

# Case study: Statistics Sweden

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

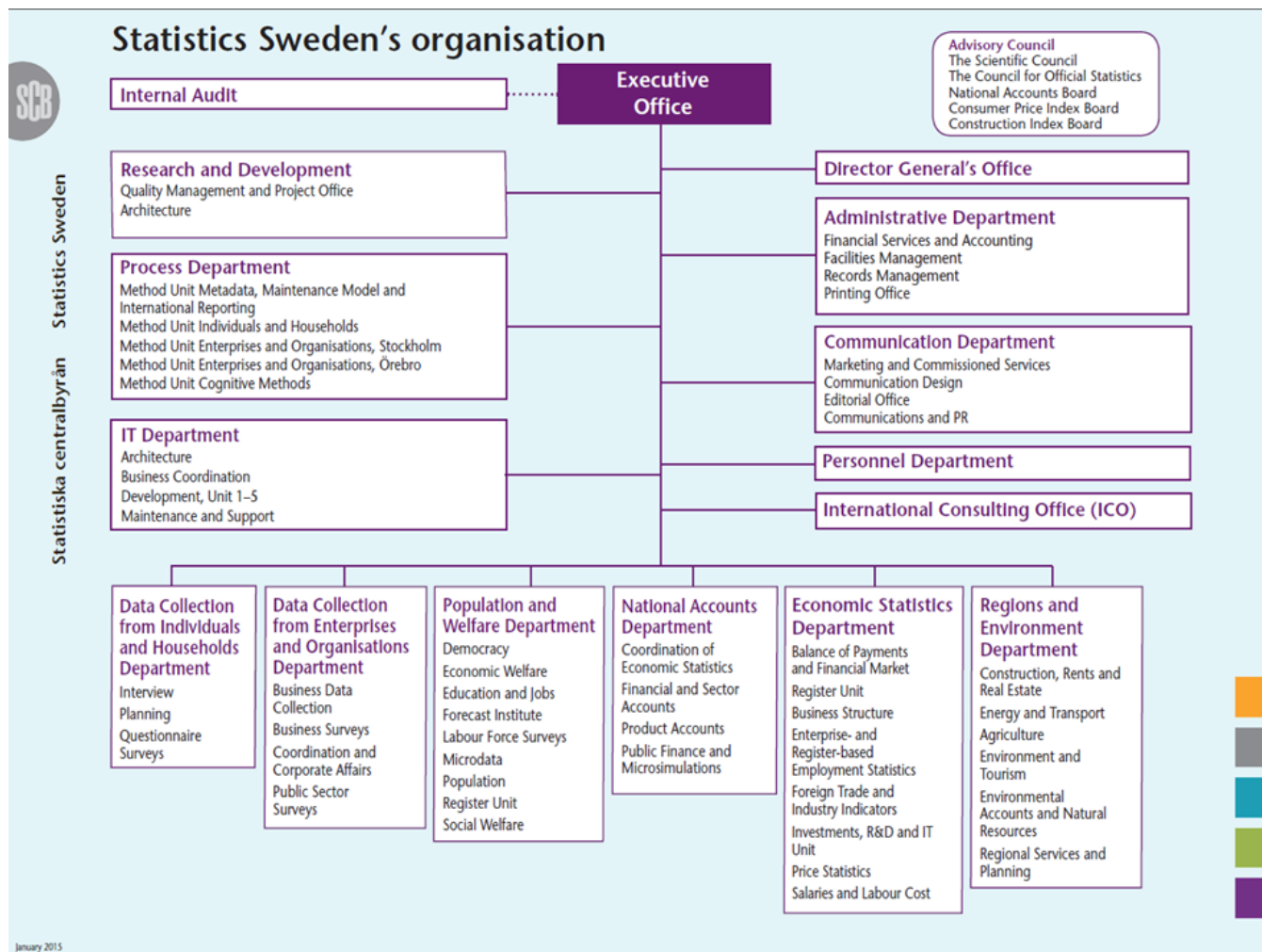
Contact person*	Klas Blomqvist, Process Department
Job title	Senior Advisor, Metadata
Email	klas.blomqvist@scb.se
Telephone	+46 8 50694352

Summary\*

.

Metadata strategy

Organisation



## Metadata strategy

Statistics Sweden has been active in the metadata field for a long time and has developed several metadata systems and templates that contain a lot of metadata. A metadata model for covering the whole production process was created in the late 80's and early 90's (SCBDOK). In 2007 Statistics Sweden introduced a major reorganization program and a new organization following the "Lotta"-project. The new way of working is process oriented production with an overall strategy that focuses on customers, efficiency and standardisation. There is however no separate strategy for the metadata management area. Metadata is handled by several organisational bodies within statistics Sweden. Classifications and documentation is by an area in the process department. Process managers are responsible for metadata related to their processes, the communication department is responsible for metadata related to publication in the statistical databases. The Research and development department is responsible for developing the business architecture and its encompassing metadata. The long term goal is to have common metadata for the whole organisation and that the golden rule of metadata, that it should be collected when it is created is incorporated in the production system in the future.

