



# seriss

SYNERGIES FOR EUROPE'S  
RESEARCH INFRASTRUCTURES  
IN THE SOCIAL SCIENCES

Deliverable Number: D3.10

Deliverable Title: Guidelines for the management of ELSST content

Work Package: Number and Title: Number and Title: WP3: Maximising equivalence through translation

Deliverable type: Report

Dissemination status: Public

Submitted by: CESSDA AS

Authors: Lorna Balkan (UK Data Archive)

Date Submitted: July 2017

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654221.





[www.seriss.eu](http://www.seriss.eu)  @SERISS\_EU

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. (2017) Guidelines for the management of ELSST content, Deliverable 3.10 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](http://www.seriss.eu/resources/deliverables)

## Summary

These guidelines describe how to manage the content of the European Social Science Thesaurus (ELSST), following international standards and best practice. It is aimed primarily at ELSST translators and content developers, but will also be of interest to end-users.

The guidelines take account of ISO 25964-1, the latest international standard on thesaurus construction, published by the International Organization for Standardization, as well as the work of other knowledge organisation experts and thesaurus developers.

The guidelines are divided into different parts. Within each part, best practice for thesauri in general is described (and shown in italics), followed by specific guidelines for ELSST.

The first part of the guidelines contains an overview of and background information on ELSST, including access and copyright issues. The second part of the guidelines introduces the main linguistic elements of ELSST: concepts, terms, relationships and notes. The third part looks at the management structure adopted for ELSST. The fourth part covers the construction and maintenance of the thesaurus, while the fifth part focuses on the translation process. The sixth part deals with quality control issues.

## Contents

Summary .....	3
Document scope .....	7
Document status .....	7
PART 1: ELSST overview and background information.....	8
1 Purpose of ELSST .....	8
2 Target Audience of ELSST .....	8
3 History of ELSST.....	8
4 Scope of ELSST.....	9
5 Languages of ELSST.....	9
5.1. Language versions.....	9
5.2. Language varieties.....	9
5.3. Language status.....	10
6 Versioning .....	10
7 Accessing ELSST .....	11
7.1. Online access.....	11
7.2. Offline access.....	11
8 Copyright.....	12
9 Standards.....	12
10 Contributing suggestions .....	12
PART 2: Linguistic content and form .....	13
11 Linguistic content of ELSST.....	13
11.1. Overview .....	13
11.2. Concepts.....	13
11.3. Terms .....	13
11.3.1. Preferred Terms (PTs) .....	13
11.3.2. Use For terms (UFs).....	13
11.4. Relationships.....	14
11.4.1. Overview.....	14
11.4.2. Hierarchical relationship.....	14
11.4.3. Associative relationship.....	14
11.4.4. Equivalence relationship .....	15
11.5. Note fields .....	15
11.5.1 Overview.....	15
11.5.2. Scope Note (SN) .....	16
11.5.3. Scope Note Source (SNS) .....	16
11.5.4. Use Note (UN).....	16
11.5.5. History Note (HN).....	16

12. Form of ELSST terms .....	17
12.1. Part of speech .....	17
12.2. Word order .....	17
12.3. Acronyms and abbreviations.....	17
12.4. Singular or plural forms.....	17
12.5. Qualifiers .....	18
12.6. Diacritics.....	19
12.7. Gender .....	20
12.8. Spelling.....	20
12.9. Punctuation within terms.....	21
PART 3: Management structure .....	22
13. Management structure .....	22
13.1. Overview .....	22
13.2. Roles and responsibilities .....	22
13.2.1. Overview.....	22
13.2.2. Language spokespersons.....	23
13.2.3. Translators .....	23
13.2.4. Reviewers.....	24
13.3. Internal communication and decision making.....	24
13.4. Education and training of team members .....	24
PART 4: Thesaurus construction and maintenance.....	25
14. Creating and updating content.....	25
14.1. Overview .....	25
14.2. Vocabulary resources .....	25
14.3. Addition of new concepts and terms.....	25
14.3.1. Criteria for inclusion of concepts.....	25
14.3.2. Selecting a Preferred Term.....	26
14.3.3. Selecting Use For terms.....	26
14.3.4. Selecting Broader Terms, Narrower Terms, and Related Terms.....	27
14.3.5. Selecting Scope Notes.....	27
14.3.6. Scope note source .....	28
14.3.7. Choosing Use Notes .....	28
14.4. Deletion of existing concepts .....	28
14.5. Changes to existing concepts and terms.....	28
PART 5: Translation management .....	29
15. Pre-translation process .....	29
15.1. Overview .....	29
15.2. Preliminary actions.....	29

15.3. Feedback mechanism .....	30
15.4. Tracking progress .....	30
16 Translation process .....	30
16.1. Overview .....	30
16.2. Overall aims and approach .....	30
16.3. Some language-specific issues .....	31
16.3.1. Overview .....	31
16.3.2. Ambiguity .....	31
16.3.3. Singular or plural forms .....	31
16.3.4. Qualifiers .....	31
16.4. What to translate .....	32
16.4.1. Overview .....	32
16.4.2. Preferred terms .....	32
16.4.3. Use For terms .....	34
16.4.4. Scope Notes .....	34
16.4.5. Scope Note Source .....	34
16.4.6. Use Note .....	34
16.4.7. History Note .....	34
17 Post-translation process .....	34
PART 6: Quality control .....	35
18. Quality control .....	35
18.1. Overview .....	35
18.2. Quality criteria for ELSST .....	35
18.3. Quality assurance methods .....	35
References .....	37
Other references consulted .....	39
Appendix 1: Vocabulary resources for translation .....	40
Appendix 2: Quality criteria for ELSST .....	46
Glossary .....	47
Abbreviations and acronyms .....	51

## Document scope

These guidelines contain recommendations for the management of the content of the European Social Science Thesaurus (ELSST). It is aimed primarily at content developers and translators of ELSST, but will also be of interest to end-users.

The guidelines cover the following subjects: overview and background information; linguistic structure and content; management structure; construction and maintenance, the translation process, and quality control.

The guidelines also contain a glossary and a list of acronyms. Terms referred to in the glossary are highlighted in bold when they first appear in the text.

Excluded from these guidelines is guidance on the thesaurus management system, which is covered in the related document 'ELSST Thesaurus Management System: User Guide for Translators'.

## Document status

This document was first created as one of the key deliverables of the Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS) project, funded under the European Union's Horizon 2020 research and innovation programme GA No: 654221.

# PART 1: ELSST overview and background information

## 1 Purpose of ELSST

The European Social Science Thesaurus (ELSST) is a broad-based, multilingual **thesaurus** for the social sciences.

A thesaurus is a **controlled vocabulary** consisting of **concepts, terms** and relationships. Concepts represent 'units of thought'. Each concept is represented by one or more terms in each language of the thesaurus, and the relationships between concepts are shown explicitly.

The main purpose of a thesaurus is to improve **indexing** and information retrieval. It does this by guiding the user to use the same term for the same concept. From the set of terms used to represent each concept, one is chosen for each language as the **Preferred Term**, and only it may be used for indexing. The other terms, which are **synonyms** or **Use For terms**, are used to guide the indexer to the Preferred Term. This in turn makes searching more efficient, since a query using the Preferred Term or any of its Use For terms will find the same set of objects.

Searching using thesaurus terms, rather than free-text, also improves search **precision**, since, unlike everyday language, Preferred Terms must not be ambiguous.

The explicit relationships between concepts also help the searcher by offering useful suggestions for **query expansion**.

A multilingual thesaurus promotes information retrieval across languages: search terms formulated in one language of the thesaurus can be used to find objects indexed with terms in any other language of the thesaurus.

ELSST is available in **Simple Knowledge Organization System** (SKOS) format. SKOS has become increasingly valued throughout the thesaurus community as a means of incorporating thesauri into existing applications, working with and encoding thesauri, and using in **Linked Data**.

## 2 Target Audience of ELSST

The primary users of ELSST are social science researchers and policy makers. Other users include social science libraries and data archives (especially the **Consortium of European Social Science Data Archives** (CESSDA) who are the main developers of ELSST), information specialists who are dealing with the needs of these users, publishers, and the general public.

## 3 History of ELSST

ELSST was originally developed in 2000 as part of the EU-funded Language Independent Metadata Browsing of European Resources (LIMBER) project in order to aid cross-language information retrieval of social science datasets in the CESSDA data portal. It has been further enhanced and extended through additional funding from the UK's Economic and Social Research Council (ESRC), the University of Essex, and through subsequent EU grants including Multilingual Access to Data Infrastructures of the European Research Area (MADIERA) and the CESSDA-Preparatory Phase Project (CESSDA-PPP). Most recently it has been funded as part of the ESRC-funded CESSDA-ELSST project and the SERISS project.

ELSST was originally based on the Humanities and Social Science Electronic Thesaurus (HASSET), the foremost British English social science thesaurus, which is owned and managed by the UK Data Archive at the University of Essex. The thesauri are still closely related: shared, or ‘core’ concepts have the same Preferred Term labels and **Broader Terms**.

## 4 Scope of ELSST

*Thesauri are usually domain-specific. The scope of a thesaurus should be clearly stated.*

ELSST concepts cover the core social science disciplines: politics, sociology, economics, education, law, crime, demography, health, employment, information and communication technology and, increasing environmental science. The concepts reflect the holdings of the CESSDA archives but do not claim to provide exhaustive coverage of any one subject area. Concepts that are culture-, language-, or institution-specific are excluded.

## 5 Languages of ELSST

### 5.1. Language versions

*Each concept in a multilingual thesaurus should be represented in every language of the thesaurus (ISO 25964-1, p.50).*

Currently ELSST is available in 12 languages: Danish, Czech, English, Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Spanish, and Swedish. The Slovenian translation is under way.

Some language versions are more mature than others, but at least 70% of terms of each language are available in each published version of the thesaurus.

### 5.2. Language varieties

*Different varieties or dialects of a language may either be treated as separate languages of a thesaurus, or be accommodated within the same language version.*

In ELSST, both approaches are possible. Currently, however, different varieties of the same language are contained within the same language version. For example, the German version of ELSST contains terms from both Germany and Austria. However, all Preferred Terms should be chosen from the same language variety. All other varieties should be included as Use For terms.

