

# Case study: Statistics Norway

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

Contact person*	Jenny Linnerud
Job title	Senior advisor
Email	jal@ssb.no
Telephone	+47 40 91 22 99

## Summary\*

## Metadata strategy

Statistics Norway has in the course of time developed many different metadata systems. This led to the same information being stored several times in several places making the availability of updated and consistent information difficult. In later years, there has been a strong focus on the need to link existing systems and a requirement that new metadata systems should not be built in isolation. To facilitate this, Statistics Norway developed a metadata strategy, which was approved early in 2005. The strategy focuses on establishing a conceptual framework, clear roles and responsibilities, and a stepwise development involving integration and linkage of systems.

For more information on our metadata strategy see: [http://www.ssb.no/english/subjects/00/doc\\_200508\\_en/doc\\_200508\\_en.pdf](http://www.ssb.no/english/subjects/00/doc_200508_en/doc_200508_en.pdf)

Almost all of the goals in our metadata strategy from 2005 are now accomplished, so in 2015 we are working on a new information management plan.

## Current situation

The objective of Statistics Norway's current work on metadata is to develop an integrated metadata system that will contribute to effective statistics production and dissemination, in addition to improved quality of statistics. Different metadata systems are being linked together making the metadata more easily accessible for all users. Metadata should be updated only in one place.

Metadata projects completed after 2005 include:

1. Documentation of key metadata concepts
2. Metadata portal on the Internet and the intranet
3. Variables documentation system
4. About the statistics as a content management system instead of just a text document
5. Analysis of the end-to-end creation and re-use of metadata in one production cycle for one statistic.
6. A service library for master metadata systems
7. Improved access to micro-data for researchers (About the data collections)
8. Administrative system for projects, products and processes

Current/planned metadata projects include:

- Creation of a master system for codelists and an upgrade to our master system for classifications

## Metadata Classification

The only metadata classification we have needed in-house so far is the distinction between conceptual and contextual metadata. This probably reflects the fact that we are only just beginning to try to integrate the metadata in the production lifecycle.

We are starting to use the GSIM grouping of information objects: Business, Concepts, Exchange, Structures

## Metadata system(s)

### Datadok - File descriptions (implemented)

We document all permanent archive data files in our file documentation database Datadok. The database was built in 1998 but wasn't mandatory until 2002.

### **Vardok - Variables documentation system (implemented)**

The overall purpose of the variables documentation system is to document variables in a central location, accessible by all, and to function as a tool for harmonising names and definitions.

There is a two way link between Vardok and Datadok (file descriptions database), a one-way link from Vardok to Stabas (standard classifications database), a two way link between Vardok and StatBank (dissemination database), a two way link between Vardok and Metadb (system for documentation of event history data) and a one way link from About the statistics, About the data collections and the statistical metadata portal to Vardok, via web services.

2006 was the last year in the development phase for the Vardok-project.

### **Stabas - Standard classifications database (implemented, but being replaced 2015-2016 by Klass that will also include codelists)**

The overall aim of Stabas is:

- To make work with and the use of standards simpler and more efficient
- To ensure systematic use of standards across different statistical areas

One main task is to make approved versions of the central statistical classifications available in a database system where they can be taken out at different aggregation levels, together with texts in different languages and relevant documentation, and where the classifications can be exported to other IT tools.

2004 was the last year in the development phase for the Stabas-project.

### **Service library for metadata systems (implemented)**

The purpose of this project was to

- Create a library of services for the master systems Vardok, Datadok, Metadb and Stabas.
- Define a framework for the description and formulation of SSB's metadata based on international metadata models (e.g. Neuchâtel) and standards (e.g. ISO/IEC 11179).

The project began in 2005 and ended in 2008.

### **Metadata portal (implemented)**

The overall purpose of the metadata web page is to make Statistics Norway's metadata systems more accessible and easier to use. Both internal and external users will get easier access to the metadata by displaying the contents of these systems in a common web page. The project began in 2005 and ended in 2009.

