

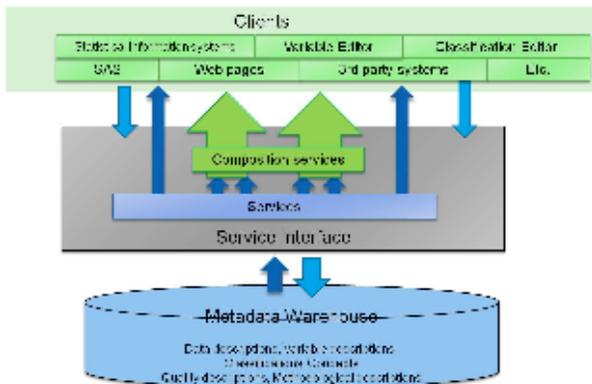
4. System and Design Issues (Finland)

Next section

4.1 IT Architecture

Statistics Finland's common metadata system is implemented according to the principles of service-based architecture.

Services meeting the needs of different user groups and client systems have a key role in the service-based architecture. The picture below shows the service interface to be built on top of the metadata warehouse, whose services produce the required data from the documents in the metadata warehouse, and also attend to storing of data to the warehouse.



The content of the metadata warehouse is maintained and it can be made available in client systems by ordering services through the service interface from the metadata warehouse. The service interface is implemented in line with the REST architecture (Representational State Transfer). The basic structures of the application are carried out according to the layer style. The business logic layer is formed of REST service interfaces, their processing logics and data transfer modules offered by the interfaces to client software. The function of data transfer modules is to offer data from the XML data warehouse to client software with an easy-to-use entity structure.

4.2 Metadata Management Tools

See Section 2.2.

4.3 Standards and formats

Statistics Finland has developed a Common Structure of Statistical Information (CoSSI) based on xml. It is a modular data model for describing statistical tables, classifications, concepts, variables, general information on statistical documents, quality descriptions, etc. CoSSI was designed in accordance with international standards such as the Dublin Core and CALS. If needed, CoSSI can be expanded; new elements, e.g. for data descriptions have already been integrated into it. In its ITC strategy, Statistics Finland has provided guidelines for the use of the CoSSI model. The data models of the classifications and concepts in use have been developed in the 1990s, and the elements they contain are presently part of CoSSI.

The basic structure and content of statistical information is defined in the CoSSI data model. It describes the information structure of the statistical data to be produced. The way in which data are produced, that is, the production steering system, is not described in the CoSSI data model. The definition of the data and content required by the production steering system was left to the future development phase of the model.

The data model comprises a description of basic information of data sets for the production and editing of statistical data and distribution of statistical information. At the moment, the model's parts to be extended and checked due to changed content requirements are as follows:

- Quality description of statistics
- The classification information model
- Supplementing the metadata part (docmeta) concerning the data record with data required by archiving
- Methodological description of editing
- Attaching source system metadata as part of statistical metadata
- Metadata of questions and questionnaires.

Preliminary examinations indicate that the CoSSI data model offers an adequate basis for producing content description data of statistical information following the GSIM data model (Generic Statistical Information Model, version 0.4/ 5.2012). A preliminary outline has been made to the CoSSI model of the structure that would cover the needs of Eurostat's different quality reports.

CoSSI documentation on the web: http://www.stat.fi/org/tut/dthemes/drafts/cossi_en.html

4.4 Version control and revisions

The versioning of the classifications can be seen as a 4-level hierarchy:

Level 1 (the highest level) is the classification name. It is a logical element in a hierarchy and its purpose is to aggregate all the statistical versions of a classification. In the classification database the classification name is expressed as a short technical name.

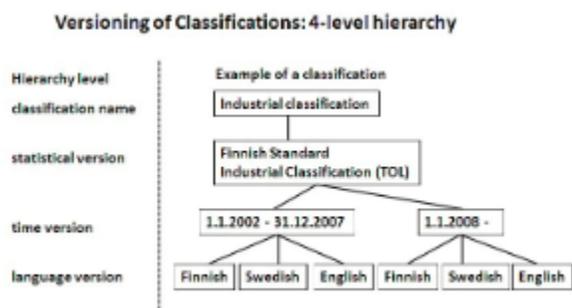
Level 2 consists of the statistical versions of a classification. There can be one or more statistical versions per classification. For example, statistical versions for the classification name "Industrial classification" are, for example, the Finnish Standard Industrial Classification TOL and the Industrial Classification for business services statistics. In the classification database the statistical versions are separated from one another by version number.

Level 3 contains the time versions of a classification. In the classification database the time versions are separated from one another by period of validity. When a classification changes, a new time version is made on the classification database. The old versions remain in the database and they can be used, e.g. in archiving.

Level 4 includes the language versions of a classification. The language versions (Finnish, Swedish and English) share the basic information with their mother classification or concept.

The concepts are being versioned in the same way as the classifications. The concepts usually stay the same for a long time but sometimes modifications take place, for example, due to amendments to legislation.

A configuration on the versioning of the classifications (see picture below).



(click on the thumbnail to view full size image)

4.5 Outsourcing versus in-house development

The user interfaces and the applications for the databases have been mainly developed and built in-house. The applications developed at Statistics Finland can in principle be shared free of charge with other statistical organizations. Where necessary, details regarding test use and access to more precise descriptions etc. may be agreed upon separately.

4.6 Sharing software components of tools

4.7 Additional materials Links to additional materials such as data models.

[Next section](#)