Report of DDI Concept Management and Harmonization Group

Place: Schloss Dagstuhl

Dates: 12-09-2011 --- 16-09-2011

Participants:

Franck Cotton (INSEE)

Dan Gillman (US BLS)

Jannik Jensen (DDA)

Wendy Thomas (MPC)

Problem – Determine how concepts, concept systems, and classification systems are the same and differ in DDI, SDMX and its RDF expression in the Data Cube Model, and SKOS. Code sets are seen as kinds of classification systems. Classification systems are kinds of concept systems. Concept systems are the highest level object on SKOS.

We found there are few differences, so the question became whether SKOS was adequate to handle the semantics of concept systems for statistics, in particular whether the relations in SKOS are rich enough. For instance, the relations available in DDI as expressed in the Code Scheme section are not contained in SKOS.

The analysis of SKOS revealed the following deficiencies:

* The relations broader than, narrower than (and their transitive variants), and related to are not as semantically rich as the relations defined in ISO 1087-1 – *Terminology work*, Part 1: *General Vocabulary* and the *Neuchâtel Classification Model*.
* Levels for statistical classification schemes are not adequately described
* Other minor additions were identified

We identified several use cases to help prove the need for some deficiencies found in SKOS. These use cases came from the INSEE, US BLS, and others.

None of the changes identified were seen as major changes to SKOS, all were additions, such as added detail. Specific additions to SKOS are as follows:

* Add *part of* and *generic* relations to be able to specify hierarchies more explicitly. Both directions are required. A transitive option will be added for each.
* Add *sequential*, *temporal*, and *causal* association relations under the general relations section. Both directions are required.
* Add the ability to number, name, and associate meaning to levels of statistical classification systems.
* Add the ability to describe the set of concepts in a level (defined as a sub-type of the SKOS Collection) as exhaustive and mutually exclusive with respect to the associated meaning.

There are a number of open problems that we did not have the time to address:

* Correspondence tables (i.e., harmonization maps between the concepts in concept systems) and item changes need to be accounted for in SKOS. We have a proposed solution to this, but it has not been fully agreed upon.
* Determine whether the Comparison Model in DDI is sufficient to express the Correspondence Tables in *Neuchâtel*.
* Identify all DDI elements that are conceptual in nature.
* Model concepts as they are used in statistical studies and data and express the identified conceptual elements in this model.

**Bibliography**:

* DDI – <http://www.ddialliance.org>
* ISO 1087-1 – *Terminology work*, Part 1: *General Vocabulary* – Available through ISO at <http://www.iso.ch>
* *Neuchâtel Classification Model* – <http://www1.unece.org/stat/platform/download/attachments/14319930/Part+I+Neuchatel_version+2_1.pdf?version=1>
* RDF – <http://www.w3.org/RDF>
* SDMX – <http://www.sdmx.org>
* SKOS – <http://www.w3.org/2009/08/skos-reference/skos.html>
* Statistical Data Cube – <http://publishing-statistical-data.googlecode.com/svn/trunk/specs/src/main/html/cube.html>