Metadata portal: <http://www.ssb.no/english/metadata/>

### **Metadb - metadatabase for event history data (implemented)**

Metadata for FD-Trygd (Social security database) and NUDB (Norwegian national Education Database).

FD-Trygd: details on demography, social conditions, social security, employment, search for employment, government employees, income and wealth. Data from 1992 to the present. Continuous regulatory and technical changes.

NUDB : All individually based statistics on education from completed lower secondary education to tertiary education from 1970 to the present.

### **Administrative system for projects, products and processes (implemented)**

This administrative system can be used to take out reports that combine manhours and other administrative information. It includes important information on all products in Statistics Norway such as financing, response burden, responsible division and person, response rates, frequency, laws, EEA requirements, subject field etc. This system contains both metadata and data.

### **About the data collections (implemented)**

Researchers frequently use data collections from Statistics Norway for their research. However, the process from finding out what you need, to actually getting the data, may be long and troublesome, especially for inexperienced researchers. Statistics Norway has therefore (with support from the Research Council of Norway) developed a website to make information about this process more easily available. Among other things, this page provides the users with documentation of several data collections. Each data collection has a general description e.g. of data quality, and it also contains a list of relevant variables, including variable documentation from Vardok. A new system is being scoped, hopefully with even more automatic solutions.

This system will be replaced in 2017 by [RAIRD](#) (Remote Access Infrastructure for Register Data).

### **About the statistics (implemented)**

About the statistics is metadata that describes each statistics that is published by Statistics Norway. It contains administrative information, information about statistics production, variables, concepts, sources of errors and uncertainty, comparability, coherence and availability. About the statistics now uses a CMS (Content Management system)-platform. CMS makes it possible to link About the statistics to Vardok and Stabas.

### **StatBank - dissemination database (implemented)**

StatBank Norway is a service where you may select scope and content of each table, and then may export the result in various formats to your own PC. This system contains both metadata and data.

## Costs and Benefits

### Examples of costs:

#### Metadata strategy:

A total of 1420 man-hours have been used in preparing the metadata strategy with ca. 35% of resources from IT.

#### Vardok:

A total of 12690 man-hours have been used in development with ca. 70% of resources from IT. A total of 476 man-hours from standards were used in 2007 for continued harmonisation of names and definitions, and training of personnel in the six new divisions. 294 IT man-hours were used in 2007 for maintenance and minor changes to the system.

#### Stabas:

A total of 7200 man-hours have been used in development 2002-2004 with ca. 75% of resources from IT. However these man-hours do not include the development performed by Statistics Denmark on the editing application. A rough guess for this would be 2500 man-hours. The system required approximately 1000 man-hours in production each year from 2005-2007 with ca. 70% from IT. We are now planning a new version of the editing application that we hope will be more flexible and less costly in production.

#### Metadata portal (man-hours used):

	2005	2006	2007	2008	Total
Senior adviser	200	300	325	230	1055
System architect	200	200	210	140	750
IT developer	150	1310	975	470	2905
Web designer	-	390	370	180	940
Total	550	2200	1880	1020	5650

### Benefits:

#### [IT-strategy for Statistics Norway 2014- 2017:](#)

#### Statistics Norway shall have easy access to data sources

IT shall develop better and more integrated metadata systems to help to enable data to be collected and reused across different sources and collection channels to a greater extent.

#### Statistics Norway shall be an effective and knowledge-based organization

IT shall contribute to the development of common solutions that ensure standardisation and automation of work processes, use the best statistical methods and create consistent quality indicators and link metadata and data in

the production of statistics in order to ensure good storage and reuse of data.

#### Data and metadata

Good descriptions of Statistics Norway's data, methods and processes are fundamental to the understanding of statistics and reuse of data. These metadata shall be systematically stored during the production of statistics, and be well integrated with Statistics Norway's data and statistical products. Further development of an effective and comprehensive system that ensures this shall be prioritised. Statistics Norway's description of data shall be based on national and international standards and models. This will make it easier for Statistics Norway to apply solutions developed by other producers of statistics or generally available software based on the same standards. Statistics Norway's statistical definitions and classifications shall be easily available for external use. Further development of the metadata systems will help Statistics Norway to disseminate open data with adequate documentation for reuse. Good metadata systems are also necessary for Statistics Norway's data archive to be harmonised and structured. This in turn will help Statistics Norway to effectively provide data for research and analysis.