Figure 1 shows the current situation for ELSST.

Note: Currently all language varieties of a language are assigned the same language code, so cannot be distinguished from each other. For example, all German terms are assigned the code ‘de-DE’.

Language	Language variety of Preferred Terms	Additional language varieties of Use For Terms
Czech	Czech	
Danish	Danish	

English	British English	
Finnish	Finnish	
French	French French	Swiss French (culture-specific terms only (e.g. political institutions))
German	German German	Austrian German
Greek	Greek	
Lithuanian	Lithuanian	
Norwegian	Norwegian	
Romanian	Romanian	
Slovenian	Slovenian	
Spanish	Spanish Spanish	
Swedish	Swedish	

Figure 1 Language varieties in ELSST

### 5.3. Language status

*In a multilingual thesaurus, all languages should have equal status (ISO 25964-1, p.50).*

ISO 25964-1 (p.92) recognises three methods of constructing a thesaurus:

- a) translation of a monolingual thesaurus
- b) merging of several distinct monolingual thesauri
- c) simultaneous construction of the various language versions of the multilingual thesaurus

In ELSST, approach (a) was originally adopted, since ELSST was derived from HASSET. English remains the sole **source language** from which all **target languages** are derived. In this way, ELSST aims to promote consistency in translation across languages. However, the addition of new concepts and revision to existing concepts are carried out in consultation with target language developers. In this way, it is hoped to avoid bias towards source language concepts and structure.

## 6 Versioning

*New versions of a thesaurus should be released on a regular basis, and users informed of changes since the last version.*

ELSST aims to release new versions on an annual basis, or more frequently, if major changes have been made or a new language added.

The thesaurus management system contains two versions of ELSST: the 'Current Version' and the 'Next Version'.

The Current Version is the latest released version of ELSST, and is available for all to search

and browse online.

The Next Version, which is only viewable to registered users behind a login, is continually being updated, and becomes the new Current Version on its release.

## 7 Accessing ELSST

### 7.1. Online access

*To promote re-use of a thesaurus, it should be possible to view the thesaurus freely online, and obtain a copy of its contents, possibly under licence. Permissions to make changes to the thesaurus must be carefully managed.*

ELSST has four different categories of user:

- general public
- guests: guests are registered users of the UK Data Service who have logged in to ELSST
- translators
- administrators

The Current Version of ELSST is available for all to search or browse online. The Next Version of ELSST is only available to view behind a login.

To obtain an ELSST login, users need to be registered with the UK Data Archive.

Read access for the different user types is shown in Figure 2 below.

	Current version			Next Version		
	ELSST suggestions	ELSST history	View Concept	ELSST suggestions	ELSST history	View Concept
General public	X	X	✓	X	X	X
Guests	✓	✓	✓	✓	✓	✓
Translators	✓	✓	✓	✓	✓	✓
Administrators	✓	✓	✓	✓	✓	✓

Figure 2 Read access for ELSST users

Write access to the Next Version of ELSST is via controlled permissions only:

- translators may edit their own language only
- administrators may edit all languages, including the source language

### 7.2. Offline access

The full content of the thesaurus may be provided to individuals or organisations who hold a valid ELSST licence. See Section 8 on how to apply for a licence.

Thesaurus files are normally distributed in SKOS format.

## 8 Copyright

ELSST is copyright-protected. The complete thesaurus, or any substantial part of the thesaurus, may not be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without the permission of the copyright holder.

The copyright in the natural language translations is held by the individual translators or the relevant translating organisation; the database rights in the organisation and collection of the data are held by the University of Essex. The rights in the underlying application are also held by the University of Essex.

Organisations who are interested in translating, adapting, or using the complete thesaurus, or any substantial part of it in their own organisations, should discuss licensing options for ELSST by contacting the ELSST Administrator via the link on the ELSST website.

## 9 Standards

*To promote **interoperability** with other **Knowledge Organization Systems (KOS)**, it is recommended that common standards be followed when constructing a thesaurus.*

ELSST seeks to comply with the latest international guidelines on the construction of monolingual and multilingual thesauri: ISO 25964-1 (see ISO 25964-1 (2011)). SKOS is also a widely used standard, recommended by the World Wide Web Consortium (W3C), to represent and promote the interoperability of thesauri and other Knowledge Organization Systems.

## 10 Contributing suggestions

*A thesaurus should provide a means for thesaurus developers and other users to contribute comments and make suggestions for changes.*

In ELSST, the general public can use an online form on the management application to send their suggestions.

Registered users and developers of ELSST may use the Suggestions interface in the Next Version of ELSST.

## PART 2: Linguistic content and form

### 11 Linguistic content of ELSST

#### 11.1. Overview

ELSST consists of a set of concepts, represented by terms, organised in such a way that the relationships between the concepts are made explicit. Terms may also be associated with note fields.

#### 11.2. Concepts

*Concepts represent 'units of thought' (ISO 25964-1, p. 3). In a multilingual thesaurus, all concepts should be relevant to all or most languages of the thesaurus.*

In ELSST, concepts should be culture-neutral: they are intended to reflect a European perspective, rather than be country-, language- or institution-specific.

All ELSST concepts are associated with a **Globally Unique Identifier** (GUID).

#### 11.3. Terms

ELSST follows thesaural conventions in defining two types of term: Preferred Terms (PTs) and Use For terms (UF).

##### 11.3.1. Preferred Terms (PTs)

*Preferred Terms (PTs) represent the concepts contained in the thesaurus and are used for indexing purposes. There should be one Preferred Term per language for each concept of the thesaurus.*

In ELSST, Preferred Terms are unique to each language version of ELSST (i.e. there may be no duplicate Preferred Terms within any one language version).

##### 11.3.2. Use For terms (UFs)

*Use For terms (UFs) are also known as **Non-Preferred terms**. They are not used for indexing, but guide the user to their associated Preferred Term. They support more efficient indexing and information retrieval since the user needs to formulate only one query, using either the Preferred Term or any of its Use For terms, in order to retrieve objects about the same concept.*

Use For terms can be any of the following:

- synonym or **near-synonym** (including a **lexical variant**), for example:  
SOCIAL HOUSING  
    **UF** COUNCIL HOUSING  
    **UF** PUBLIC HOUSING
- **antonym**, for example:  
SATISFACTION  
    **UF** DISSATISFACTION
- **upward posting**, i.e. a specific term subsumed by a broader concept. For example:  
PARENTAL BENEFITS

## UF MATERNITY ALLOWANCES UF PATERNITY BENEFITS

In ELSST, Use For terms:

- must be unique to each language version of ELSST (i.e. there can be no duplicate UFs within any one language version)
- cannot be the same as a Preferred Term in the same language version
- may be language- or culture-specific (unlike Preferred Terms)

### 11.4. Relationships

#### 11.4.1. Overview

**Hierarchical relationships** and **associative relationships** exist between concepts in a thesaurus.

ELSST is a **symmetrical** thesaurus, which means that all language versions share the same structure of hierarchical and associative relationships.

Additionally, terms (not concepts) may be related to each other via an **equivalence relationship**.

#### 11.4.2. Hierarchical relationship

*Preferred Terms may be 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. For example:*

NURSES  
    **BT** HEALTH PROFESSIONALS  
    **NT** MIDWIVES

Broader Terms are paired with Narrower Terms in a **reciprocal relationship**. For example:

MIDWIVES  
    **BT** NURSES

The display of hierarchical relationships enables indexers and searchers to find terms at the correct level of specificity, and allows searchers to widen or narrow their search.

Thesauri may permit **polyhierarchical structures**, where a Preferred Term has more than one Broader Term. For example:

SOCIAL HOUSING  
    **BT** HOUSING  
    **BT** SOCIAL WELFARE

In ELSST, polyhierarchical structures should be kept to a minimum to promote ease of browsing.

The top node of a hierarchy is referred to as the **Top Term**, and is also the name of the hierarchy.

#### 11.4.3. Associative relationship

*Preferred Terms may stand in an associative, but non-hierarchical relation to each other. This is represented by **Related Terms (RTs)**. For example:*

SOCIAL HOUSING  
RT HOUSING POLICY  
RT HOUSING TENURE  
RT RENTED ACCOMMODATION

Associative relationships are also reciprocal. For example:

HOUSING POLICY  
RT SOCIAL HOUSING

The display of associative relationships enables users to extend their search to related terms.

#### 11.4.4. Equivalence relationship

*In a multilingual thesaurus, an **equivalence relationship** exists between two terms in a thesaurus that both represent the same concept. In ordinary discourse, terms that are quasi-synonyms may represent slightly different concepts. After inclusion in the thesaurus, however, the equivalence clarifies that both are regarded as representing the same concept. When two or more such terms are in the same language within a monolingual or multilingual thesaurus, one of them is designated a **Preferred Term**, and the other(s) **Non-Preferred terms** [i.e. Use For terms]; when two or more such terms are in the different languages of a multilingual thesaurus, each of them may be a Preferred Term in its own language respectively, and the relationship is known as **cross-language equivalence** (ISO 25964-1, p.4).*

In ELSST, examples of equivalence in the same language are given in Section 11.3.2.

An example of cross-language equivalence is the En(lish) term SOCIAL ATTITUDES and the Fr(en)ch term ATTITUDES SOCIALES:

EN SOCIAL ATTITUDES  
FR ATTITUDES SOCIALES

### 11.5. Note fields

#### 11.5.1 Overview

*Thesaurus terms are commonly associated with note fields, including scope notes, history notes and definitions. Additional notes may also be used for administrative purposes. Not all note types need be displayed to users. Some note fields, such as the date the term was created or modified, may be generated automatically by the thesaurus management system.*

Originally ELSST contained scope notes only. Following ISO 25964-1 (p.20), these were used to clarify the meaning of the term as it is used in the thesaurus, and to provide advice on the usage of the term either to the indexer or the searcher. Additionally, ELSST scope notes were also used to provide information about the source of the scope note, and changes to the term over time.

