

# Case study: Statistics Finland

Case study: Statistics Finland Statistics Finland: use of GSBPM Statistics Finland: use of GSIM

Contact person*	Essi Kaukonen
Job title	Planning Officer / Metadata services
Email	essi.kaukonen@stat.fi
Telephone	+ 358 50 314 2311

## Summary\*

## Metadata strategy

Statistics Finland does not have a specific metadata strategy but metadata tasks are steered by the objectives of Statistics Finland's operational strategy such as the usability and reusability of data, reliability of statistics and standardization of processes. A common metadata system supports these objectives by providing a uniform way of describing statistical resources, such as classifications, data sets and variables.

The main principle is that metadata will be created and maintained in the common metadata system and will be made available to whatever use it is needed for in business processes from data collection to data dissemination. Common statistical metadata system will be further developed in order to rationalize and support the harmonization of statistical business processes.

The planning and building of the metadata systems are based on the principles of service oriented architecture included in Statistics Finland's ICT strategy. The metadata system will be developed block by block starting from the classification system and continuing with variables and concepts.

When developing the metadata systems, the national and international standards (e.g. GSIM, GSBPM) will be taken into account and utilized instead of creating own solutions. By means of this, the integration between the different systems will be strengthened; the data transmission gets easier; and customer service will be improved. Statistics Finland participates in national and international metadata cooperation and projects.

Metadata are developed in the projects as a part of Statistics Finland's other information architecture work. The smaller scale developing tasks are planned and coordinated in the change management groups.

## Current situation

At the moment, we are developing our metadata system and standard ways to utilize it in several projects.

## Metadata Classification

*The following categorization of metadata will be further developed in national ISAACUS Project. See Ongoing projects.*

### 1) Statistical metadata.

Statistical metadata consist of :

- descriptions and definitions of statistical data and variables
- classifications
- variable formulas and unit of measurement

### 2) Statistical data quality.

Statistical data quality reports consist of :

- statistical method descriptions
- relevance of data
- validity, reliability and accuracy of data

The former elements of the report are evaluated by quality indicators which are based on international recommendations.

### 3) Metadata of statistical documents or products.

Document and product metadata consist of information about:

- producers
- publication information
- identification knowledge of the publications or products
- field or subject area glossary
- keywords

### 4) Process metadata. Process metadata are divided into technical and conceptual metadata:

#### a) technical metadata

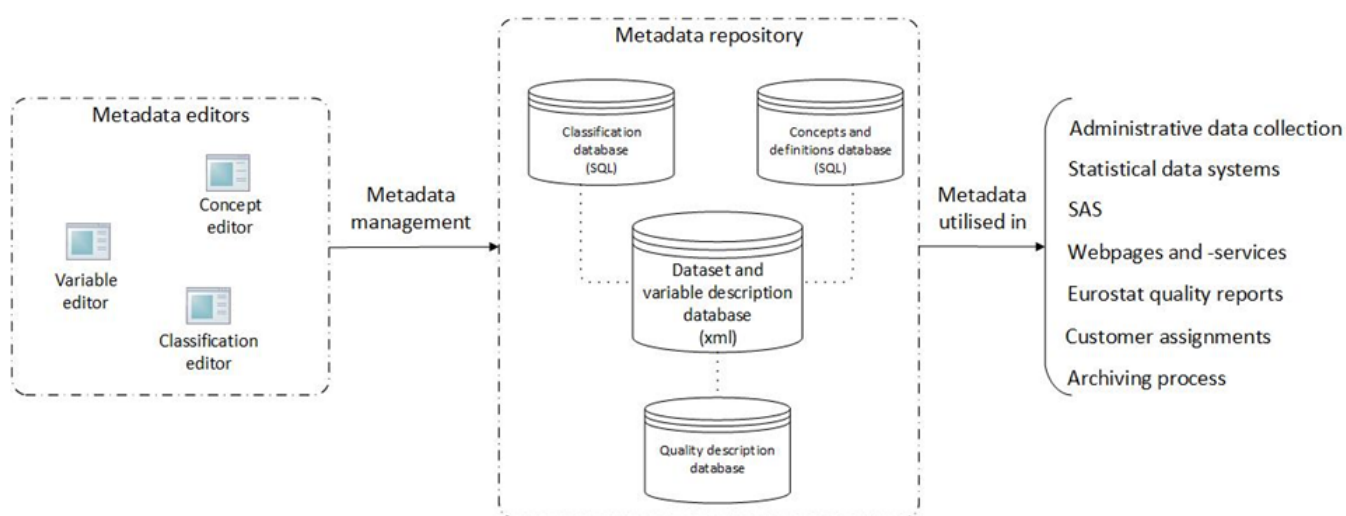
- technical metadata guide the process of data production: data collection, data management and data dissemination. For instance, it makes it possible to follow data production phase by phase. It also documents the process.