Metadata development is to be seen as included in the overall strategy for Statistics Sweden's data architecture which can be seen in the picture below. Here metadata has the role of driving and delivering information between different process steps.

The image below shows the architecture vision for a Data warehouse:

As mentioned above there is currently no overall metadata strategy. There is however an unofficial metadata vision, that is in line with the architecture in the picture above:

## Current situation

### Implemented:

SCBDOK was implemented in 1994.

Description of the statistics including the Quality Declaration.

MetaPlus was implemented in 2007. The Classification database is a subsystem of MetaPlus.

The Process support system first stage was implemented 22 September 2008, the second stage was released in September 2009.

Triton - A new platform for data collection integrated with editing tools. It is a common data collection tool for the whole production process. It integrates metadata and provides possibilities for metadata driven production.

The information model project - creating an information model for use in future process development projects. The first step of this work is going to be tied to the work within Triton as a part in defining system requirements.

The statistical databases contain separate metadata. It uses the PC-axis model.

The Product database.

## **Ongoing projects:**

DUR - Datawarehouse and register development program, a major redevelopment project

MetaPlus development - A project for improving Metaplus functionality is planned for release during 2013. This includes multilingual functionality and improved export and administration functionality.

## **Metadata Classification**

Statistics Sweden does not use any metadata classification other than the distinction between reference metadata and formalized metadata. The reference metadata are used for descriptions of the production process and for the published statistics, these are published on the internet. The quality framework mentioned above is for handling quality metadata. Macro metadata is handled in the statistical databases, but the long term plan is for this system to be connected to and fed by the common metadata repository.

## **Metadata system(s)**

Statistics Sweden has several metadata systems, some new and some old. The long term vision is to have one common system that provides metadata for all parts of the production process so that metadata is only stored once and reused in the other parts of the production process.

## **Description of the statistics**

Descriptions of the statistics are according to the Swedish Statistics Act to be available for all official Swedish statistics. The purpose is to provide brief information on the quality of the statistics and other basic facts. The descriptions serve an important function as a quality declaration for the published statistics. The documentation is in free text format following a standard template containing the main chapter's *General information* and *Quality declaration*.

## **SCBDOK**

Supports the whole production process. In 1994 Statistics Sweden decided that all observation registers and production systems under its responsibility should be documented in the SCBDOK system. The purpose is to provide a detailed account of the process of creating a statistical register, from data collection to presentation. The SCBDOK documentation is created in free text format following a standard template containing the following six chapters: 0. General information, 1. Contents outline, 2. Data collection, 3. Final observation registers, 4. Statistical processing and presentation, 5. Data processing system, 6. Log file.

## **MetaPlus**

MetaPlus was implemented in 2007, and was initially created to document microdata, supporting the process phase. The Classification database is a subsystem of MetaPlus. Previously Statistics Sweden used the in-house developed Metadok system to document final observation registers in a formalised way, and the Classification Database (KDB) to document national and international classifications. From 2007 both those systems were replaced by the new MetaPlus system. Chapter 3 of the SCBDOK documentation is derived from MetaPlus. MetaPlus is primarily developed for documenting final observation registers, describing the micro data. This means that at the moment it is mainly used in the statistical process phase. However the model is general and can be used for all stages in the production process. It can be used to describe raw data, in the data collection phase and to describe aggregated data used in tables in the statistical computation phase. MetaPlus can also be used in the design phase, for example to get variable definitions and construct the sample. How MetaPlus is going to be used in the production is not clear at this point, several projects are looking into that at the present and more work will be done in this field during 2010.

## **KDB, the classification database**

KDB is a Neuchâtel terminology based classification data base and contains classifications used at Statistics Sweden. It is a part of the MetaPlus system.

## **Macro meta, metadata for the statistical database**

Macro meta is used by the software system supporting the dissemination or output databases of Statistics Sweden, which are available on the Internet. The model is based on PC-Axis. Macro-Meta is the standard tool for entering and modifying the metadata in tables in the Statistical Database.

## **The product database**

The product database's primary objective is to keep track of Statistics Sweden's and other authorities responsible for official statistics statistical products. The product database has a central role in the Swedish statistical system and is an important coordination tool. The official names and identities from the product database should be used in all other systems that handle products. The data are used in Statistics Sweden's planning and monitoring process.

## **The process support system**

The process support system describes Statistics Sweden's standard methods, tools and work practices. It contains information on the different sub processes according to the Swedish process model. It contains technical documentation, manuals, templates, checklists, etc. for carrying out the statistical production process. The process support system is structured according to the definitions in the business process model adopted by Statistics Sweden. Each sub-process and activity is outlined in a hierarchical structure with links to more detailed information about the various methodologies, tools and work methods.

## **FMOD (governance model)**

FMOD is Statistics Sweden's governance model; it is to be applied in the management of processes related to the statistical production process. FMOD is scheduled to be introduced gradually into operation from 2009. FMOD is intended primarily for statistical products and some of the common processes.