Work is currently under way to separate the ELSST scope note field into four distinct fields: Scope Note, Use Note, Scope Note Source, and History note. The content of each of these separate note fields is described below. Where scope notes have not yet been subdivided, they contain all types of information as before.

### 11.5.2. Scope Note (SN)

In ELSST, a Scope Note (SN) is used to explain the meaning of a Preferred Term in the thesaurus where the Preferred Term's label or place in the hierarchy is not enough to disambiguate it.

A Scope Note defines or clarifies the semantic boundaries of a concept as it is used in the thesaurus. A Scope Note is particularly useful where:

- a term can have several meanings in normal usage: the Scope Note is used to restrict the concept to only one of these meanings
- the meaning of the term in normal usage can be interpreted too broadly or too narrowly
- the scope of a concept in the thesaurus is not always the most common meaning associated with its Preferred Term (or any of its Use For terms) in ordinary discourse

An example of a Scope Note in ELSST is:

POLITICAL INFLUENCE

**SN** THE EXTENT TO WHICH THE PUBLIC CAN AFFECT POLITICAL DECISION-MAKING.

### 11.5.3. Scope Note Source (SNS)

A Scope Note Source (SNS) records the (digital) bibliographic reference for a term.

An example of a Source Note Source in ELSST is:

MARRIAGE

**SNS** UK DATA ARCHIVE

### 11.5.4. Use Note (UN)

A Use Note (UN) is used in ELSST to provide additional information about a Preferred Term for indexers. It may refer to other concepts in the thesaurus that are included in or excluded from the scope of the concept in question, or to distinguish between Preferred Terms that have overlapping meanings in natural language, but not within the context of the thesaurus. For example:

CHILD LABOUR:

**SN:** WORK CARRIED OUT BY CHILDREN OF SCHOOL AGE, USUALLY PART-TIME.

**UN:** WHERE THE EMPHASIS IS ON THE EXPLOITATION OF CHILD WORKERS USE "EXPLOITATION OF CHILD LABOUR"

### 11.5.5. History Note (HN)

*A History Note records the date of introduction or the changes made to an existing concept. It may apply to a single Preferred Term, Use For term or concept.*

A History Note should be added when:

- a new Preferred Term is added to the thesaurus
- a change is made to an existing term that affects the concept's scope in different periods of application

The following is an example of a History Note in ELSST:

MARRIAGE

**UF** SAME-SEX MARRIAGE

**HN** PRIOR TO JULY 2015 'SAME-SEX MARRIAGE' WAS A UF OF CIVIL PARTNERSHIPS

## 12. Form of ELSST terms

### 12.1. Part of speech

*A thesaurus term should preferably consist of a noun or noun phrase.*

*To express activities, the commonly accepted conventions of each language should be followed. For example, gerunds may be used in languages such as English, while infinitives may be acceptable in other languages.*

*Adjectives can be useful as part of a noun phrase, but should be avoided as stand-alone terms, as they can be a source of confusion.*

*Terms may consist of a single word or be **compound terms** (i.e. a term that can be split morphologically into separate components). For example:*

HEALTH CARE can be split into 'HEALTH' and 'CARE'

SCHOOLCHILDREN can be split into SCHOOL and CHILDREN

### 12.2. Word order

*The order of words within a compound term should follow the normal word order for the language in question. For example, the following is acceptable:*

MEDICAL HISTORY

but not:

HISTORY, MEDICAL

### 12.3. Acronyms and abbreviations

*In general, abbreviations and **acronyms** should be avoided. However, they can be used as Preferred Terms when they are well established.*

e.g. LASERS

Otherwise, they should be retained as Use For terms only. For example:

BOVINE SPONGIFORM ENCEPHALOPATHY

**UF** BSE

### 12.4. Singular or plural forms

*Different conventions exist in different languages concerning the use of singulars or plurals in thesauri. Each language should follow the thesaurus conventions for their language.*

In ELSST, each language is free to follow the thesaurus conventions for their language as far as choosing singular or plural forms of nouns.

In English, the choice of whether to use a singular or plural form of a term is based on whether a particular term is a **count noun** or a **non-count noun**.

- Count nouns are names of countable entities that are subject to the question 'How many?' but not 'How much?'. For example:

DOMESTIC APPLIANCES  
BIRDS

- Non-count nouns are names of concepts such as materials or substances that are subject to the question 'How much?' but not 'How many?' For example:

COAL  
EQUIPMENT

Note that:

- Some nouns can have both count and non-count versions. The following examples are from English:
  - LAW (non-count noun)  
Meaning = 'The body of rules, whether proceeding from formal enactment or from custom, which a particular state or community recognizes as binding on its members or subjects.' (Oxford English Dictionary)
  - LAW (count noun)  
Meaning = 'One of the individual rules which constitute the 'law' [in the sense of the non-count noun] of a state or polity.' (Oxford English Dictionary)
- Both the count and non-count versions of the same noun may appear within the same language version of ELSST.

In English, the source language of ELSST, count nouns are expressed as plurals, and non-count nouns usually as singulars. All languages except Slovenian have chosen to represent count nouns in their plural form. In Slovenian, the singular form is preferred, since Slovenian has three values for grammatical number: singular, dual, and plural.

## 12.5. Qualifiers

**Qualifiers**, shown as words in parentheses after a term, are used to disambiguate **homographs** (words that have the same spelling but different meanings) within the same language in the thesaurus. A qualifier should be added to each homographic term. For example:

- LABOUR (WORK)
- LABOUR (PREGNANCY)

*Even if there is only one homographic term within the thesaurus, a qualifier may be added to promote interoperability. For example:*

- COFFEE (CROP)
- AIDS (DISEASE)

*In any language that contains both the singular and plural form of a term that refer to different concepts, the distinction between them may be reinforced by adding a scope note (in the case of Preferred Terms) and if possible a qualifier.*

*For example, ISO 25964-1 (p.29) recommends distinguishing the French term STATISTIQUE (meaning 'statistics' in English) from STATISTQUES (meaning 'data' in*

English) by adding the following qualifiers:

- STATISTIQUE (SCIENCE)
- STATISTIQUES (DONNÉES)

Qualifiers should not, however, be overused, as they make terms slightly cumbersome to apply and some electronic systems have difficulties in applying them.

## 12.6. Diacritics

A multilingual thesaurus must be able to support all accents and diacritics.

ELSST uses UTF-8 encoding. Letters containing diacritics may form part of the alphabet, in some languages. Other diacritics or accents may be added to existing letters.

Language	Alphabet used in ELSST	Additional accents which are not part of the alphabet and which are used in ELSST
Czech	Czech alphabet (42 letters)	none
Danish	Danish/Norwegian alphabet (29 letters)	none
English	Basic Latin alphabet (26 letters)	none
Finnish	Finnish alphabet (29 letters)	none
French	basic Latin alphabet (26 letters)	acute, grave, circumflex, diaeresis and cedilla
German	basic Latin alphabet (26 letters)	all umlauts (Ä, Ö, Ü), plus AE, OE, UE variants in non-preferred terms
Greek	Greek alphabet (24 letters)	diaeresis and acute
Lithuanian	Lithuanian alphabet (32 letters)	none
Norwegian	Danish/Norwegian alphabet (29 letters)	none
Romanian	Romanian alphabet (31 letters)	none
Slovenian	Slovenian alphabet (25 letters)	none
Spanish	Spanish alphabet (27 letters)	none
Swedish	Swedish alphabet (29 letters)	none

Figure 3: Alphabets and accents in ELSST

## 12.7. Gender

There is no general consensus on how to deal with gender within thesauri.

In ELSST, where a language marks a term for gender (e.g. German distinguishes between male teachers (LEHRER) and female teachers (LEHRERINNEN), the following strategies for choosing a preferred term can be adopted:

- use a general, neutral umbrella term  
e.g. LEHRPERSONAL in German, meaning TEACHING PERSONNEL
- use the most common form, be it either masculine or feminine  
e.g. MIDWIVES in English
- use a compound of the form 'X AND Y', or its equivalent in the target language, where both X and Y are introduced as UFs:  
e.g. LEHRER UND LEHRERINNEN in German  
**UF LEHRER**  
**UF LEHERINNEN**

The preferred option depends on the language and term concerned, and should be decided on a case-by-case basis.

## 12.8. Spelling

*To promote consistency, a single authoritative source should be used for the spelling of Preferred Terms. Alternative spellings can be added as Use For terms.*

In ELSST, the sources used for the spelling of Preferred Terms in each language are shown in Figure 4.

Language	Spelling source
Czech	Slovník spisovného jazyka českého
Danish	Dansk Sprognævn
English	Oxford English Dictionary
Finnish	MOT Kielitoimiston sanakirja
French	Dictionnaire de français Larousse
German	Duden Online-Wörterbuch
Greek	Λεξικό της κοινής νεοελληνικής
Lithuanian	tbc
Norwegian	tbc

Romanian	Dicționarul Explicativ al Limbi Române (DEX)
Spanish	Diccionario de la lengua española
Swedish	Svenska Akademiens ordlista

Figure 4: Spelling conventions in ELSST

## 12.9. Punctuation within terms

*Punctuation marks in thesauri should be kept to a minimum, since some systems find them difficult to handle.*

In ELSST, the following should not be used:

- Full-stops (.)
- Dashes (–)
- Colons (:)
- Semi-colons (;)

However, the following may be found:

- Apostrophes: apostrophes should be used in Preferred Terms. Variants without apostrophes should be added as Use For terms

e.g. WOMEN'S RIGHTS  
**UF WOMENS RIGHTS**

- Hyphens: hyphens in Preferred Terms should follow the spelling conventions for each language. Variants with or without hyphens should be added as Use For terms. For example:

CROSS-BORDER COOPERATION  
**UF CROSS-BORDER CO-OPERATION**

- Parentheses: parentheses are used for qualifiers, and cannot appear elsewhere in a term. For example:

LABOUR (WORK)

- Commas: commas may be used rarely in Preferred Terms and Use For terms. For example:

PROPERTY, OWNERSHIP AND TENURE

- Slashes: slashes can be used rarely in both Preferred Terms and Use For terms. For example:

INFORMATION/LIBRARY RESOURCES

## PART 3: Management structure

### 13. Management structure

#### 13.1. Overview

*A management structure for a multilingual thesaurus should be set up to oversee content development, translation, quality assurance, training, licensing, promotion and distribution of the thesaurus. Ideally, a representative from each language and language variety of the thesaurus will be involved in the content management of the thesaurus.*

The ELSST Thesaurus Management Team consists of representatives from each member organisation of ELSST. The UK Data Archive representative chairs the Management Team and is also the ELSST Administrator.