## Implementation strategy

All our metadata projects are based on a step-wise approach.

### IT Architecture

Statistics Norway's technical solutions shall be built mainly upon the principles of service-oriented architecture. Guidelines on this are presented in *Norway's eGovernment plan*. All solutions for external users and most solutions for internal users shall:

- Have support for *open standards*.
- Be *platform independent*.

- Be *component based*.
- Have support for the packing in of data and functions in the form of *services* (web services).

These are central principals in service-oriented architecture. By applying these principles, applications and services can reuse existing functionality /components completely independent of the system they were developed in. In addition, by use of this technique, we can extend the lifetime of older applications, which have important functionality we wish to expose, just by creating a service layer on top of these. This increases the possibilities for collaboration between old and new applications in a completely new way, which gives benefits in the form of shorter development time, increased reuse and more consistent systems. This also enables us to replace systems behind the scenes, because communication with these is not directly exposed to the users.

For more information see: [http://www.ssb.no/english/about\\_ssb/strategy/it\\_strategy.pdf](http://www.ssb.no/english/about_ssb/strategy/it_strategy.pdf)

### **System architects**

System architects are introduced for each of the following areas in the top-level information architecture: data collection, metadata and dissemination. The mandate for this role supports the system architect's responsibility to ensure that IT development projects are in line with the IT strategy.

### **Solution architects**

Solution architects are appointed for each new development project to ensure that the systems development is in accordance with the current IT-strategy, architecture principles, recommended tools and practices.

## **Metadata Management Tools**

All our master metadata systems have Oracle databases. Links are mostly hard-links between these databases. Our web applications obtain metadata from our service library for metadata systems.

## **Standards and formats**

Our current classifications system is an implementation of Neuchâtel Terminology Model Part 1 Classifications v2.0 with the addition of an attribute on the Correspondence Item for Item Change from v2.1.

We are developing a statistical classifications and codelists system based on the GSIM v1.1 Statistical Classification and Codelist information Objects.

Our variables system is a partial implementation of Neuchâtel Terminology Model Part 2 Variables. The extent to which we follow ISO/IEC 11179, is best described by the figure in the attachment to this chapter where the number of instances of the objects per 2012 are given in the brackets. The figure in the attachments illustrates that there is little re-use of data elements or value domains in the current archive. This is a situation that we hope to address in the coming years.

We are considering using DDI in connection with micro-data for researchers.

We have used definitions of key metadata concepts from SDMX MCV where possible.

We have contributed to the development of GSIM (Generic Statistical Information Model) v1.1.

We contributed to a task force that looked at the flow of GSIM v1.0 information objects in GSBPM v4.0.

## **Version control and revisions**

Metadata have valid to, valid from and last updated. Program code for web-services is checked in and out of subversion.

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

- Improved editing tool for our classification database - previous editor was developed out-of-house.
- Analysis of the end-to-end creation and re-use of metadata in one production cycle for one type of statistics - no system development required (purely analysis).
- Service library for master metadata systems was developed in-house.
- Metadata portal on the Internet and intranet was developed in-house.
- Variables documentation system was developed in-house
- About the statistics as a content management system instead of just a text document was developed in-house
- Improved access to micro-data for researchers (About the data collections) was developed in-house
- Administrative system for projects, products and processes was mainly being developed in-house, but we had substantial help from external consultants on-site.
- Continuing development of a master metadatabase for questionnaires - in-house development, but with help from external consultants on-site.

## **Sharing software components of tools**

## **Overview of roles and responsibilities**

Subject matter statistician, survey manager, metadata manager, senior advisers in standards, IT developers, IT and information architects, solution architects and web designers are all important roles in metadata/statistical lifecycle management in Statistics Norway. List of metadata contacts Jenny Linnerud (jenny.linnerud@ssb.no) Anne Gro Hustoft (agt@ssb.no)

## **Metadata management team**

Our core team for metadata systems consists of one information architect and one IT architect for metadata systems in Development. Three programmers in the Department of IT. When necessary we draw on other IT expertise from the same department.

