strong></td>
<td><strong>LEFTWING EXTREMISTS</strong></td>
<td>V48</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LINKSEXTREMISMUS</strong></td>
<td><strong>EXTREMIST DE STÂNGA</strong></td>
<td><strong>LEFTWING EXPREMISM</strong></td>
<td>V48</td>
<td></td>
</tr>
<tr>
<td>Greek Term</td>
<td>English Term</td>
<td>Page</td>
<td>Comments</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------</td>
<td>-------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Ακροδεξιός</td>
<td>RIGHTWING EXTREMISTS</td>
<td>V50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Πολύτεκνοι</td>
<td>MANY CHILDREN</td>
<td>V51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ψυχοπαθείς</td>
<td>PSYCHOPATHS</td>
<td>V52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GUTE BEZAHLUNG</td>
<td>GOOD PAY</td>
<td>V69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ευχάριστοι συνεργάτες</td>
<td>PLEASANT WORK COLLEAGUES</td>
<td>V70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARBEITSKOLLEGEN</td>
<td>COLEGI COLLEAGUES</td>
<td>V70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRESS</td>
<td>STRESS</td>
<td>71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Πρωτοβουλία</td>
<td>INITIATIVE</td>
<td>74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Χρήσημη</td>
<td>USEFUL</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NUTZEN FÜR DIE ALLGEMEINHEIT</td>
<td>USE TO THE PUBLIC</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GROßZÜGIGKEIT</td>
<td>GENEROSITY</td>
<td>76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BERUFLICHE VERANTWORTUNG</td>
<td>RESPONSIBILITY AT WORK</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Σημαντικά</td>
<td>IMPORTANT</td>
<td>V79</td>
<td>etc</td>
<td></td>
</tr>
<tr>
<td>Ενδιαφέρον</td>
<td>INTERESTING</td>
<td>V80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERESSANTE BERUFLICHE TÄT</td>
<td>INTERESTING WORK</td>
<td>V80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TÄGIGKEIT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANWEISUNGEN DES VORGESETZT</td>
<td>FOLLOW SUPERIOR'S INSTRUCTIONS</td>
<td>V101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENDE BEOFOLGEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOTTESDIENST</td>
<td>CHURCH SERVICE</td>
<td>V109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EINZELNE PERSONEN</td>
<td>INDIVIDUALS</td>
<td>V115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEUTSCHE WORTER</td>
<td>GRIECHISCH/HEBRÄISCH/LATINISCH</td>
<td>DEUTSCHE ÜBERSETZUNG</td>
<td>KÖRPERICHER/GEISTIGER/EMOTIONALER BEDÜRFNISSE</td>
<td>V117</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------</td>
<td>----------------------</td>
<td>-----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>GEISTIGE UND SPIRITUELLE BEDÜRFNISSE</td>
<td>ΠΝΕΥΜΑΤΙΚΕΣ ΑΝΑΓΚΕΣ</td>
<td>SPIRITUAL NEEDS</td>
<td>GOTT/GOTTESGLAUBE/GOTT</td>
<td>V119</td>
</tr>
<tr>
<td>Glaube an Gott</td>
<td>DUMNEZ EU</td>
<td>GOD</td>
<td>GOTT</td>
<td>V119</td>
</tr>
<tr>
<td>Leben nach dem Tod</td>
<td>Μετά θάνατον ζωή</td>
<td>LIFE AFTER DEATH</td>
<td>VIATĂ DUPĂ MOARTE</td>
<td>V120</td>
</tr>
<tr>
<td>Höhle</td>
<td>Κόλαση</td>
<td>HELL</td>
<td>IAD</td>
<td>V121</td>
</tr>
<tr>
<td>Himmel</td>
<td>Παράδεισος</td>
<td>HEAVEN</td>
<td>RAI</td>
<td>V122</td>
</tr>
<tr>
<td>Sünde</td>
<td>Αμαρτία</td>
<td>SIN</td>
<td>PĂCAT</td>
<td>V123</td>
</tr>
<tr>
<td>Reinkarnation, Wiedergeburt</td>
<td>Μετενσάρκωση</td>
<td>REINCARNATION</td>
<td>REÎNCAR NARE</td>
<td>V124</td>
</tr>
<tr>
<td>Bedeutung/Bedeutsamkeit</td>
<td>%ΕΤΕΝΣΑΡΚΩΣΗ</td>
<td>IMPORTANCE</td>
<td>SPRIJIN</td>
<td>V129</td>
</tr>
<tr>
<td>Kraft</td>
<td></td>
<td>STRENGTH</td>
<td></td>
<td>V130</td>
</tr>
<tr>
<td>Trost</td>
<td></td>
<td>COMFORT</td>
<td></td>
<td>V130</td>
</tr>
<tr>
<td>Gebet</td>
<td>Προσευχή</td>
<td>PRAYER</td>
<td></td>