Members are mostly drawn from CESSDA and associated archives (see Figure 5). Where a language is represented by more than one organisation, the lead partner is indicated:

Language	Archive	Lead partner
Czech	Czech Social Science Data Archive (CSDA), Czech Republic	
Danish	Danish Data Archive (DDA), Denmark	
English	UK Data Archive, University of Essex, UK	
Finnish	Finnish Social Science Data Archive (FSD), Finland	
French	The Swiss Centre of Expertise in the Social Sciences (FORS), Switzerland	
German	Leibniz-Institute for the Social Sciences (GESIS), Germany	GESIS
	The Austrian Social Science Data Archive (AuSSDA), Austria	
Greek	National Centre for Social Research (EKKE), Greece (part of Social Data Network (So.Da.Net))	
Lithuanian	Lithuanian Data Archive for Humanities and Social Sciences (LiDA), Lithuania	
Mexican	Instituto Nacional de Estadística y Geografía (INEGI), Mexico	
Norwegian	NSD - Norwegian Centre for Research Data (NSD), Norway	
Romanian	Romanian Social Data Archive (RODA), Romania	
Slovenian	Social Science Data Archives (ADP), Slovenia	
Swedish	Swedish National Data Service (SND), Sweden	

Figure 5: Members of ELSST management team

#### 13.2. Roles and responsibilities

##### 13.2.1. Overview

*Thesaurus content development and translation work should be carried out by a multidisciplinary team consisting of linguists, translators, and subject experts. All need to be*

*familiar with the subject area of the thesaurus, as well as the aims and structure of a thesaurus. Ideally, content developers and translators will also have experience in, or an understanding of, the indexing process.*

In ELSST, development work and translation is carried out by language spokespersons for each language of ELSST, and any translators that they appoint. Subject experts should also be involved in the process, to review and give advice on the terms and concepts of the thesaurus.

### 13.2.2. Language spokespersons

Each language version and variety of ELSST is represented by a language spokesperson.

The source language spokesperson, i.e. the ELSST Administrator, is responsible for maintaining the source language and for co-ordinating the translation work. Other responsibilities include:

- managing permission rights for the thesaurus management system
- managing internal communications, including email lists and online meetings
- delivering translator training
- managing ELSST licences and issuing offline versions of the thesaurus to licensed users
- keeping thesaurus documentation up-to-date, including information on the ELSST website
- keeping track of changes to ELSST, and reporting changes to translators when required
- overseeing and co-ordinating new releases of ELSST
- contributing to the management and enhancement of the Thesaurus Management System

Target language spokespersons are responsible for co-ordinating the translation of ELSST into their language. They may undertake the translation work themselves, or appoint one or more translators to do it for them. Their responsibilities include:

- ensuring the timely update of their language version prior to an annual release
- appointing one or more translators for their language, where required
- ensuring that they, and any translators they appoint, receive translator training for ELSST
- liaising with other target language spokespersons who wish to contribute Use For terms to their language version of ELSST
- quality control of their language version, including making strategic decisions on language-specific issues for their language (e.g. use of diacritics)
- signing off a translation before the next release

### 13.2.3. Translators

*Translators should be native speakers of the target language, with an excellent knowledge of the source language. Formal qualifications in translation or experience of translation are desirable, but not essential.*

Translators of ELSST should also:

- have a good understanding of social science data
- be familiar with ELSST principles and guidelines
- attend ELSST translator training before starting to translate ELSST

Translators are responsible for:

- contributing to discussion of changes to ELSST
- translating ELSST into their language
- entering their translations into the thesaurus management system, if required

#### **13.2.4. Reviewers**

*Reviewers should be subject experts in, and native speakers of, the terms they are asked to review.*

Reviewers can be used at any time during the thesaurus development process in order to advise on the currency of concepts and Preferred Terms, the accuracy of scope notes, and the appropriateness of the thesaurus structure.

### **13.3. Internal communication and decision making**

*Good communication channels and decision making processes should be established between content developers. All meetings should be minuted and decisions documented.*

In ELSST, internal communication is conducted via a dedicated email list and via ELSST translator meetings that are held quarterly and chaired by the ELSST Administrator. All language spokespersons and translators may join the email list and attend translator meetings. The ELSST Administrator is responsible for drawing up the agenda of translator meetings, in collaboration with translators, for minuting the meetings, and for circulating the minutes to all members of the list.

Decisions are taken by consensus of the majority, either via email, or at the meetings.

### **13.4. Education and training of team members**

*All involved in the production of a thesaurus (content developers, translators and reviewers) should be fully versed in the structure and content of the thesaurus, the translation process, and how to access relevant documentation and the thesaurus management system.*

In ELSST, translator training is provided by the ELSST Administrator at regular intervals, and on request. It is recommended that all translators undertake training before starting to translate.

## PART 4: Thesaurus construction and maintenance

### 14. Creating and updating content

#### 14.1. Overview

This section contains recommendations on how to create new terms and concepts, and keep existing concepts up-to-date.

*Thesaurus maintenance must be carried out on a continual basis, since the meaning and usage of terms may change over time, and new terms emerge, while others become obsolete. Without regular updating, thesaurus terms and concepts may lose their currency and relevance.*

*Changes to concepts, terms and structure need to be carefully managed in order to preserve the integrity of the thesaurus and the data it is used to index. All changes to the thesaurus must be recorded and any records that are dependent on those changes, corrected.*

In ELSST, suggestions for new terms and concepts, and changes to existing terms and concepts, are received from both ELSST developers and end-users on a continuous basis (see Section 10), and reviewed at regular intervals. They also feed into the review of concept hierarchies which forms part of ongoing maintenance work.

#### 14.2. Vocabulary resources

*ISO 25564-1 (p.89) and information experts (Broughton (2006) and Hedden (2010)) recommend various resources that can be used to help select terms, Scope Notes, and relationships. They include:*

- *existing vocabularies, including other thesauri and classification schemes, provide suggestions for synonyms and relationships.*
- *published work such as books, journals, conference papers, and reports. Journals and conference papers may indicate new research trends and related terms (Broughton 2006).*
- *reference works such as dictionaries, terminologies and encyclopaedias. These can be consulted to confirm definitions of terms and their place in a hierarchical structure (ISO 25964-1, p.90).*
- *web resources such as navigational menus and feedback from web search logs are a good source of terms*
- *subject experts can provide term and Scope Note suggestions.*

A list of existing vocabularies of particular relevance to ELSST is provided in Appendix 1.

#### 14.3. Addition of new concepts and terms

##### 14.3.1. Criteria for inclusion of concepts

*When a new term is suggested, checks should be made to ensure that its scope does not substantially overlap with an existing concept in the thesaurus, since this would lead to unwanted redundancy within the thesaurus. The source(s) of the term should be referenced in order to establish the exact meaning of the term and determine its place within the thesaurus.*

*Where a new term represents the same concept as an existing Preferred Term, it may be added as a Use For term for that existing term (or substituted for it, if it is a more appropriate term). Otherwise, if the term represents a concept that is new to the thesaurus, it may be added as a new Preferred Term.*

In ELSST, a new concept can be added as a Preferred Term provided that it:

- does not overlap with an existing concept to such an extent that it cannot be clearly distinguished from it
- is not culture-, language- or institution specific

and if any of the following holds:

- it has data to support it
- it is a useful emerging concept (emerging concepts within a specific academic discipline pose provenance issues as they may be terms used by only one researcher which do not gain currency in the literature)
- it is a useful 'bucket' concept to hold other concepts and to ensure structural integrity of the thesaurus, even if the concept itself does not currently have data associated with it. Bucket concepts will require 'do not use' text in the Use Note field and carry a brief definition since they are particularly broad, generic concepts. Because of their breadth, 'bucket' concepts should be located at Level 1 or Level 2 in the hierarchy.

### 14.3.2. Selecting a Preferred Term

*Where a new concept is required, all synonyms for the concept should be considered, and the Preferred Term selected. It is important that the Preferred Term that is chosen:*

- *is current: Currency can be assured by reference to up-to-date subject dictionaries, or, if these are not available, by consulting subject experts.*
- *is unambiguous: Strategies for ensuring this are discussed in Section 14.3.2.1 below*
- *contains no political or moral bias (see Broughton, p.72)*

#### 14.3.2.1. Clarification and disambiguation of thesaurus terms

*In ordinary discourse, a single term might have more than one meaning, In a thesaurus, however, terms in both the source and target languages should be as unambiguous as possible.*

The intended scope of a term is conveyed by the context provided by the hierarchy of broader and narrower concepts linked to the concept in question. ISO 25964-1 p.21 recommends that, if this is insufficient to clarify the intended scope, additional measures should be applied:

- the thesaurus terms (especially Preferred Terms) should be expressed as unambiguously as possible  
  
For example, the multi-meaning term DEPRESSION could be reformulated as ECONOMIC DEPRESSION or METEOROLOGICAL DEPRESSION as appropriate
- a qualifier may be added to the term
- an explicit scope note should be used

### 14.3.3. Selecting Use For terms

*It is useful to add as many Use For terms as possible, to help in information retrieval.*

In ELSST, unlike Preferred Terms, Use For terms may be language- and culture-specific. Note that, like Preferred Terms, they should be expressed as unambiguously as possible (see Section 14.3.2.1.).

