

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Luxembourg, 9-11 April 2008)

Topic 2 (iii) Metadata and the statistical cycle and Implementation

EXPERIENCES FROM DISTRIBUTED REGISTERING OF METADATA IN METAPLUS

Submitted by Statistics Sweden¹

I. INTRODUCTION

1. Having one central metadata repository is of great value to a national statistical organization. The value added from this can in the long run give a more effective production – thereby cutting costs and reducing the respondent burden by using already collected data more efficiently. A central metadata repository provides an overall picture of the data collected and can be used to support different types of technical solutions for producing statistics.

2. MetaPlus was originally constructed for registering metadata for microdata. The actual registration of metadata in MetaPlus is carried out by the subject matter statisticians. This paper mainly deals with the support organisation built up for this purpose and the reasons for choosing this organisation. It presents a number of identified advantages and problems that arise from the chosen solutions.

II. Statistics Sweden's metadata system

A. MetaPlus²

3. MetaPlus is one of several parts that form Statistics Sweden's metadata system. It is a tool that was originally developed to primarily replace an old system for documenting microdata. However, the development project had a somewhat higher ambition: to create an open model that could be used in a wider perspective in the future. Thus the model is compliant with ISO 11179 making it possible to develop the system further to handle all metadata aspects of the statistical life cycle.



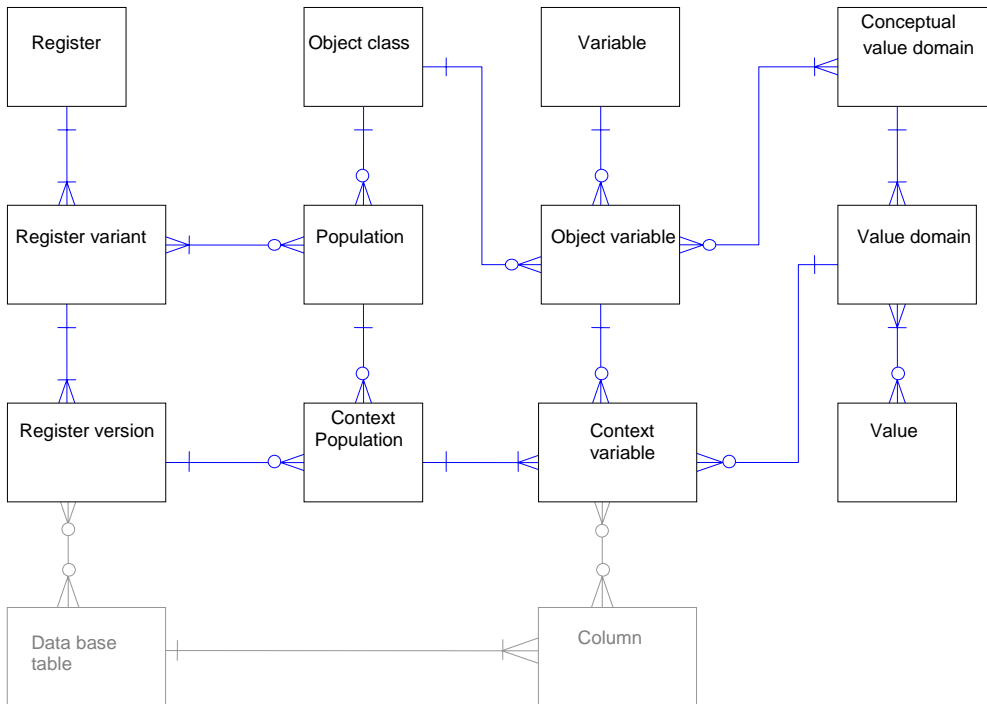
¹ Prepared by Klas.Blomqvist klas.blomqvist@scb.se and Lars-Göran Lundell larsgoran.lundell@scb.se.

² For more information about MetaPlus see "METAPLUS" by Blomqvist, Lundell, Karling and Svensson and "Developing a system for description of microdata at Statistics Sweden" by Blomqvist and Kristiansson

4. MetaPlus has been in use since the beginning of 2007 and the functionality is constantly being improved. Several projects concerning connections to the survey life cycle are planned or have already been launched.