<td>V131</td>
</tr>
<tr>
<td>Wichtigkeit</td>
<td></td>
<td>IMPORTANCE</td>
<td></td>
<td>V136</td>
</tr>
<tr>
<td>Treue</td>
<td>Πίστη στο/στη σύζυγο</td>
<td>FIDELITY</td>
<td></td>
<td>V136</td>
</tr>
<tr>
<td>Angemessene Bezahlung</td>
<td></td>
<td>APPROPRIATE SALARY</td>
<td></td>
<td>V137</td>
</tr>
<tr>
<td>Übereinstimmung</td>
<td></td>
<td>AGREEMENT</td>
<td></td>
<td>V141</td>
</tr>
<tr>
<td>Schwiegereltern</td>
<td>Πεθερικά</td>
<td>PARENTS-IN-LAW</td>
<td></td>
<td>V142</td>
</tr>
<tr>
<td>Sexuelle Beziehung</td>
<td></td>
<td>SEXUAL RELATIONSHIP</td>
<td></td>
<td>V143</td>
</tr>
<tr>
<td>Viata Sexual</td>
<td></td>
<td>SEX LIFE</td>
<td></td>
<td>V143</td>
</tr>
<tr>
<td>Teilen</td>
<td></td>
<td>SHARING</td>
<td></td>
<td>V144</td>
</tr>
<tr>
<td>Unterstützung</td>
<td></td>
<td>SUPPORT</td>
<td></td>
<td>V144</td>
</tr>
<tr>
<td>Category</td>
<td>German Phrase</td>
<td>English Translation</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>---------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>LEBENZUFRIEDENHEIT</td>
<td>Ολοκληρωμένος</td>
<td>FULFILLED</td>
<td>V149</td>
<td></td>
</tr>
<tr>
<td>ÜBERHOLTE/MODERNE EINRICHTUNG</td>
<td>Ξεπερασμένος</td>
<td>OUTDATED</td>
<td>V150</td>
<td></td>
</tr>
<tr>
<td>GUTE MANIEREN</td>
<td>Καλοί τρόποι</td>
<td>GOOD MANNERS</td>
<td>V170</td>
<td></td>
</tr>
<tr>
<td>UNABHÄNGIGKEIT/SELBSTÄNDIGKEIT</td>
<td>Περιπολή δουλειά</td>
<td>HARD WORK</td>
<td>V172</td>
<td></td>
</tr>
<tr>
<td>VORSTELLUNGSKRAFT/PHANTASIE</td>
<td>Φαντασία</td>
<td>IMAGINATION</td>
<td>V174</td>
<td></td>
</tr>
<tr>
<td>RESPEKT/ACHTUNG</td>
<td>Σεβασμός</td>
<td>RESPECT</td>
<td>V175</td>
<td></td>
</tr>
<tr>
<td>SPARSAMKEIT</td>
<td>Αποταμίευση</td>
<td>THRIFT</td>
<td>V176</td>
<td></td>
</tr>
<tr>
<td>ENTSCHLOSSENHEIT</td>
<td>Αποφασιστικότητα</td>
<td>DETERMINATION</td>
<td>V177</td>
<td></td>
</tr>
<tr>
<td>AUSDAUER</td>
<td>Επιμονή</td>
<td>PERSEVERANCE</td>
<td>V177</td>
<td></td>
</tr>
<tr>
<td>SELBSTLOSIGKEIT</td>
<td>Αλτρυσμός</td>
<td>SELFLESSNESS</td>
<td>V179</td>
<td></td>
</tr>
<tr>
<td>WOHLWOLLEN</td>
<td>Εγνωστής</td>
<td>SELFISH</td>
<td>V179</td>
<td></td>
</tr>
<tr>
<td>UNTERSCHRIFTENSAMMLUNG/UNTERSCHRIFTENAKTION</td>
<td>γραπτή διαμαρτυρία</td>
<td>PETITION</td>
<td>V187</td>
<td></td>
</tr>
<tr>
<td>BOYKOTT/BOYKOTTIERUNG</td>
<td>Κατάληψη</td>
<td>SQUATTING</td>
<td>V191</td>
<td></td>
</tr>
<tr>
<td>PERSÖNLICHE FREIHEIT</td>
<td>Εγαλίτα</td>
<td>EQUALITY</td>
<td>P192</td>
<td></td>
</tr>
<tr>
<td>Aristerá</td>
<td>LEFT</td>
<td>V192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δεξιά</td>
<td>RIGHT</td>
<td>V193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUNDESWEHR</td>
<td>ARMY</td>
<td>V206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>εκπαιδευτικό σύστημα</td>
<td>EDUCATIONAL SYSTEM</td>
<td>V207</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POLITIE</td>
<td>POLICE</td>
<td>V210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΠΑΡΛΑΜΕΝΤ</td>
<td>PARLIAMENT</td>
<td>v211</td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>Δημόσια Υπηρεσία</td>
<td>CIVIL SERVICE</td>
<td>V212</td>
<td></td>
</tr>
<tr>
<td>κρατικές παροχές</td>
<td>BENEFICI SOCIAL</td>
<td>SOCIAL BENEFITS</td>
<td>v233</td>
<td></td>
</tr>
<tr>
<td>CONSUM DE DROGURI</td>
<td>DRUG USE</td>
<td>V236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LÜGE</td>
<td>Minciună</td>
<td>LIE</td>
<td>V237</td>
<td></td>
</tr>
<tr>
<td>ZAHLUNG VON SCHMIERGELDERN</td>
<td>PAYMENT OF A BRIBE</td>
<td>V239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>δωροδοκούματα/δωροδοκία</td>
<td>ACCEPTING BRIBE/BRIBE</td>
<td>V239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCWARZFAHREN</td>
<td>FARE DODGING</td>
<td>V247</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Appendix 6 New term suggestions for ELSST from SHARE