Use For terms are particularly important for system-specific hierarchies like educational

levels, social security etc. where the source Preferred Terms have been based on international, standardised classifications (like International Standard Classification of Education (ISCED)). This makes the terms functional cross-nationally but they are not necessarily the terms information seekers would use for searching, therefore Use For terms are important.

For example, in the EDUCATIONAL INSTITUTIONS hierarchy, there is a term HIGHER EDUCATION INSTITUTIONS. Each language will add its country-specific Use For terms for the institutions of that level. In Spanish, for example, terms like ESCUELAS SUPERIORES, ESCUELAS TECNICAS SUPERIORES (ETS), ACADEMIAS may be included, and for French, GRANDES ECOLES, ECOLES NORMALE SUPERIEURES (ENS) and so on.

#### 14.3.4. Selecting Broader Terms, Narrower Terms, and Related Terms

As noted above, ELSST is a symmetrical thesaurus, which means that all language versions share the same structure of hierarchical and associative relationships.

*Care should be taken in a symmetrical thesaurus to ensure that relationships are valid for all languages of the thesaurus.*

*A term that has no BT is referred to as an **orphan term**. Orphan terms are not as useful for browsing, so should be avoided where possible.*

In ELSST, where more than one Broader Term is assigned to a term, they should generally not belong to the same hierarchy.

A term's Related Term may not be its Broader Term, Narrower Term, or **sibling term**, and should preferably not be from the same hierarchy. Related terms should share the same degree of specificity as the term to which they are related.

Reciprocal relationships between Broader Terms and Narrower Terms, and between Related Terms are established automatically by the system.

#### 14.3.5. Selecting Scope Notes

*If necessary, an explicit Scope Note may be added to the term, to help disambiguate it.*

In ELSST, Scope Notes:

- are provided wherever possible in the source language, in order to help both translators and end-users
- must be concise, to the point, and internationally applicable

The following hierarchy of sources should be used for populating the Scope Note field:

1. specialist published dictionaries/encyclopaedias etc.
2. seminal sources in the appropriate discipline that have been identified by subject specialists, including monographs and journal articles
3. other thesauri
4. other sources, including Wikipedia
5. UK Data Archive (i.e. writing the Scope Note from scratch)

Sources (1) and (2) are particularly useful for technical terms in the social sciences. Note, however, that a dictionary entry need not be quoted verbatim.

#### 14.3.6. Scope note source

*A Scope Note Source should only be provided if it defines the term in the same way as the scope note in the thesaurus.*

In ELSST, a Scope Note Source should be provided for all Scope Notes.

#### 14.3.7. Choosing Use Notes

Use Notes in ELSST should:

- be added to a Preferred Term only when necessary
- be reciprocal: when reference is made to another concept within a use note, a reciprocal use note should generally be provided for each concept mentioned.

### 14.4. Deletion of existing concepts

It may be necessary, when revising the thesaurus, to remove a concept. Reasons to delete whole concepts include redundancy (i.e. they overlap with existing concepts in the thesaurus), or if they are felt to be too language-specific. If appropriate, the terms of a deleted concept may become Use For terms of other existing concepts.

Any record dependent on the Preferred Term of a deleted concept (e.g. a catalogue record that has been indexed with the Preferred Term) must be corrected.

### 14.5. Changes to existing concepts and terms

Changes may be made to the Preferred Term, Use For term, note fields, or relationships of an existing concept.

Preferred Terms may require to be changed if they are no longer current, or because they pose problems for translators. Deleted Preferred Terms should normally be retained as Use For terms, unless they are politically incorrect.

Changes to a Preferred Term and/or its scope note often affect its hierarchical and associative relationships, which should be reviewed in the light of its changes.

History Notes need to be updated to reflect all changes. Scope Notes or Use Notes that refer to the Preferred Term should be updated with its new name.

Where a change is made to a Preferred Term, any record dependent on the Preferred Term (e.g. catalogue record that has been indexed with the Preferred Term) must be corrected.

Records may also need to be corrected if a change has been made to a term's Scope Note, as the Preferred Term may no longer be valid.

## PART 5: Translation management

### 15. Pre-translation process

#### 15.1. Overview

The translation of a thesaurus needs to be carefully planned.

This section looks actions that need to be carried out prior to the translation of a thesaurus.

#### 15.2. Preliminary actions

Prior to the translation of ELSST, the following actions need to be carried out.

The ELSST Administrator should:

- provide translators with the requisite guidelines and training
- provide translators with the necessary rights to the thesaurus management system
- provide the concepts that need to be translated or updated
- provide off-line files if requested
- agree a mechanism for providing feedback (see Section 15.3.)
- create placeholders in the management system for a new language
- include, in these ELSST guidelines, strategic decisions made by the Target Language Spokespersons on the use of: singular or plural, diacritics, spelling preferences etc.

The ELSST language spokesperson for the target language should:

- appoint a suitably qualified translator or translators, if required
- ensure that translators are given the requisite guidelines and training
- ensure translators have the necessary rights to the thesaurus management system
- be responsible for the translation quality of their language version
- ensure that the translation is executed in time
- sign off the translation for the next release
- make strategic decisions for their language. This includes use of:
  - singular or plural
  - diacritics
  - spelling preferences
- inform the ELSST Administrator of these strategic decisions so that they can be added to the ELSST guidelines.

The ELSST translator should:

- sign up for ELSST translator training
- apply to the ELSST Administrator for translator rights if they need write access to the thesaurus management system

Where a new variety is to be added to an existing language, the new Target Language Spokesperson should discuss with the Lead Target Language Spokesperson for that language whether to add their terms as UFs to an existing version, or create a separate version of ELSST for their language version.

In ELSST, a list of vocabulary resources that might be useful for the translator is provided in Appendix 1.

### 15.3. Feedback mechanism

*It is likely that during the translation process, translators will encounter problems finding a translation due to errors or inconsistencies in the source language. They need a mechanism for communicating these problems.*

In ELSST, translators can raise any queries they have regarding the translation of terms at any time by (a) adding them to the Suggestions area of the ELSST software (b) raising them as agenda items at quarterly translators' meetings (c) emailing the ELSST Administrator.

### 15.4. Tracking progress

Translators can track the progress of a concept's translation status as they translate. A translation can have one of three statuses:

- Not Started:

Untranslated terms are automatically assigned the status 'Not Started' by the system. 'Not started' terms begin with XX\_, where XX is the 2-letter language code of the target language, and TERM is the source language term. For example, CS\_CAREGIVERS is the untranslated Czech version of the source language term CAREGIVERS.

Once a term's XX\_ prefix is removed, it is considered to be complete, unless marked as 'Incomplete'

- Incomplete: the translator can mark any problematic term, or incomplete translation that they wish to come back to, as 'Incomplete'.

Note: As long as a term contains its XX\_ prefix, it should not be marked as 'Incomplete' – terms should have one Translation status only. Once translators are happy with their translation, they should untick the 'Incomplete' tag.

- Complete: The status 'Complete' is automatically assigned by the system to terms that are neither 'Incomplete' nor 'Not Started'.

An additional note field, 'Supporting data' is available for adding information about the datasets indexed with the term in question, to provide evidence on how the term has been used. This field is not visible to the general public, but is visible to logged-in guest users.

## 16 Translation process

### 16.1. Overview

This section looks at the overall approach to the translation process itself. It begins with a discussion of the overall aims and approach, then discusses some language-specific issues, before listing which elements need to be translated, and how to translate them.

### 16.2. Overall aims and approach

Since multilingual thesauri can be derived by a number of methods, including translation (see Section 5.3), ISO 25964-1 avoids the use of the term 'translation', and talks instead of establishing 'cross-language equivalence' between terms across languages:

*"In a multilingual thesaurus, a bi-directional cross-language equivalence relationship should be shown between preferred terms that represent the same concept in each of the languages" (ISO 25964-1, p.50).*

*In the ideal situation, for each concept a Preferred Term can be found in each of the languages, and these terms are exact equivalents in normal discourse. In reality, this is not*

*always the case, so strategies have to be found for how to preserve one-to-one cross-language equivalence relationships between Preferred Terms in a multilingual thesaurus.*

In ELSST, strategies for translation, i.e. establishing cross-language equivalence, are discussed in Section 16.3 below.

To promote quality control, translators must:

- use the English version of ELSST as the starting point for all ELSST translations. This practice will prevent any potential translation errors from multiplying.
- base their translation on the English concept, as expressed via the Scope Note, or, in the absence of a Scope Note, the English Preferred Term.
- translate Preferred Terms within their hierarchies, since their Broader and Narrower Terms help to elucidate their intended meaning.
- as well as aiming for cross-language equivalence, ensure that they follow the linguistic guidelines in Parts 2 and 4 above, which are valid for all languages.

## 16.3. Some language-specific issues

### 16.3.1. Overview

The following language-specific points should be borne in mind when translating terms.

### 16.3.2. Ambiguity

Ambiguity is language-specific: a term that is not ambiguous in the source language may be ambiguous in the target language. Strategies for how to avoid ambiguity are discussed in Section 14.3.2.1.

### 16.3.3. Singular or plural forms

As noted above in Section 12.4, the choice of whether to use singular or plural forms of nouns is language-specific.

If translating a new language, the target language spokesperson for that language should contact the ELSST Administrator before translation begins to discuss what form of terms to use, so that the ELSST guidelines can be updated. In all cases, rules must be followed consistently within a language version.

Note that, even where two languages adopt the same approach to representing count nouns, it is possible to find cases where a term in the singular in the source language has an equivalent in the plural form in a target language, or vice versa, because a target language equivalent with the same count as the source language is either unavailable or deemed inappropriate. The following examples are from Romanian:

- SL count → TL non-count  
e.g. FACILITIES (plural) → INFRASTRUCTURĂ (singular)
- SL non-count → TL count  
e.g. OVERTIME (singular) → ORE SUPLIMENTARE (plural)

### 16.3.4. Qualifiers

Qualifiers are language-specific and may not require translation into the other language. Conversely, qualifiers may be required in the target language term even if they are not present in the source language term.

## 16.4. What to translate

### 16.4.1. Overview

The sections below cover the translation of concepts, terms and note fields. Since the relational structure in ELSST is the same for all languages, it is translated automatically.

