



Enhancing the multi-lingual Research Infrastructure (RI)

Work Package 4 of the 2008-2009 CESSDA PPP under the 7th Framework Programme
(See <http://www.nsd.uib.no/cessda/project/index.html> for information on the CESSDA PPP.)

WP4 concentrates on strategies that will enhance information transfer across languages in the CESSDA RI and thus promote the functionality of the multi-lingual, multi-organisational infrastructure.

Multi-lingual thesaurus – IPR, extension and management

Thesauri are controlled vocabularies in which concepts are represented by preferred terms, organised in a way that makes paradigmatic relationships between the concepts are explicit. The CESSDA multi-lingual thesaurus has been developed through two previous EU-funded projects, LIMBER and MADIERA. The thesaurus is currently available in nine languages. CESSDA members use the thesaurus for indexing data, and the thesaurus functions as an important multi-lingual search tool in the CESSDA data portal. It forms a valuable resource for other European RIs, and various research infrastructures outside of CESSDA, in different countries, have expressed interest in using the thesaurus.

However, the copyright and intellectual property rights (IPR) issues for the multi-lingual resource, a joint product of many organisations, have not been addressed in the previous projects. Hence the conditions and licensing for use within and outside of the CESSDA infrastructure have not been formalised. This forms a barrier to its use outside of CESSDA. One of the key tasks of WP4 is to determine the intellectual property rights of the thesaurus.

WP4 will also give thought to what kind of role the thesaurus will have in the forthcoming, upgraded CESSDA RI, and prepare policy recommendations on the issue. The potential of extending the thesaurus to the native language of every member of the CESSDA RI is under investigation. Managerial and maintenance processes for the thesaurus need to be developed to ensure its long-term viability. Committing to these processes will become part of criteria for membership of the CESSDA RI for those members who are responsible for a language version of the thesaurus. Software tools will be needed for thesaurus maintenance and for adding new languages. The work package has already finalised functionality specifications for the thesaurus software which WP12 will develop during the PPP project.

There are also plans to use the thesaurus as a search tool for linking questions or variables with concepts in the online conversion keys platform WP9 is developing. WP4 will take a closer look at thesaurus terms related to concepts where conversion is often needed, such as occupation, education, socio-economic status. Potential terms will be compared to international standard classifications, with the aim to produce non-system specific, pan-European terms to aid searching for particular variables across countries.

Other controlled vocabularies to harmonise metadata

WP4 will contribute to data harmonisation and comparability by providing a strategy for the use of other controlled vocabularies in data documentation. We will provide a scenario of metadata elements which require a controlled vocabulary, and create source controlled vocabularies in English to be translated into other languages later. The work will be done in

close co-operation with the international DDI Alliance Controlled vocabularies group. In addition to enhancing metadata harmonisation, the vocabularies will make multi-lingual data documentation easier and quicker by providing an approved translation of certain metadata elements. Such specialist terminology is not easily found in dictionaries and glossaries.

Enhancing multi-lingual data search and documentation

The CESSDA data portal provides access to European social science data. WP4 prepares both functional specifications for software and strategy proposals for CESSDA which aim to ensure that a) searchers will find out relevant data in the portal no matter which language they used for searching and b) that they will understand the metadata that their searches have yielded.

One of the strategies currently under consideration relating to point b) is whether all CESSDA members should provide a minimum study description in English for datasets published in the portal. A survey of CESSDA members revealed that surprisingly many members already produce metadata both in their own language(s) and in English. Another strategy proposal under consideration, arising from the results of the survey, is whether the CESSDA should provide tools for easy multilingual data documentation. WP4 will be cooperating closely with WP5 regarding this issue.

Participation in Work Package 4 (contact persons at the national partner archives)

The project team of WP4 is comprised of staff from different CESSDA members.

Contact persons at each organisation:

Taina Jääskeläinen (FSD, WP leader); taina.jaaskelainen[at]juta.fi

Ken Miller (UKDA); millk[at]essex.ac.uk

Pam Miller (UKDA); millp[at]essex.ac.uk

Domingo Scisci (ADPSS); domingo.scisci[at]unimib.it

Christina Fredzu (EKKE); cfredzu[at]jekke.gr

Pilar Rey-del-Castillo (ARCES/CIS); prey[at]cis.es

Iris Alfredsson (SND); Iris.Alfredsson[at]jssd.gu.se