<table>
<thead>
<tr>
<th>German term</th>
<th>Greek term</th>
<th>English equivalent</th>
<th>Question number</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALLGEMEINER GESUNDHEITSZUSTAND</td>
<td>χρόνια ασθένεια</td>
<td>HEALTH STATUS</td>
<td>PH003</td>
</tr>
<tr>
<td>LANGZEITIG/CHRONISCH</td>
<td>περιορισμός</td>
<td>LONG-TERM/CHRONIC</td>
<td>PH004</td>
</tr>
<tr>
<td>EINSCHRÄNKUNG</td>
<td>παθήσεις / πάθηση</td>
<td>RESTRICTION/LIMITATION</td>
<td>PH005</td>
</tr>
<tr>
<td>ORGAN</td>
<td>ΜΕΡΟΣ ΤΟΥ ΣΩΜΑΤΟΣ</td>
<td>ORGAN</td>
<td>PH006</td>
</tr>
<tr>
<td>JAHR</td>
<td>έτος</td>
<td>YEAR</td>
<td>PH008</td>
</tr>
<tr>
<td>ZEITPUNKT/DATUM</td>
<td>IN DER VERGANGENHEIT/ΚÚΡΖLICH</td>
<td>POINT IN TIME/DATE</td>
<td>PH076</td>
</tr>
<tr>
<td>MONAT</td>
<td>μήνας</td>
<td>MONTH</td>
<td>PH077</td>
</tr>
<tr>
<td>IN DER VERGANGENHEIT/ΚÚΡΖLICH</td>
<td>παθήσεις / πάθηση</td>
<td>DISEASES/DISORDERS</td>
<td>PH071</td>
</tr>
<tr>
<td>ZURZEIT/MOMENTAN</td>
<td>ένταση του πόνου</td>
<td>PAIN INTENSITY</td>
<td>PH089</td>
</tr>
<tr>
<td>WÖCHENTLICH</td>
<td>φαρμακευτική αγωγή</td>
<td>MEDICINAL DRUG TREATMENT</td>
<td>PH085</td>
</tr>
<tr>
<td>NÄHRUNGSERGÄNZUNGSMITTEL</td>
<td>CURRENTLY</td>
<td>DIETARY SUPPLEMENTS</td>
<td>PH082</td>
</tr>
<tr>
<td>GROßE MENGEN</td>
<td>WÖCHENTLICH</td>
<td>LARGE QUANTITIES</td>
<td>PH082</td>
</tr>
<tr>
<td>GEWICHTSREDUZIERUNG/GEWICHTSABNAHME</td>
<td>χάσιμο βάρους</td>
<td>WEIGHT LOSS</td>
<td>PH095</td>
</tr>
<tr>
<td>BEWEGGRUND</td>
<td>BEWEGGRUND</td>
<td>MOTIVATION</td>
<td>PH066</td>
</tr>
<tr>
<td>ART/SORTE</td>
<td>TYPE/SORT</td>
<td>WEITSICHTIGKEIT</td>
<td>PH069</td>
</tr>
<tr>
<td>WEITSICHTIGKEIT</td>
<td>LONGSIGHTEDNESS</td>
<td>SHORTSIGHTEDNESS</td>
<td>PH043</td>
</tr>
<tr>
<td>KURZSICHTIGKEIT</td>
<td></td>
<td></td>
<td>PH044</td>
</tr>
<tr>
<td>HÖRGERÄT</td>
