2. Statistical metadata systems and the statistical business process (Norway)

Next section

2.1 Statistical survey life cycle

Our statistical life cycle model is based on the work done by Statistics New Zealand, by Statistics Sweden and by METIS.

The phases we are currently using are as follows: 1. Specify Needs, 2. Design, 3. Build, 4. Collect, 5. Process, 6. Analyse and 7. Disseminate. We also have two overarching phases 8 Quality management - evaluate and feedback and 9 Support and infrastructure.

Comparison with the Generic Statistical Process Model v4.0 from April 2009 by the UNECE secretariat: Phase 8. Archive is covered in our model by sub-processes. The quality management part of the overarching phase Quality management/Metadata management is covered by our phase 8 Quality management - evaluate and feedback, while the metadata management part is covered in our model by sub-processes.

For more information refer on Statistics Norway's Business Process Model see http://www.ssb.no/english/subjects/00/90/doc_200817_en/doc_200817_en.pdf

2.2 Current 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)

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 from1992 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.

System for questionnaires and rules and

Systems exist but are being replaced.

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.

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.

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
unlike all the other systems described above.

2.3 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:

Statistics Norway's Strategies 2007 emphasise systematic quality control of products and processes. Statistics Norway's IT-strategies 2007 emphasise that
- metadata systems contribute to simplifying, improving and re-use of work processes
- data that are disseminated and exchanged must in addition to an agreed structure have sufficient metadata to give them meaning
- use of metadata systems are a pre-condition for the development of efficient data capture solutions according to Statistics Norway's data capture strategy.

2.4 Implementation strategy

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

Next section