



Title	Recommendations and Guidelines Report (D4.4)
Work Package	WP4
Authors	Taina Jääskeläinen, Lorna Balkan
Source	
Date	9 March 2010
Dissemination Level	PU (Public)

Summary/abstract

Background

The recommendations and guidelines have been made on the assumption of their being a new legal entity (referred to in this report as the cessed-ERIC) which will replace the existing CESSDA organisation and which will include existing members of CESSDA and potentially, new member organisations whose languages have not already been included in the thesaurus.

Chapter 1

Source language of the CESSDA multilingual thesaurus ELSST;
 Use and licensing of the thesaurus;
 Obligations of cessed-ERIC members regarding the thesaurus;
 Management and maintenance structure of the thesaurus;
 Protection of its IPR;
 Access to the thesaurus.

Chapter 2

Maintenance and management of the Topics Classification.

Chapter 3

Extension of the thesaurus to new languages.

Chapter 4

CESSDA multilingual thesaurus and international standard classifications.

Chapter 5

Translation of key DDI elements of catalogue records published in the cessed-ERIC portal into English.

Chapter 6

Other, related thesauri work.

1. CESSDA multilingual thesaurus

1.1 Source language

CESSDA has agreed that the source language for the thesaurus will be English and that all candidate concepts and discussions will be submitted and recorded in English.

1.2 Use and licensing of the multilingual thesaurus

The cessda-ERIC should hold the copyright of the multilingual thesaurus and all its language versions. Only the cessda-ERIC should issue licences for the thesaurus.

The thesaurus should form a single product, complete with all language versions. Licences will be issued only for this product.

CESSDA members can freely use the thesaurus for indexing and as a search tool within the cessda-ERIC and their own local systems. The thesaurus will be freely available to the general public for browsing within the Data Portal.

A fee should be charged to cover at least the administrative costs if the licence is issued to a non-commercial organisation which is not a CESSDA member. An annual administrative fee will be charged if the licensee wants the annual updates of the thesaurus.

Commercial opportunities can be investigated at a future stage.

1.3 Obligations and recommendations regarding the thesaurus

1.3.1 Obligations of cessda-ERIC members

For each language version there will be one designated CESSDA member who is responsible for it. This member may share the translation task and/or use peer reviews but will bear overall responsibility for ensuring that the language version is produced and maintained within the stipulated time frame.

Full members of the cessda-ERIC shall ensure that all terms and keywords are translated into agreed native language(s) within nine months of entering full membership. Associate members shall present a translation plan and timetable, and the translation must be finalised within a maximum of 5 years.

The organisations responsible for a language version shall ensure that the language version is maintained by translating the changes and new concepts annually within the maintenance cycle mentioned in 1.4.1. If changes have not been made for three years, the cessda-ERIC will take the agreement up for review. The changes made to the source thesaurus during the Madiera and CESSDA PPP projects shall be translated by the end of September of the year the cessda-ERIC is established.

The organisations responsible for a language version shall ensure that the copyright of the language version is assigned to the cessda-ERIC - if necessary, by means of separate agreements with the initial copyright owners (translators). WP4 recommends that the cessda-ERIC provides lawyer-reviewed templates for these agreements, particularly for situations where translators are staff members.

The organisations responsible for a language version shall ensure that the translation follows thesaurus translation guidelines provided by the cessda-ERIC.

The organisations responsible for a language version shall nominate a representative to the Thesaurus Management Team, should it be requested by the cessa-ERIC.

1.3.2 Recommendations for best practices

Translation workshops should be organised by the cessa-ERIC, and it is strongly recommended that translators attend before beginning their work.

It is recommended that member organisations use the thesaurus for indexing data published in the CESSDA portal and that they change the keywords in their data descriptions accordingly after each official version. Indexing data with CESSDA thesaurus terms may become mandatory at a later date, particularly at question and variable level, to meet the requirements of the Question Bank and the Harmonisation tool (3CDB).

Local extensions are allowed. Local terms can be used in searches but they will not be visible in the thesaurus tree structure. It is recommended that local extension managers review which local terms might be useful to add to the multilingual thesaurus, adding them as candidate terms (in English) to the Comments database. Criteria for reviewing the usefulness of a term:

- 1) The concept is not covered by an existing term in the multilingual thesaurus;
- 2) The concept is not too culture-specific or system-specific.