#### b) conceptual process metadata

- conceptual process metadata consist of the technical information of data and variables which are used in producing data. For example, they can be minimum or maximum values, various calculation rules or use of certain classification values.

## Metadata system(s)

Statistics Finland's Metadata System includes metadatapools of concepts, classifications and variables. The data is maintained using metadata editors. The content of the centralized metadata system is used in different parts of statistical production process, in different systems like SAS as well as in the internet. Nevertheless, there is still work to be done to ensure that all relevant metadata used in the process will be maintained in the central system.



## Concepts

The concepts and definitions published on the stat.fi website in the concepts service are produced from the concepts database. In the variable editor and the archiving database system, concepts can be retrieved from the concepts database and they can be combined to variable descriptions.

## Classifications

The content of the classification warehouse has long been utilised in the statistical production process, for instance as SAS formats and also in .NET applications and databases. Nevertheless, a lot of classifications are still stored within specific production systems and are thus not visible to other users. In the variable editor and in the archiving system, classifications connected to certain instance variables can be retrieved from the classification warehouse and they can be linked into the data description. The classifications published on the classifications webpages on the stat.fi service are also produced from the classification warehouse.

## Data and variable descriptions

The data descriptions stored in the metadata warehouse are created with the variable editor. Metadata maintained in other systems (classification and concepts database, the operational guidance and planning system STOJ) are linked into the data descriptions. Data description warehouse includes data descriptions from the published macro data sets as well as micro data sets for researchers.

The data descriptions in the metadata warehouse are also utilised in trilingual tabulation in SAS and PX-Edit.

## Ongoing development projects

The new metadata system aims to enhance the connections of the metadata warehouse to the statistics production process by making the metadata maintenance tools easy to use, by improving the connections of the statistical information systems to the centralised metadata warehouse, and by increasing services related to the use of metadata contents.

## **ISAACUS: Towards a national data description system**

ISAACUS-project (2016-2017) is a joint project together with the National institute for health and welfare (THL) and the Finnish social science data archive (FSD). The main aim is to develop a national information model suitable for register holders and other data producers. To achieve this there first need to be a common understanding at the conceptual level. GSIM is used to achieve this common understanding. For the two data producers in the project (Statistics Finland and THL) the objective is also to connect metadata more closely with the business process (GSBPM). For the long run this ideally means unified metadata from the data collection all the way to the dissemination.

## **Quality reports and SIMS**

Statistics Finland is developing its quality reporting system according to SIMS model. The main objective of the action is to create a common metadata warehouse which can be used to provide necessary user and producer oriented quality reports and requested metadata elements for Eurostat and other purposes. Currently the national quality reporting recommendation by Official Statistics of Finland (see [http://www.stat.fi/meta/svt/laatuseloste\\_en.html](http://www.stat.fi/meta/svt/laatuseloste_en.html)) does not contain all aspects of the ESQRS or ESMS/SIMS. The aim of the project is to align the basic publishing of quality information with the SIMS standard and to make a proper repository for various reporting purposes. The warehouse system might later be offered to other National Authorities (ONA) as well as other institutions of the Official Statistics of Finland.

## **New archiving system for microdata (unit level data), implementation project**

In the project that was executed in years 2015-2016, Statistics Finland renewed the process for archiving microdata and built some new tools for the process. The new process is based on combining data description (made with Variable editor) with data sets (SAS files) to an archive-ready xdf file.

The work now goes on as an implementation project. The most essential tasks to carry out during this Project are to make archiving task a solid part of the statistical process; to make clear instructions on the selection criteria, storing and disposal of archived statistical data; to ease planning and quality control for archived data and to make archived data qualified enough for immediate active use in research and other purposes.