<td>HEARING AID</td>
<td>PH045</td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>δυσκολία</td>
<td>DIFFICULTY</td>
<td>PH048</td>
<td></td>
</tr>
<tr>
<td>TÄGLICHE ROUTINE/VERRICHTUNGEN</td>
<td>DAILY ROUTINE</td>
<td>PH048</td>
<td></td>
</tr>
<tr>
<td>GESUNDHEITLICHE PROBLEME</td>
<td>HEALTH ISSUES/PROBLEMS</td>
<td>PH049</td>
<td></td>
</tr>
<tr>
<td>AKTIVITATEN</td>
<td>ACTIVITIES</td>
<td>PH049</td>
<td></td>
</tr>
<tr>
<td>ANWENDUNG</td>
<td>χρησιμοποίηση</td>
<td>PH048</td>
<td></td>
</tr>
<tr>
<td>TÄGLICH</td>
<td>DAILY</td>
<td>BR001</td>
<td></td>
</tr>
<tr>
<td>TABAKKONSUM</td>
<td>TOBACCO CONSUMPTION</td>
<td>BR005</td>
<td></td>
</tr>
<tr>
<td>προϊόντα καπνίσματος</td>
<td>SMOKING PRODUCTS</td>
<td>BR005</td>
<td></td>
</tr>
<tr>
<td>E-ZIGARETTEN</td>
<td>E-CIGARETTES</td>
<td>BR005</td>
<td></td>
</tr>
<tr>
<td>STANDARDEINHEITEN ALKOHOL</td>
<td>UNITS OF ALCOHOL</td>
<td>BR040</td>
<td></td>
</tr>
<tr>
<td>ANSTRENGUNG</td>
<td>VIGOUR</td>
<td>BR015</td>
<td></td>
</tr>
<tr>
<td>δραστηριότητα ες χαμηλής κατανάλωσης ενέργειας?</td>
<td>LOW ENERGY ACTIVITIES</td>
<td>BR016</td>
<td></td>
</tr>
<tr>
<td>ΔΙΑΤΡΟΦΙΚΕΣ ΣΥΝΗΘΕΙΕΣ</td>
<td>EATING HABITS</td>
<td>BR026</td>
<td></td>
</tr>
<tr>
<td>MILCHPRODUKTE</td>
<td>DAIRY PRODUCTS</td>
<td>BR026</td>
<td></td>
</tr>
<tr>
<td>HÜLSENFRÜCHTE</td>
<td>LEGUMES</td>
<td>BR027</td>
<td></td>
</tr>
<tr>
<td>FLEISCH</td>
<td>MEAT</td>
<td>BR028</td>
<td></td>
</tr>
<tr>
<td>FISCH</td>
<td>FISH</td>
<td>BR028</td>
<td></td>
</tr>
<tr>
<td>GEFLUGEL</td>
<td>POULTRY</td>
<td>BR028</td>
<td></td>
</tr>
<tr>
<td>VERZEHR</td>
<td>CONSUMPTION</td>
<td>BR029</td>
<td></td>
</tr>
</tbody>
</table>
Glossary
Many of the terms in this glossary have competing definitions and interpretations. The definitions provided below are intended to clarify their use within this document.