1.4 Management and maintenance structure of the multilingual thesaurus

1.4.1 Management Team, Thesaurus Officer, and the maintenance cycle

Management Team membership

The Management Team consists of up to 5 CESSDA members. Elections will take place every 3 years, with nominations reviewed by CESSDA Management Board/ERIC Directorate based on the spread of languages and expertise. If there are not enough volunteers, members will be nominated on a rotation basis. UKDA will be the administrator of ELSST and will always be represented in the team.

Meetings

There will be virtual meetings held as often as needed. The Management Team will meet physically once a year.

Management Team tasks

Review overall coverage of concepts;
Review candidate terms;
Review logs of most frequent failed search terms and create drafts of these;
Review structural change suggestions;
Decide position of concept within hierarchical structures;
Decide on the relationships to other thesaurus concepts;
Create definitive expression of concepts (ambiguity scope notes);

Oversee translation procedures;
Release official versions including new languages;
Review functionality within portal, publisher, question bank;
Publish and maintain guidelines for use;
Provide training;
Promote usage and uptake - within the CESSDA organisations and contribute to promotion to organisations outside of CESSDA.

Thesaurus Officer tasks

(The cessda-ERIC may decide to transfer some of the tasks below to Thesaurus Management Team leader)

Leadership of the management team;
Preparatory work for the management team tasks (research into terminology, identify useful sources, suggest comprehensive scope notes, create drafts of amendments based on the comments database and logs of failed search terms);
Organisation of management team meetings;
Recording of decisions;
Organisation of each stage of the annual maintenance cycle;
Introduction of agreed amendments into the source thesaurus;
Maintenance and coordination of comments on thesaurus terms;
Responding to queries about the thesaurus and related matters;
Production of the reports required by the cessda-ERIC.

Annual maintenance cycle

Candidate concepts and structural change suggestions are added to the comments database.

Entries in the comments database at the 31st December (cut-off point) will be dealt with by the Management Team according to the timetable below. Any items entered after this cut-off date will be dealt with in the following review cycle (i.e. at the end of the following year).

Timetable:

Months 1-3: Management Team reviews suggestions for candidate terms, changes, and logs of most frequent failed search terms and creates drafts of these.

Month 4: Management Team drafts will be disseminated to CESSDA members for comments.

Month 5-6: Management Team review comments and make final decisions, agreed changes and additions are then implemented in the ELSST source thesaurus.

Month 7-9: CESSDA member organisations translate changes and new concepts.

Month 10-11: Documentation and IT tasks, made centrally by UKDA.

30th November Official release of new version.

Resources required

Thesaurus officer and management team leader tasks require in total 7 person months (PMs) annually. The division of PMs between these two roles depends on the division of work, and include training provided by them. It is assumed that the work done by

other management team members (2-3 weeks a year) will be resourced by their parent organisations.

Thesaurus maintenance and management software system was developed during the CESSDA-PPP. The software requires small-scale further development, altogether 1 PM in the first year of the cessda-ERIC, to guarantee easy usability and to add the reporting tool for annual reports. After that, the maintenance requires about 0.5 PM annually.

1.4.2 Procedures for new concepts and structural changes

New concepts

New concepts will be offered for consideration by individual organisations, using the comments database. In addition to the native language, the suggestions should also be presented in English, and should include a suggested scope note (unless there is clearly no need), and suggested position within a hierarchy plus Broader Term relationship to an existing term.

Synonyms (UF terms) can also be suggested.

All suggestions will be approved or rejected by the management team, and drafts offered for comments to other CESSDA members. (see timetable at 1.4.1)

UKDA (as the English language expert) should be responsible for the final wording of the terms and the ambiguity scope notes in the source thesaurus.

Edits to existing terms, removal of terms, structural changes

Edits and removal of existing terms and structural changes will be offered for consideration by individual organisations, using the comments database. Suggestions should be in English. These will be approved or rejected by the Management Team, and drafts offered for comments to other CESSDA members (see timetable at 1.4.1). Obsolete terms will remain in the thesaurus but will not be available for indexing.

1.4.3 Release of new languages

A (new) language can be included in the official release even if not all concepts have been translated at the time of the release. Any concepts not translated will appear in the English, preceded by the relevant two letter ISO country code.