## **New classification system, implementation project**

The new GSIM-based classification editor and the classification services as well as the classification warehouse were launched 16.02.2016. The system is already used for classification maintenance. The classification data is synchronized into the old classification system. During 2016 plans will be made about how and when all the other systems like, SAS and Internet, will be connected with the new system via classification services.

## **Metadata web service development**

Metadata web service development project is a project that aims to improve discoverability of statistics through metadata web services [http://www.stat.fi/index\\_en.html](http://www.stat.fi/index_en.html). Project was divided into three parts.

- 1) Planning and implementation of the research catalogue Taika. Taika is displaying the micro-datasets that are accessible (permit required) to researchers through the remote access service. (<http://taika.stat.fi>). The service is built upon Statistics Finland's information model CoSSI. Taika was published with a trilingual interface in the end of September 2016.
- 2) Planning and implementing the update of concepts in Statistics Finland web service ([http://tilastokeskus.fi/meta/kas/index\\_en.html](http://tilastokeskus.fi/meta/kas/index_en.html)). The metadata contents was renewed and more metadata about concepts was made visible to the general public. The improvement of metadata quality inside the concept warehouse is an ongoing task, and it continues beyond this project. The updated web service was released in the end of September 2016.
- 3) Final phase in this project is going to be the renewing of the classification web service ([http://www.stat.fi/meta/luokitukset/index\\_en.html](http://www.stat.fi/meta/luokitukset/index_en.html)). The new classification system is going to enable new metadata content and new possibilities in terms of discoverability. Also Open data issues are going to be examined during this phase. New classification web service is planned for a release in spring 2018.

## **Other data systems related to metadata**

TILKUT is a description database of statistical programs that contains basic data, like name, description, topics, keywords, publication frequency and contact person. The data from the TILKUT database are used in the stat.fi web service, the operational guidance and planning system STOJ, and in the variable editor.

The operational guidance and planning system STOJ includes information about publication times and contact persons of publications. The contact details of persons needed for data descriptions are retrieved from STOJ to the variable editor.

Starting from 2006, the data collection register contains data related to Statistics Finland's data collections. The system was originally built to serve metadata needs connected to direct data collections. The system was later extended to cover administrative data sets as well. In principle, the register should have all Statistics Finland's data collections described, but especially for administrative data sets, this objective has not been reached. The data are used in stat.fi's services to data providers, in the register of enterprise respondents and in Statistics Finland's planning and monitoring process. The register contains estimates of the burden caused by an individual data collection.

The register of enterprise respondents is a register intended for managing data collections to which samples, response data and respondent data are stored. The register is used to control whether a response has been received from a data provider and a rough estimate is given of the response burden.

For personal data collections, data on samples have already been collected for some time, but there is no actual register of them.

## Costs and Benefits

.

## Implementation strategy

The implementation strategy of the metadata system is step-wise.

## IT Architecture

Statistics Finland's common metadata system has been built ever since from beginning of 90's. The strategy today is to move towards service-based architecture step-wise.

The different editors (interfaces) have been built during several decades and, therefore, also the technologies used differ. We currently have .NET applications and powerBuilder applications. The newest, i.e classification editor, is a web-application.

created with ASP.NET MVC - framework. Editor uses classification services which is created with ASP.NET Web API (2.0) framework.

The databases behind are both SQL databases (SQL Server) and XML databases.(eXist).

## Metadata Management Tools

Statistics Finland's metadata system is operated by following tools:

- The Concept Editor (in use since 1990's, will be renewed soon)
- The Classification Editor (renewed system and editor are in production since February 2016)
- The Variable Editor (in use since 2013, will be developed further in following years)

These tools are used for metadata management. The metadata itself (concepts, classifications, variables) are used in several systems in statistical production process.

## Standards and formats

Statistics Finland aims to implement UNECE standards (GSBPM, GSIM, CSPA, LIM) in all development projects. The standards, especially GSIM and LIM, will be used as the base for the information architecture developed in upcoming years. Statistics Finland is planning to renew the data description system starting from 2016 onwards. The conceptual base for this work is GSIM. The main idea is to re-use metadata elements and show interactions between core metadata elements in different process phases.

Currently, Statistics Finland is using its own information model CoSSI "a Common Structure of Statistical Information" in most parts of the metadata system. GSIM Statistical Classification Model is used in the new Classification System.