## **Costs and Benefits**

At the moment Statistics Sweden has several metadata systems that are not formally connected. This has a lot of drawbacks as the same information has to be loaded manually into several systems. The vision for the future is to tie the different parts together into an integrated system of systems. This does not mean one giant system but subsystems that are linked to one another.

## **Implementation strategy**

A stepwise implementation approach is used for all systems.

## **IT Architecture**

The IT architecture at Statistics Sweden aims at maintaining a comprehensive Swedish statistics system. It should be focusing on:

- Consistency, structured and well documented data warehouse(s)
- Use registers when possible
- Minimize duplicate data and data transports
- Support process oriented production
- Support efficient data collection
- Support flexible dissemination

The foundation of the system is 3 base registers, individuals, business/organizations and real estate. The base registers are statistical register that are:

- Managing basic object types
  - Defines object types and connections
- Center of "sphere of interest"
- Autonomous
  - Has links to other base registers
- Containing stock variables (mainly)
- Containing time-stamped data

An overview of the architecture vision:

## **Metadata Management Tools**

MetaPlus contains an administrative tool. In addition a Super cross database for identifying areas for harmonization and outputs for using metadata in production has been created.

## **Standards and formats**

The classification database KDB is an implementation of Neuchâtel Terminology for Classifications. The variables part of the MetaPlus system is based on ISO/IEC 11179.

## **Version control and revisions**

All metadata in MetaPlus have a "valid to", "valid from" date stamp with a possibility to add time series and historical information. All metadata is logged and can be recreated. All reference metadata in SCBDOK and Description of the statistics have a reference time.

## **Outsourcing versus in-house development**

All metadata systems are developed in-house. SSD metadata is developed within the PC-Axis cooperation.

## **Sharing software components of tools**

Statistics Sweden can share the MetaPlus system and the business process model with other organizations. The MetaPlus model is available in English and a crude translation of the MetaPlus application is available in English.

## **Overview of roles and responsibilities**

Statistics Sweden has decentralized responsibility for registering metadata except for the statistical databases for which registering metadata is centralized.

## **Guidelines and follow-up**

There is a general director decision on what is to be documented and when the documentation should be published. This is used as a base for documentation follow-up. Statistics on documentation for the subject matter departments are presented every four months.

## **The Process department (PCA)**

The Process department is responsible for approving appropriate functionality for the production process this includes IT-tools, methods, documentation, support and education for the processes.

The process managers - are responsible for the business process model and the Process support system. This includes responsibility for management and operation for all sub processes in the production process, for developing, supporting and implementing the common processes and tools. These common methods, tools and working methods are established by General Director decisions, and they should be used in all situations where they can be applied.

The KMI area (Classifications, metadata and content harmonization) is an area in the process department responsible for metadata and classification harmonization as well as support for and follow up of documentation work. This includes responsibility for managing and providing training for the classification database, MetaPlus, SCBDOK and the description of the statistics.

## **The Research and development department**

The enterprise architects in the Research and development department are responsible for integrating metadata in the overall system architecture and developing the long term view.

The quality management area is dealing with issues concerning the overall responsibility for Statistics Sweden's quality management system.

## **The subject matter departments**

Product managers in the subject matter departments are responsible for the production of the statistical products. This also means responsibility for documenting the statistics, all statistics (in principle) should be documented in SCBDOK, MetaPlus and have a description of the statistics.

Documentation managers - each department has to have at least one. The role of the documentation manager is to coordinate the needs for training and support within the department and to review and approve finalized documentations.

## **The communication department**

The communication department is responsible for the website and the Statistical databases and it's metadata.

## **Metadata management team**

.

## **Training and knowledge management**

### **The KMI area**

The Process department is responsible for maintaining and developing the different metadata systems and templates. This also includes providing training and instructions. The department has an overall responsibility for contents in the metadata system. The organisational area in the process department responsible for this is the classifications, metadata and content harmonisation (KMI) group. The tasks involve administering and developing MetaPlus, providing training in metadata systems and templates and development in the field.

### **The documentation network**

All subject matter departments have at least one representative in the documentation network (the documentation managers), it is chaired by KMI. It is used as a means for distributing and collecting information on the documentation system. The group plans the contents of training sessions and discusses relevant problems that arise.

The documentation network discusses training issues. Training is offered by the KMI-group and provided by experts in the field, usually experts on the tool or template in question together with methodologists.

### **The classification group**

The responsibility for classifications is decentralised at Statistics Sweden. The purpose of the network is to create a forum where classification related issues can be raised and discussed. Exchange of experiences is one of the goals, but the long term main goal is harmonisation of classifications. It is used as a means for distributing and collecting information related to classifications, the group is chaired by KMI.

## **Partnerships and cooperation**

Statistics Sweden are part of PC-axis Nordic cooperation and were taking part in the Neuchâtel group

## **Other issues**

.

## **Lessons learned**

When building metadata systems:

- Involve different types users at an early stage, we used different groups for:
  - Methodologists
  - Subject matter statisticians
  - System developers
- Make it a content, not an IT-driven project
- Make simple prototypes as early as possible to get input from users, written requirements are to abstract

Be aware of that it is a complicated task and takes a lot of time to get things right.

**Links:**

## Attachments