1.5 Protection of the thesaurus IPR

The thesaurus is a database

The CESSDA multilingual thesaurus is a database as there has been substantial investment in obtaining and presenting its content. The thesaurus database content is protected by the EU directive 96/9/EC on the legal protection of databases (Chapter III, sui generis right, similar to copyright). The cessda-ERIC should become the rightholder of the present thesaurus, all its language versions and its future versions. The copyright of the previous versions will not be clarified. The right provided by the directive runs

for fifteen years. There will be a new official release of the thesaurus database each year, with amended content. The protection period of 15 years starts from each release.

Lawyer's advice: To ensure the maintenance of copyright with CESSDA, it would be best if the decision-makers, especially regarding the selection and arrangement of the material for the thesaurus, are employed by the cessa-ERIC.

IPR of related thesauri

The CESSDA thesaurus was originally based on HASSET which was originally based on a paper version of the UNESCO thesaurus 20 years ago. UNESCO no longer holds any copyright in either HASSET or CESSDA thesaurus since both have undergone extensive changes since then. UKDA holds the copyright of HASSET.

Thesaurus database components

The cessa-ERIC should become the rightholder of all components of the database (indexing and searching tool, the maintenance interfaces, the comments database, and user guidelines).

Lawyer's advice: It would be wise to have a separate contract for authorised users, stating their rights and duties expressively.

1.6 Access to the thesaurus

Access to the web maintenance of the ELSST thesaurus will be via authenticated login name and password for all staff at organisations of the cessa-ERIC and the available functionality will be dependant on their level of authorisation. All other users will be asked to login using their email address.

The available functionality will consist of the:

- a) Addition, editing and deletion of ELSST concepts;
- b) Addition, editing and deletion of scope notes;
- c) Addition, editing and deletion of synonyms;
- d) Addition and deletion of the relationships between concepts;
- e) Translation of concepts, scope notes and synonyms;
- f) Maintenance and management of local extensions;
- g) Review of next version changes and local extensions;
- h) Suggestion of candidate concepts and structural changes

The levels of authorisation are as follows:

- a) Level 1: (General usage – non cessa-ERIC)
Only the present version of ELSST will be available and only suggestions for new concepts can be made.
- b) Level 2: (cessa-ERIC – including indexers)
Present, Next and Local Extensions available but only suggestions for new concepts can be made.
- c) Level 3: Reviewers

Present, Next and Local Extensions available and suggestions for new concepts can be made. Has access to comments interface to also suggest Scope notes, synonyms and structural changes.

d) Level 4: Translator

Present, Next and Local Extensions available and suggestions for new concepts can be made. Has access to comments interface to also suggest Scope notes, synonyms and structural changes. For their specific language they also have the ability to edit an ELSST concept term and to edit, add or delete scope notes and synonyms.

e) Level 5: Local Manager

All the rights of a translator plus the ability to add, edit and delete Non-ELSST concepts to their specific local extension and to add and delete relationships to non-ELSST concepts with any other concept, either local or ELSST.

f) Level 6: ELSST Manager

All the rights of a translator plus the ability to add, edit and delete ELSST concepts and to add and delete relationships between ELSST concepts in the underlying English version. Also have the ability to change the content and status of the suggestions made via the comments interface.

g) Level 7: Administrator

All the rights of ELSST and every Local Manager.

2. Maintenance and management of the CESSDA topics classification

Topics classification will be maintained in the same way as the thesaurus. The Thesaurus Management Team will also function as the management team of the topics classification, and the same maintenance procedures will be followed.

3. Extension of the thesaurus

The statutes of the cessa-ERIC state that full members shall maintain their local language(s) within the multi-lingual thesaurus. This report specifies in 1.3.1. that for each language version there is one designated CESSDA member who is responsible for it. Full members of the cessa-ERIC shall ensure that all terms and keywords are translated into agreed native language(s) within nine months of entering full membership. Associate members shall present a translation plan and timetable, and the translation should be finalised within 5 years at maximum.

At present, the multilingual thesaurus contains the following language versions: English (source thesaurus), Danish, German, Spanish, French, Finnish, Greek, Norwegian and Swedish. German and Spanish language versions are not complete, and the non-translated concepts and scope notes should be added.

It is estimated that the translation of ELSST into a new language requires seven PMs. This estimate is based on the experiences gained from a prior EU project called Madiera.