antonym: An antonym is a word that has the opposite meaning to another.

associative relationship: An associative relationship holds between concepts that are not hierarchically related, but are otherwise closely related.

Broader Term (BT): A Broader Term is a Preferred Term representing a concept that is broader than the one in question.

controlled vocabulary: A controlled vocabulary is a prescribed list of terms used for indexing and/or searching for resources. A controlled vocabulary typically includes Preferred Terms and Use For terms, and is restricted to a specific domain. A thesaurus is an example of a controlled vocabulary.

data collection: See study.

data documentation: Data documentation explains how data were created or digitised, what data mean, what their content and structure are, and any manipulations that may have taken place.

Data Documentation Initiative (DDI): The DDI is an international metadata standard originally designed to describe social, behavioural, and economic data.

data lifecycle: The data lifecycle is the stages data go through from their initial planning to re-use.

harmonised variable: A harmonised variable is a standardised variable designed to improve comparability of different surveys and measures.

hierarchical relationship: A hierarchical relationship in a controlled vocabulary is a relationship between a Broader Term and a Narrower Term.

hierarchy: A hierarchy is a structure in a controlled vocabulary where terms are arranged in order of their specificity, from the broadest to the most specific level. All terms in a hierarchy, except the top term, have one or more Narrower Terms. The number of hierarchies in a controlled vocabulary can vary, as can the number of levels within each hierarchy.

History Note: a History Note describes the history of a term.

homograph: A homograph is one of two or more words that are written in the same way, but have different meanings.

index term: Index terms, or keywords, are allocated to studies in order to describe subject content. They may either come from a controlled vocabulary such as a thesaurus, or be freely assigned (uncontrolled). The term ‘index term’ is used in this document to refer to both types.

indexing: Indexing is the process of evaluating information and assigning index terms to it.
Indexing language: An indexing language is used for assigning terms to describe the content of resources. Terms may belong to a controlled vocabulary or be freely assigned, i.e. the indexer can choose any term they like to describe the resource.

metadata: Metadata describe the content, context and provenance of datasets in a standardised and structured manner, and are used for resource discovery.

Narrower Term (NT): A Narrower Term is a Preferred Term representing a concept that is narrower than the one in question.

near synonym: A near synonym is one of two or more terms whose meanings are generally regarded as different in ordinary usage but which may be treated as labels for the same concept in a given controlled vocabulary. Also referred to as a quasi-synonym.

Nesstar: Nesstar is a software system for data publishing and online analysis.

Non-preferred term: see Use For term

precision: Precision is a measure of a search system's effectiveness in terms of retrieving only relevant results.

Preferred Term (PT): A Preferred Term is used to represent a concept when indexing. A Preferred Term is sometimes referred to as a descriptor.

Related Term (RT): A Related Term has an associative relationship with the one in question.

round: see wave

scale: A scale is a collection of related questions that measure an underlying concept (CLOSER project glossary. Available at https://wiki.ucl.ac.uk/display/CLOS/Glossary)

Scope Note (SN): A Scope Note defines or clarifies the semantic boundaries of a concept.

Scope Note Source (SNS): A Scope Note Source describes the source of a Scope Note.

Source Language (SL): The Source Language is the language that serves as a starting point in translation or a search for term equivalents

study: a set of data files and accompanying documentation. It is also known as a data collection. (UK Data Archive (2017): Data collection indexing procedures, Version 6.00, October 2017)

survey: A survey is an investigation about the characteristics of a given population by means of collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology. (see the OECD Glossary of Statistical Terms, https://stats.oecd.org/glossary/detail.asp?ID=2620, accessed 27 June 2018)

synonym: A synonym is one of two or more terms denoting the same concept
**synonymy**: Synonymy is a type of semantic relation in which two words or terms have the same or very similar meaning.

**Target Language** (TL): A Target Language provides a translation or an equivalent for a term in the source language.

**thesaurus**: A thesaurus is a controlled vocabulary in which concepts are represented by terms organised so that relationships are made explicit, and **Preferred Terms** are accompanied by **Use For terms** for synonyms or near-synonyms.

**top term**: a top term is the highest level term of a hierarchy.

**Use For term** (UF): A Use For term is a term that is not used for indexing, but which guides the user to a Preferred Term instead. Use For terms are typically synonyms or near synonyms, but may also be antonyms. They are sometimes referred to as Non-preferred terms.

**Use note**: A Use Note provides guidance for indexers to know when to apply a term, and how to distinguish it from closely related terms.

**wave**: A wave is a round of data collection in a longitudinal survey (e.g., the base year and each successive follow-up are each waves of data collection) or cross-sectional survey; a wave can also be referred to as a 'round'.
Abbreviations and acronyms

**BT**: Broader Term

**CESSDA**: Consortium of European Social Science Data Archives (CESSDA)

**DDI**: Data Documentation Initiative

**ELSST**: European Language Social Science Thesaurus

**ESS**: European Social Survey

**EVS**: European Values Study

**HASSET**: Humanities and Social Science Electronic Thesaurus

**NT**: Narrower Term

**PT**: Preferred Term

**RT**: Related Term

**SHARE**: Survey of Health, Ageing and Retirement in Europe

**SL**: Source Language

**SN**: Scope Note

**SNS**: Scope Note Source

**TL**: Target Language

**UF**: Use For Term

**UN**: Use Note