If, when translating, the translator finds that the structure is not appropriate for their language, they should report the problem to the ELSST Administrator, so that the structure and associated concepts can be reviewed.

### 16.4.2. Preferred terms

#### 16.4.2.1. Overview

During preparation of a multilingual thesaurus ISO 25964-1 (p. 51) recognises four degrees of equivalence that are most frequently encountered between source language and target language terms, and which present challenges for finding cross-language equivalents.

These types of equivalence, which should be understood to be points along a spectrum of possibilities rather than as distinct relationship types (ISO 25964-1, p.51), are described below, together with strategies for how to deal with them.

#### 16.4.2.2. Exact equivalence

*Exact equivalence is where “preferred terms are semantically and culturally equivalent” (ISO 25964-1, p.51)). In this case, there is no problem in establishing cross-language equivalence between them.*

#### 16.4.2.3. Inexact or near-equivalence

*Inexact or near-equivalence is where “the terms found in different languages carry a small difference in scope, perhaps due to differences of culture, connotation or appreciation (similar to the case of quasi-synonyms within one language.) If, however, the terms are accepted as being close enough to represent the same concept when used in the thesaurus, then a preferred term for each language should be selected, equivalence between them established, and the scope of the concept understood as including all the cultural variations.” (ISO 25964-1 p.51).*

In ELSST, inexact equivalents can be treated as if they were exact equivalents.

#### 16.4.2.4. Partial equivalence

*Partial equivalence, also known as ‘**broader/narrower equivalence**’ arises when “one language has no term that accurately represents the concept for which a label is sought, but does have a term that is normally considered to represent a broader or narrower concept. If the scope of the concept represented by one term falls completely within the scope of the concept represented by the other, the terms are said to be partially equivalent. If the difference in scope is small enough, it might be acceptable to admit the terms to the thesaurus, treating the two as equivalents that both represent the same concept (similar to the case of specific terms subsumed in a broader concept, within one language.)” (ISO 25964-1, p.51).*

An example in ELSST is:

SCIENCE in English → WISSENSCHAFT in German

Although the German term is broader than the English term, they are treated as equivalent.

Where the target language term chosen is broader/more ambiguous than the source language term, it may be possible to add a qualifier to narrow its meaning.

For example, in ELSST:

ADVICE in English → CONSEIL (AVIS) in French

*ISO 25964-1 (p.51) defines a special case of partial equivalence called “**compound equivalence**” where “one of the languages offers two or more partial equivalents that, in combination, represent the whole of the concept represented by one term in the source language.”*

In ELSST this type of equivalence is handled by making a combined term in the target language that consists of the two target language terms that represent the concept linked by ‘and’ (or more precisely the TL translation of ‘and’). This is more useful for information retrieval than creating an entirely new term, as information seekers would not know the new term and would thus not use it as a search term. Examples in ELSST are:

FUELS in English → COMBUSTIBLES ET CARBURANTS in French

SINGLE-SEX SCHOOLS in English → JUNGENSCHULEN UND MÄDCHENSCHULEN in German

The individual parts of the combined term (in the second example, JUNGENSCHULEN and MÄDCHENSCHULEN) may be added as Use For terms.

#### **16.4.2.5. Non-equivalence**

*Non-equivalence occurs when “no term in one language can be found that even partially or inexactly represents a concept that is needed in the thesaurus” (ISO 25964-1, p. 52). This may arise because either (a) the concept is totally unknown in the target language or (b) the concept is known in the target culture but is not lexicalised.*

In ELSST, the ELSST Thesaurus Management Team has tried to exclude from ELSST very culture-specific terms as preferred terms. However, in rare cases non-equivalence occurs. Translators have several options:

- import the source language Preferred Term into the target language as a loan term, i.e. use the source language term as the target language
- create an equivalent (coined) term in the target language.
- very occasionally, the translation of a source language Use For term can act as the target language equivalent. For example, the term HOMELESSNESS has no direct equivalent in French and Spanish, whereas its non-preferred term HOMELESS PERSONS does. Thus the translation of HOMELESS PERSONS is used as the target language term in French and Spanish.

#### **16.4.2.6. Summary**

Where a suitable term is not available in one language, it may be possible to modify the scope of the concept so that it can be presented by a suitable term in each of the thesaurus languages. Alternatively, as discussed above, an inexact or partial equivalent may be accepted in one language, or a gap in one of the languages may be filled with a loan term or a coined term.

ISO 25964-1 (p. 52) notes that there is rarely a unique correct solution to dealing with equivalence problems, and that it is important to weigh up the merits of the alternatives in each case and consider what is best for most retrieval situations. This is the approach taken in ELSST, what is important is that a source language term and its target language equivalent should allow you to retrieve comparable datasets.

#### **16.4.2.7. Revisions to Preferred Terms**

Where the label of a Preferred Term changes in the source language, its new Preferred Term should be reviewed and retranslated if necessary. As in the source language, the old Preferred Term should normally be retained as a Use For term, unless there is good reason

not to (if, for example it is not politically correct). Where data descriptions have not been updated, due to lack of resources, retaining the old Preferred Term as a Use For term enables users to find the data using the old Preferred Term.

### 16.4.3. Use For terms

*It is not obligatory to translate Use For terms, since Use For terms are language-specific: it may be possible to find Use For terms of target language terms even if the source language term has no Use For terms. Conversely, the source language may have a Use For term where the target language does not need any. Thus the number of Use For terms in the source language and the target language may vary.*

*It is important to add as many Use For terms as possible, in order to improve information retrieval using the thesaurus.*

In ELSST, an example of a source language Use For term that has no target language equivalent in French is

ADOLESCENTS  
UF TEENAGERS

Note: Where translators add a target language Use For term to ELSST that is not an exact synonym of the target language Preferred Term or a translation of a source language Use For term, they should report this to the ELSST Administrator. This Use For term can then be evaluated and suggested to other translators for inclusion in the source language version and translation into other languages, if appropriate.

### 16.4.4. Scope Notes

In ELSST, it is not obligatory to translate Scope Notes, but if they are, they must express the same meaning as the source language scope note (no more and no less). However, they do not need to be a literal translation.

Translators should not add a Scope Note in the target language if there is none in the source language. If they require a Scope Note in the target language, they should contact the ELSST Administrator to request one.

If the source language Scope Note does not work in the target language, the translator should leave it untranslated (marking the translation as 'Incomplete' if the Preferred Term has been translated) and inform the ELSST Administrator, so that they can review it. It is likely that other translators will have had similar problems with the Scope Note, and the Scope Note and/or concept may need to be amended.

### 16.4.5. Scope Note Source

Translation of the source note source is optional.

### 16.4.6. Use Note

Translation of Use Notes is optional.

### 16.4.7. History Note

History Notes should be translated.

## 17 Post-translation process

After the translation process, the linguistic guidelines may need to be updated. Statistics on changes to the translations also need to be recorded for reporting purposes. Where a translator makes changes to the Preferred Term or Scope Note of an existing translation, records indexed with that term may need to be corrected (see Section 14.5 above).

## PART 6: Quality control

### 18. Quality control

#### 18.1. Overview

This section presents a set of quality criteria for ELSST content, with a focus on translation quality. It also contains recommendations for how to achieve these criteria. Quality criteria for thesaurus software or general project management issues are outside the scope of this document.

For a discussion on more general quality criteria for thesauri and other Knowledge Organization Systems, see Soergel (2001).

A quality checklist can be found in Appendix 2.

#### 18.2. Quality criteria for ELSST

*To ensure that a multilingual thesaurus can be applied consistently, regardless of the language of its terms, the thesaurus should meet a set of defined quality criteria.*

The SERISS project recently elaborated a set of translation quality criteria for ELSST as follows (see Balkan (2017):

- Semantic adequacy
- Pragmatic adequacy
- Formal adequacy

Semantic adequacy covers issues such as:

- Are the terms clear and unambiguous?
- Are the terms clearly differentiated from each other?
- Do the target language terms denote the same concept as their source language equivalents?
- Are all concepts language and country-neutral?
- Are the hierarchical relationships valid for all languages?

Pragmatic adequacy:

- Are the Preferred Terms acceptable to the user?

Formal adequacy covers:

- Do the terms and concepts of the thesaurus meet the thesaurus guidelines?

Most of these criteria are equally relevant to the source language content. They have been expanded and presented as a checklist of questions, with suggested responses and corrective actions in Appendix 2. They are intended for translators, but can equally be used for source language content developers. Questions indicated with an asterisk apply to the target language only.

#### 18.3. Quality assurance methods

In order to ensure semantic adequacy in ELSST, the following measures are recommended:

- English is used as the source language for each target language

- translators must be native speakers of the target language
- all translators must follow the ELSST guidelines on how to establish cross-language equivalence and avoid ambiguity
- all new translators must receive translator training before they begin translation
- ideally, the translation will be undertaken by at least two translators

Translations should be reviewed by a subject expert, especially if they have been carried out by only one translator. Pragmatic adequacy is assured by involving subject experts in the term creation process. It is up to the target language spokespersons to decide how and when their translations are reviewed.

Formal adequacy is assured by:

- all translators must follow the ELSST guidelines
- all new translators must receive translator training before they begin translation.

## References

Balkan, Lorna (2017): Report on findings from re-translation of ELSST terms and their use in the CESSDA Portal. Deliverable D3.9 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](http://www.seriss.eu/resources/deliverables)

Broughton, Vanda (2006): Essential thesaurus construction, Facet Publishing, London.

CESSDA-ELSST project. Available at <https://www.ukdataservice.ac.uk/about-us/our-rd/cessda-elsst>

CESSDA-Preparatory Phase Project (CESSDA-PPP). Available at <http://www.data-archive.ac.uk/about/projects/cessda-ppp>

Consortium of European Social Science Data Archives (CESSDA). Available at <https://cessda.net/>

Dansk Sprognævn (Danish Language Council). Available at <https://dsn.dk/om-os/about-the-danish-language-council>

Diccionario de la lengua española. Available at <http://www.rae.es/>

Dicționarul Explicativ al Limbi Române (DEX). Available at <https://dexonline.ro/>

Dictionnaire de français Larousse. Available at <http://www.larousse.fr/dictionnaires/francais>

Duden Online-Wörterbuch. Available at <http://www.duden.de/woerterbuch>

Hedden, Heather (2010): The accidental taxonomist, Medford, New Jersey. Information Today, Inc.