Should the cessa-ERIC include full members whose native language is not one of the above mentioned languages, these language versions shall be added to the thesaurus (this may concern, for example, Hungarian, Czech, Slovenian, Italian and Romanian). Associate members shall present a translation plan and timetable.

In principle, cessda-ERIC members are expected to fund the translation of the thesaurus into their local language with national funding. There may be cases where the cessda-ERIC deems it beneficial to wholly or partly fund the translation into a member language or into some other language mentioned in the next paragraph.

Looking at the roadmap for widening the cessda-ERIC (WP7), the cessda-ERIC may consider whether funding can be found to add Polish and the widely spoken Russian to the thesaurus. On the global level, French and Spanish which are spoken in several countries outside Europe, are already included. Portuguese is the first language of seven countries, and one of the official languages in two others, which makes it a feasible candidate language. The inclusion of other major languages Chinese, Japanese and Arabic (spoken in several countries) would ensure easier access to European social science data resources to researchers speaking those languages.

If any extensions are done with cessda-ERIC funding, European languages, including Russian, would be the first priority.

4. International standard classifications and the multilingual thesaurus

During the CESSDA PPP, WP4 amended a number of hierarchies in the multilingual source thesaurus to make the terms in these hierarchies more functional for comparative research in a cross-national setting. The aim has been to base the terminology on international standard classifications, such as the International Standard Classification of Education (ISCED). The amended hierarchies include Educational institutes, Educational levels, Offences, Fields of study, Marital status, Social security. The amendments will facilitate harmonisation of metadata.

WP4 also made decisions on the new concepts offered for consideration.

The amendments, new concepts and scope notes added to the source thesaurus during CESSDA PPP will require that members responsible for a language version shall translate them into their language at the latest by the end of September of the year the cessda-ERIC is established.

5. Translation of key DDI elements of CESSDA catalogue records into English

5.1 Obligations of cessda-ERIC members

WP4 proposes that Resource Providers shall provide at least minimum study descriptions in English for all new data they publish in the CESSDA portal from the point of entering Full Membership. A transition period of six months may be considered if needed. Associate members should provide the minimum study description within two years of entering Associate Membership.

5.2 Recommendations regarding study description in English

It is recommended that, in due course, members provide at least minimum study descriptions in English for previously published data. WP4 recommends that the cessda-ERIC provides a documentation tool to facilitate multilingual study descriptions.

The minimum English study description includes the following elements:

- a. Title;
- b. Abbreviated abstract;
- c. Universe (at study level);
- d. Geography (nation, ISO code);
- e. Time period (date);
- f. Mode of collection;
- g. CESSDA topics classification;
- h. CESSDA thesaurus terms for indexing;
- i. Sampling procedure (probability, non-probability or no sampling).

The elements a-c should be translated by the Resource Provider. It is expected that the translation of elements d.-i. can be produced automatically through the use of controlled vocabularies and the documentation tool.

When the cessda-ERIC metadata model is finalised, it may become mandatory to provide more detailed geographical information in a standardised format.

5.3 Reasons behind the WP4 proposal

In a survey among CESSDA federation members in March 2009, more than 90% of the members whose local language is not English reported their intention to provide at least minimum study descriptions in English for data they publish in the portal.

In fact, most members intend to provide more than the minimum. One third said they would provide a full study description in English, equivalent to the one in the local language. A quarter aimed to provide a study description with more elements than the required minimum. No respondent supported the option that the cessda-ERIC would neither recommend nor require the minimum study description in English.

While the ideal would be a full study description both in the local language(s) and in English, this is not possible for many members because of resource implications. The use of the multilingual thesaurus as a search/recovery tool in the data portal will yield search results in languages that the searcher does not know. The goal of the minimum English study description is to provide enough information to the searcher to know whether it is worthwhile to contact the cessda-ERIC member directly for more information about the dataset.

6. Other related thesauri work

WP4 found that there does not seem to be any other, related thesauri work ongoing in other ESFRI projects. Should any such thesauri work appear within humanities and social science ERICs in the future, the cessda-ERIC should welcome co-operation.

WP4 has been in contact with the CLARIN project.

Three WP4 participants (Ken Miller, Taina Jääskeläinen and Chryssa Kappi) have participated actively in the work of the DDI Controlled Vocabulary group. WP4 has

FP7-212214

also provided expertise for the preparation of DDI Best Practice Guide on Controlled Vocabularies.