CoSSI was developed in accordance with international standards such as the Dublin Core and CALS in the beginning of this millennium. The first version was published year 2003. CoSSI is a modular information model for describing statistical tables, classifications, concepts, variables, general information on statistical documents, quality descriptions, etc. The way in which data are produced (process metadata) is not described in the CoSSI information model.

CoSSI documentation on the web: [http://www.stat.fi/org/tut/dthemes/drafts/cossi\\_en.html](http://www.stat.fi/org/tut/dthemes/drafts/cossi_en.html)

## Version control and revisions

.

## 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.

## Sharing software components of tools

.

## Overview of roles and responsibilities

A high-level organisation structure map can be found on the Statistics Finland website: [http://tilastokeskus.fi/org/tilastokeskus/organisaatio\\_en.html](http://tilastokeskus.fi/org/tilastokeskus/organisaatio_en.html)

In connection with Statistics Finland's organisational change, the Metadata Services unit was transferred on 1 January 2013 from the Information Technology Department to the new Standards and Methods Development Department. The task of the Standards and Methods Development Department is to steer the statistics production process, to support statistical methodology in statistics production, to promote uniform application of metadata and classifications and quality work in statistics production and to intensify project work.

The main guidelines for the development of the metadata systems at Statistics Finland are coordinated and processed in cooperation with the Standards and methods and IT Departments. The statistical departments lay out the main demands and needs for metadatabases and their use in various phases of the statistical processes.

The Metadata Services unit maintains classification standards, concepts and the archiving metadata system. The statistical departments maintain their own (statistical) metadata in the centralised metadata systems according to the instructions made by the Metadata Services unit. The Metadata Services unit also trains and consults statistical departments in metadata issues and is in charge of controlling quality in the metadata systems.

The CoSSI model steering group is in charge of managing and developing the model according to user needs in a manner that will not expose its main structure to risk.

## Metadata management team

## Training and knowledge management

### *Training and knowledge management of metadata experts:*

New experts working at the Metadata Services unit have been trained mainly by mentoring and guidance provided by senior experts. ESTP courses on metadata training have been provided.

### *Training and knowledge management of metadata of statistical departments:*

Statistics Finland provides structured basic courses (for new recruits) and advanced courses dealing with statistical production. These courses contain also basic information about the present metadata systems, mainly classifications, concepts and archive data management.

The Metadata Services unit provides the personnel with informative briefings whenever there are major modifications made to the metadata system either in content or in application development. More systematically organized training is needed when new tools are brought into use. Half-day metadata seminars are organised yearly to present topical metadata issues.

As an example, in connection with the implementation project of the variable editor, an extensive training programme was carried out during which a one-day training was planned and organised for statistical experts on making of data descriptions and use of common tools for it. During the project, training days were held around twice a month. In the future, new modes of training, especially self-training by following LYNC-recordings and special clinics where you can work with your own material, are to be considered.

In practice, much of the training today for users of the metadata system is still done side by side. This leads to good results but is, in fact, ineffective as each client is trained individually and also resources used here would be needed in the development of metadata systems.

## Partnerships and cooperation

Statistics Finland cooperates with organisations and participates in working groups that define standard classifications or standards on both international and national levels. Metadata experts attend regularly Eurostat's Metadata Working Group and Classification Group meetings as well as METIS meetings. Statistics Finland has also representatives in the PC-Axis Reference group and Eurostat's Quality Working Group. Spatial metadata experts contribute in INSPIRE metadata work and attend national working groups in the implementation of the INSPIRE metadata process.

## Other issues

## Lessons learned

A metadata system complying with the uniform architecture is not just a technological renewal, but its implementation will require change in work procedures, responsibilities and organisation of tasks. The change in work procedures above all means timely recording of metadata in connection with data planning and production – a move from irregular retrospective description to regular and up-to-date description. The change is a challenge to the systems and applications, because work procedures change only if the technology allows it. An optimal procedure can be realised only if the users feel the applications are easy to use and serve their work. Commitment by the Management and their support to the work is crucial for the statistical units to be able to provide the contribution needed to the development work and for ensuring that the work will be sustained. The centralised metadata system should support the harmonisation of the production of statistics to a sufficient degree, thus making it more effective, but it should also be flexible enough to a certain extent to serve what statistics specifically call for. Involving statistics in the planning is needed.

Links:

## Attachments