Humanities and Social Science Electronic Thesaurus (HASSET). Available at <https://hasset.ukdataservice.ac.uk/>

International Standard Classification of Education (ISCED). Available at <http://uis.unesco.org/en/topic/international-standard-classification-education-isced>

ISO 25964-1 (2011): Information and documentation -- Thesauri and interoperability with other vocabularies, Part 1: Thesauri for information retrieval. Available at <http://www.niso.org/schemas/iso25964/#schema>

Language Independent Metadata Browsing of European Resources (LIMBER) project. Available at <http://www.data-archive.ac.uk/about/projects/limber>

Λεξικό της κοινής νεοελληνικής (Modern Greek Dictionary). Available at [http://www.greek-language.gr/greekLang/modern\\_greek/tools/lexica/triantafyllides/](http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/)

Oxford English Dictionary. Available at <http://www.oed.com/>

MOT Kielitoimiston sanakirja

Multilingual Access to Data Infrastructures of the European Research Area (MADIERA). Available at <http://www.data-archive.ac.uk/about/projects/madiera>

Slovenski pravopis. Available at <http://www.fran.si/iskanje?FilteredDictionaryIds=134&View=1&Query=%2A>

Slovník spisovného jazyka českého. Available at <http://ssjc.ujc.cas.cz/>

Soergel, Dagobert (2001): Thesauri and ontologies in digital libraries, presented at JCDL 01, Roanoke, VA, June 2, pp.105-116. Available at <http://www.dsoergel.com/cv/B63.pdf>

Svenska Akademiens ordlista.

Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS). Available at <http://seriss.eu/>

UK Data Archive. Available at <http://www.data-archive.ac.uk/>

UK Data Service. Available at <https://www.ukdataservice.ac.uk/>

World Wide Web Consortium (W3C). Available at <https://www.w3.org/>

## Other references consulted

Aitchison, Jean, Gilchrist, Alan and Bawden, David (2000): Thesaurus Construction and Use: A Practical Manual, Fourth edition, Aslib IMI, The Association for Information Management, London.

Art & Architecture Thesaurus (AAT), Editorial Guidelines (2017): the J. Paul Getty Trust. Available at

[http://www.getty.edu/research/tools/vocabularies/guidelines/aat\\_1\\_contents\\_intro.html](http://www.getty.edu/research/tools/vocabularies/guidelines/aat_1_contents_intro.html)

Guidelines for the management and translation of SNOMED CT® Version 2.02 (2012): International Health Terminology Standards Development Organisation. Available at

<http://www.snomed.org/snomed-ct/snomed-ct-worldwide/translations-of-snomed-ct>

Lancaster, F.W. (2003): Indexing and Abstracting in Theory and Practice, 3<sup>rd</sup> edition, Facet, London.

Ryan, Catherine (2014): Thesaurus construction guidelines: An introduction to thesauri and guidelines on their construction, Dublin: Royal Irish Academy and National Library of Ireland.

Available at <http://apps.dri.ie/motif/docs/guidelines.pdf>

Shiri, AI (2012): Powering Search: The Role of Thesauri in New Information Environments, Information Today, Inc., [2012].

## Appendix 1: Vocabulary resources for translation

Examples of multilingual thesauri:

Name of thesaurus	Subject area	Languages	URL
Finnish thesaurus and ontology service (Finto)	general	Finnish, Swedish, English	<a href="https://finto.fi/en/">https://finto.fi/en/</a>
UNESCO	social sciences	English, French, Spanish	<a href="http://databases.unesco.org/thesaurus/">http://databases.unesco.org/thesaurus/</a>
Statistics Canada Thesaurus (STATCAN)	social sciences	English, French	<a href="http://www47.statcan.gc.ca/th_r000-eng.htm">http://www47.statcan.gc.ca/th_r000-eng.htm</a>
OECD	social sciences	English, French, Spanish	<a href="http://bibliotecavirtual.clacso.org.ar/ar/oecd-macroth/">http://bibliotecavirtual.clacso.org.ar/ar/oecd-macroth/</a>
EUROVOC Multilingual Thesaurus of the European Union	general	EU languages	<a href="http://europa.eu/eurovoc/">http://europa.eu/eurovoc/</a>
United Nations Bibliographic Information System Thesaurus (UNBIS)	general	English, French, Spanish, Russian, Chinese, Arabic	<a href="http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf">http://lib-thesaurus.un.org/LIB/DHLUNBISThesaurus.nsf</a>
General Multilingual Environmental Thesaurus (GEMET)	environment	37 languages, including:  Czech, Danish, English,	<a href="http://www.eionet.europa.eu/gemet/about">http://www.eionet.europa.eu/gemet/about</a>

		Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Spanish, Norwegian, Swedish	
European Multilingual Thesaurus on Health Promotion	health	Dutch, English, French, Italian, Portuguese, Spanish	<a href="http://www.vocabularyserver.com/hpmulti/en/index.php">http://www.vocabularyserver.com/hpmulti/en/index.php</a>
Learning Resources Exchange for schools thesaurus (LRE thesaurus)  (Formerly the ETB thesaurus))	education	Languages include:  Czech, Danish, English, Finnish, French, German, Greek, Norwegian, Spanish, Swedish	searchable via:  <a href="http://europeanschoolnet-vbe.lexaurus.net/vbe/(select%20'LRE%20thesaurus%20in%20Advanced%20search)">http://europeanschoolnet-vbe.lexaurus.net/vbe/(select 'LRE thesaurus in Advanced search)</a>  browsable via:  <a href="http://europeanschoolnet-vbe.lexaurus.net/vbe/browse">http://europeanschoolnet-vbe.lexaurus.net/vbe/browse</a>
Thesaurus for Education Systems in Europe (TESE)	education	Available in 15 languages, including:  Czech, English, Finnish, French, German, Greek, Lithuanian, Romanian, Spanish	Downloadable from  <a href="http://eacea.ec.europa.eu/education/eurydice/tes_e_en.php">http://eacea.ec.europa.eu/education/eurydice/tes_e_en.php</a>

POPIN	population	English, French, Spanish	<a href="http://www.cicred.org/OLD2004/thesaurus/INTEGRAL/index.html">http://www.cicred.org/OLD2004/thesaurus/INTEGRAL/index.html</a>
ILO Thesaurus	work, employ- ment	English, French, Spanish	<a href="http://www.ilo.org/thesaurus/defaulten.asp">http://www.ilo.org/thesaurus/defaulten.asp</a>
European Training Thesaurus	education	11 languages including:  Danish, English, Finnish, French, German, and Swedish	Downloadable from <a href="http://www.cedefop.europa.eu/en/publications-and-resources/publications/4114">http://www.cedefop.europa.eu/en/publications-and-resources/publications/4114</a>
European Education Thesaurus	education, learning	Languages include:  English, French, German, Swedish	<a href="http://vocabularyserver.com/tee/en/">http://vocabularyserver.com/tee/en/</a>
AGROVOC	agriculture forestry, fishery, food and related (e.g. environ- ment)	Available in 23 languages including:  Czech, English, French, German, Spanish	<a href="http://artemide.art.uniroma2.it:8081/agrovoc/agrovoc/en/">http://artemide.art.uniroma2.it:8081/agrovoc/agrovoc/en/</a>
CAB	Life sciences	equivalents for most terms in: English, Dutch,	<a href="http://www.cabi.org/cabthesaurus/">http://www.cabi.org/cabthesaurus/</a>

		Portuguese Spanish  plus lesser content in: Danish, Finnish, French, German, Italian, Norwegian Swedish	
Government of Canada (GOC) Core Subject Thesaurus	Social sciences	English, French	<a href="http://www.thesaurus.gc.ca/recherche-search/thes-eng.html">http://www.thesaurus.gc.ca/recherche-search/thes-eng.html</a>

#### Multilingual glossaries, terminologies, and vocabulary repositories

Name	Resource type	Subject area	Languages	URL
InterActive Terminology for Europe (IATE)	term bank	general	EU languages	<a href="http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load">http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load</a>
European Glossary on Education	glossary.  Also information on institutions by educational level in each country	education	EU languages	Downloadable from <a href="http://bookshop.europa.eu/en/european-glossary-on-education-pbEC3212292/">http://bookshop.europa.eu/en/european-glossary-on-education-pbEC3212292/</a>
Multilingual glossary of technical and popular medical terms	glossary	medicine	Danish, Dutch, English, French, German, Italian,	<a href="http://users.ugent.be/~rvdstich/egloss/welcome.html">http://users.ugent.be/~rvdstich/egloss/welcome.html</a>

			Portuguese, Spanish	
European Employment and Industrial Relations (EMIRE)	dictionary. Explains the national industrial relations systems of the EU member states through their terminology.	labour and employment	Languages include:  Czech, Danish, English, Finnish, French, German, Greek, Lithuanian, Norwegian, Romanian, Spanish, Swedish	<a href="http://www.eurofound.europa.eu/emire/emire.htm">http://www.eurofound.europa.eu/emire/emire.htm</a>
International Statistical Institute (ISI) glossary	glossary	statistics	29 languages	<a href="http://isi.cbs.nl/glossary/">http://isi.cbs.nl/glossary/</a>
Reference And Management Of Nomenclatures (RAMON)	metadata server	multiple	multiple	<a href="http://ec.europa.eu/eurostat/ramon">http://ec.europa.eu/eurostat/ramon</a>
The Basel Register of Thesauri, Ontologies & Classifications (BARTOC)	database of Knowledge organization systems	multiple	multiple	<a href="http://www.bartoc.org/">http://www.bartoc.org/</a>