Metadata system maintenance is carried out by three people in the Department of IT. Maintenance of the system contents, i.e. the metadata, is carried out by all statistical divisions with support from the information architect in the core metadata team.

Details per system are as follows:

Our systems for classifications and variables and our metadata portal are owned by the Department of IT. In-house development (2 developers and two customers) was been carried out by the Department of IT. Maintenance of the systems is carried out by the same department (1 developer, 1 system architect and 1 system owner) with the addition of two people from the Department of Data collection and statistical methods in the case of classifications. Maintenance of the metadata in the systems is the responsibility of all 20 statistical divisions.

Our service library for master metadata systems was developed by 7 developers from the Department of IT and is being maintained by one.

Our Administrative system for products, projects and processes is owned by the Department of administrative affairs and director general. The previous system was outsourced by the previous owner but the new system was developed by 2 developers from the Department of IT in cooperation with the new system owner and one external consultant.

Our system for file descriptions was developed and is being maintained in the Department of IT. Maintenance of the metadata in the system is the responsibility of all statistical divisions.

Maintenance of the metadata in the system for event history metadata is the responsibility of two statistical divisions (Division for Social Welfare Statistics and the Division for Education Statistics).

## Training and knowledge management

All new employees participate in a course designed to give them an overview of all areas in Statistics Norway. Most new employees attend this course within 3-6 months of starting work in Statistics Norway.

## Partnerships and cooperation

Statistics Norway participates in relevant international meetings e.g. METIS, MSIS and informal groups e.g. Neuchâtel (classifications and variables). In addition, there is a strong tradition of Scandinavian collaboration (e.g. dissemination database StatBank) and collaboration within the Statistical Network (currently Australia, Canada, Italy, New Zealand, Norway, Sweden, United Kingdom).

## Other issues

So long as metadata is not an integral part of the statistical production cycle it will be prioritised lower than the publication of statistics.

In 2016, we will start to integrate Statistical Classifications and Codelists, from our new master system Klass, into production

## Lessons learned

- Top management support is essential.
- Make a metadata strategy. It is important that we can refer to formal documents like the metadata- and IT-strategy (which has been approved by the board of directors) in our metadata work. In the same way it is useful that the list of key metadata terms promoted for use within the statistical office has an official "stamp".
- Use step-wise development of metadata systems with active user involvement and regular delivery of functionality.
- Ensure continuous follow-up of progress and quality with direct feedback to users and regular reports to middle and top management. One of the biggest challenges in management of metadata is allocating the necessary resources. Releasing good quality statistics within the planned time schedule is the primary task for the subject matter divisions and documentation will often have a lower priority. It is therefore crucial that the management stresses the importance of documentation and increases the status for this kind of work.
- Harmonising variables between subject matter divisions is also a considerable challenge and an important tool to improve the quality of metadata. Several subject matter divisions may use the same variable names, but define them differently. In some cases this is necessary because of laws and regulations, but this is not always the case. We had meetings where contact persons from divisions using variables with similar names came together and discussed the definitions, e.g. if a division could change the wording of their definition to such an extent that other divisions might use it as well, which would allow us to reduce the number of definitions to one instead of e.g. three. This is a time consuming work which requires a lot of resources, both to monitor where harmonisation is needed and to do the job.
- The possibility to release metadata on the Internet makes it easier to motivate subject matter divisions to document metadata and improve metadata quality.
- We think that to really make metadata work a natural part of everyday life in the subject matter divisions, we have to include the metadata systems in the production cycle. Then we can establish routines where the handling of metadata is included in all relevant production steps. So far the metadata work in Statistics Norway has been focused on implementing metadata systems and filling them with relevant content. Now we need to focus more on integrating the metadata(systems) in the production cycle.

Links:

## Attachments