#### Parallel corpora

Name	Resource type	Subject area	Languages	URL
Reverso	parallel corpora	all	multiple	<a href="http://context.reverso.net/translation">http://context.reverso.net/translation</a>

Linguee	parallel corpora	all	multiple	<a href="http://www.linguee.com">www.linguee.com</a>
Eur-lex	parallel corpora	law and related	EU languages	<a href="http://eur-lex.europa.eu/advanced-search-form.html">http://eur-lex.europa.eu/advanced-search-form.html</a>

## Appendix 2: Quality criteria for ELSST

Note: Questions marked with an asterisk apply to the target language only

Question ID	Question	Relevant section in present guidelines	Response/corrective action
<b>Preferred Term</b>			
1	Is the concept described by the term culture-, language- and institution neutral?		Report to the Administrator. The Administrator will review the concept.
2	Is the term:	current?	
3		distinct from other concepts in the thesaurus?	If there is an overlap of concept with another Preferred Term in the thesaurus, report to the Administrator. The Administrator will review the concept.
4		clear and unambiguous?	14.3.2.1 . If the target language term is ambiguous, update the term. If source language term is ambiguous, report to the Administrator. The Administrator will review the concept.
5		suitable as an indexing term?	
6	*Does the term express the same concept as source language Preferred Term?	16	In cases of non-equivalence, translator should leave the term untranslated, and report to the Administrator. Administrator will review the concept.
7	Does the term comply with ELSST guidelines with respect to:	spelling?	12.8.
8		punctuation?	12.9.
9		use of accents and diacritics?	12.6.
10		acronyms and abbreviations?	12.3.
11		plurality?	12.4.
12		qualifiers?	12.5.
<b>Use For term</b>			
13	Is the term clear and unambiguous?	14.3.2.1 .	
14	*Does the term correspond semantically to the Preferred Term or Use For term in the source		Report to the Administrator. The Administrator will review the Use For term for possible inclusion in the source language.
15	Does the term comply with	plurality?	12.4.
16	ELSST guidelines with respect	qualifiers?	12.5.
<b>Relationships</b>			
17	Are the following relationships valid for your language:	Broader Terms?	Report to Administrator. The Administrator will review the relationship.
18		Narrower Terms?	Report to Administrator. The Administrator will review the relationship.
19		Related Terms?	Report to Administrator. The Administrator will review the relationship.
<b>Scope Note</b>			
20	Is the Scope Note clear and precise?		If the source language Scope Note is not clear or precise, report the problem to the Administrator. The Administrator will review the Scope Note.
21	*Does the Scope Note:	express the same concept as the source language Scope Note?	16.4.4. If the source language Scope Note is not suitable for the target language, leave untranslated and report the problem to the Administrator. The Administrator will review the Scope Note.
22		have a source language equivalent?	16.4.4. Ask the Administrator for a source language Scope Note.
<b>Use Note</b>			
23	Is the Use Note clear and precise?		If the source language Use Note is not clear or precise, report the problem to the Administrator. The Administrator will review the Use Note.
24	Is there a reciprocal Use Note for all concepts referred to in the Use Note?		

Figure 6: Quality criteria for ELSST

## 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.

**acronym:** An acronym is an abbreviation formed from the initial letters of other words and pronounced as a word (e.g. EU).

**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.

**CESSDA:** CESSDA is a consortium of European social science data archives that provides large scale, integrated and sustainable data services to the social sciences.

**compound term:** A compound term is a term that can be split morphologically into separate components.

**concept:** A concept is a unit of thought.

**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.

**corpus:** A corpus is a collection of written texts.

**count noun:** A count noun is a noun that denotes a countable thing. In English, count nouns can form a plural and can be used with the indefinite article. An example of a count noun is 'book'.

**cross-language equivalence:** In a multilingual thesaurus, cross-language equivalence exists between **Preferred Terms** that are used to represent the same **concept** in the different languages (ISO 25964-1).

**descriptor:** see **Preferred Term**.

**diacritic:** A diacritic is a sign, such as an accent, which, when written over, under, or through a letter of the alphabet (e.g., the umlaut in German, such as the "ö in 'Köln'), indicates a difference in pronunciation from the same letter without the diacritic.

**equivalence relationship:** An equivalence relationship exists between two **terms** in a **thesaurus** that both represent the same **concept**. In ordinary discourse, terms that are **quasi-synonyms** may represent slightly different concepts. After inclusion in the thesaurus, however, the equivalence clarifies that both are regarded as representing the same concept. When two or more such terms are in the same language within a monolingual or multilingual thesaurus, one of them is designated a **Preferred Term**, and the other(s) **Non-Preferred terms**; when two or more such terms are in the different languages of a multilingual

thesaurus, each of them may be a Preferred Term in its own language respectively, and the relationship is known as **cross-language equivalence** (ISO 25964-1, p.4).

**Globally Unique Identifier (GUID):** A Globally Unique Identifier is a type of **Unique Identifier**.

**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.

**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).

**indexing:** Indexing is the process of evaluating information and assigning **index terms** to it.

**interoperability:** Interoperability, in the context of **controlled vocabularies**, is the ability of two or more vocabularies to exchange information, for example by mapping to each other, in order to enhance discovery.

**Knowledge Organization System (KOS):** Knowledge Organization System encompasses all types of schemes for organizing information and promoting knowledge management. Examples include classification schemes, subject headings, thesauri, and ontologies.

**lexical variant:** A lexical variant is a term that is a different word form for another term, due to spelling differences, grammatical variation, or abbreviations.

**Linked Data:** Linked Data is a pattern for hyperlinking machine-readable data sets to each other using Semantic Web techniques, especially via the use of **RDF** and **URIs**.

**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**.

**non-count noun:** A non-count noun is a noun that denotes a thing that cannot be counted. In English, a non-count noun cannot form a plural or be used with the indefinite article. An example of a non-count noun is 'poverty'.

**non-equivalence:** Non-equivalence, in the context of a multilingual thesaurus, is where no **term** in one language can be found that even partially or inexactly represents a **concept** that is needed in the thesaurus.

**Non-preferred term:** see **Use For term**

**orphan term:** An orphan term is a **Preferred Term** that has no **hierarchical relationships**.

**Persistent Identifier:** A Persistent Identifier is a long-lasting reference to a digital resource. Typically, it has two components: a **Unique Identifier**, and a service that locates the resource over time even when its location changes.

**polyhierarchical structure:** A polyhierarchical structure is where a **concept** can have more than one broader concept.

**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.

**qualifier:** A qualifier is a word or phrase in parentheses following a term used to distinguish a term in a vocabulary from otherwise identical terms that have different meanings.

**Quasi-synonym:** See **near synonym**

**query expansion:** Query expansion is the process of supplementing the original query with additional terms in order to improve retrieval performance.

**reciprocal relationship:** A reciprocal relationship exists between **concepts** in a **controlled language** where a relationship in one direction is matched by the same or different relationship in the reverse direction. For example, a reciprocal relationship exists between a **Broader Term** and its **Narrower Term (BT/NT)**, or between two **Related Terms (RT/RT)**.

**Related Term (RT):** A Related Term is a **Preferred Term** representing a **concept** that has an **associative relationship** with the one in question.

**Resource Description Framework (RDF):** Resource Description Framework is a model for encoding semantic relationships between items of data so that these relationships can be interpreted computationally.

**root term:** see **top term**

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

**Semantic Web:** The Semantic Web is part of the World Wide Web that consists of machine-readable data in **RDF** and ability to query that information in standard ways (e.g. via **SPARQL**)

**sibling term:** Sibling terms are **Preferred Terms** that share the same **Broader Term**.

**Simple Knowledge Organization System (SKOS):** SKOS is a common data model for **Knowledge Organization Systems (KOSs)**. Using SKOS, a KOS can be expressed as machine-readable data.

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

**SPARQL:** SPARQL is a semantic query language for databases that is able to retrieve and manipulate data stored in **RDF** format.

**symmetrical:** Symmetrical, in the context of a multilingual **thesaurus**, means that all language versions share the same structure of **hierarchical relationships** and **associative relationships**.

**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.

**term:** A term, in a thesaurus, represents a **concept**. Terms may be either **Preferred Terms** or **Use For terms**.

**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 **Preferred Term** representing a **concept** that has no broader concept in the thesaurus. A Top Term is sometimes referred to as the 'root term'. In ELSST, a Top Term is the broadest term in a **hierarchy**, and is also used as the hierarchy name.

**Uniform Resource Identifier (URI):** A Uniform Resource Identifier is a string of characters used to identify a resource. The most common form of URI is the **Uniform Resource Locator**.

**Uniform Resource Locator (URL):** A Uniform Resource Locator is a reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it. It is commonly referred to a web address.

**Unique Identifier:** A Unique Identifier is a number or other string that is associated with a record or piece of data, exists only once in a database, and is used to uniquely identify and disambiguate that record or piece of data from all others in the database.

**upward posting:** An upward posting is a **Use For term** that is a narrower term where its broader term functions as its **Preferred Term**, rather than its **Narrower Term**. Upward posting is a technique used to reduce the number of Preferred Terms in a thesaurus.

**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**.

## Abbreviations and acronyms

<b>Abbreviation</b>	<b>Meaning</b>
BT	Broader Term
CESSDA	Consortium of Social Science Data Archives
CESSDA-PPP project	CESSDA-Preparatory Phase Project
ESRC	Economic and Social Research Council
EU	European Union
GUID	Globally Unique Identifier
HASSET	Humanities and Social Science Electronic Thesaurus
HN	History note
KOS	Knowledge Organization System
LIMBER project	Language Independent Metadata Browsing of European Resources
MADIERA project	Multilingual Access to Data Infrastructures of the European Research Area
NT	Narrower Term
PT	Preferred Term
RDF	Resource Description Framework
RT	Related Term
SERISS project	Synergies for Europe's Research Infrastructures in the Social Sciences
SKOS	Simple Knowledge Organization System
SL	Source Language
SN	Scope Note
SNS	Scope Note Source

TL	Target Language
UF	Use For term
UN	Use note
URI	Uniform Resource Identifier
URL	Uniform Resource Locator