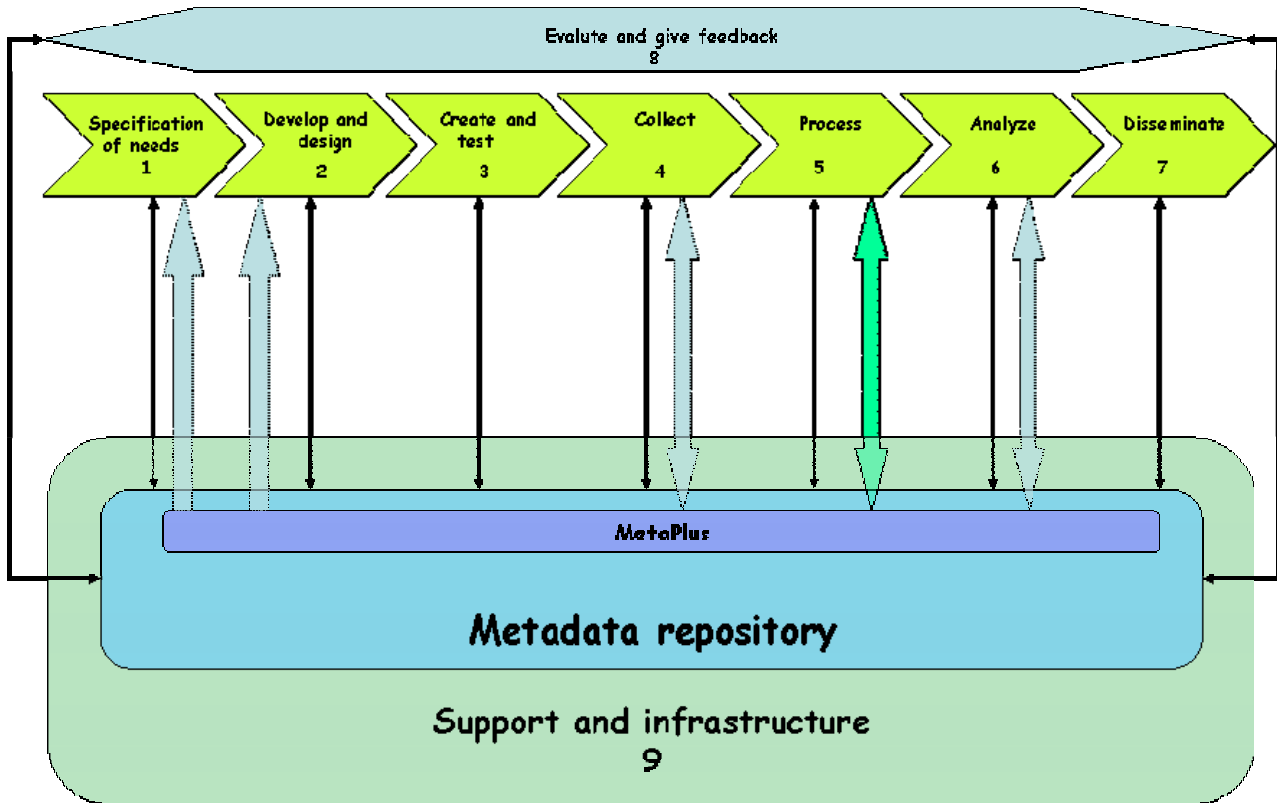
5. MetaPlus is based on a register structure and contains information on the related population and its object class. One of the main parts of MetaPlus is the variables belonging to a certain register version. Connected to the variable is the value domain, which may be based on a classification. Therefore MetaPlus includes a value domain and classification database.

Figure 1, the MetaPlus model



6. MetaPlus is currently the single most important tool to carry out central coordination and harmonisation of Statistics Sweden's metadata repository. This work is effected by the KMI Group of the Process Department. The KMI Group has an overall responsibility for MetaPlus. This includes instructing and motivating users as well as governance and further development of the MetaPlus system.

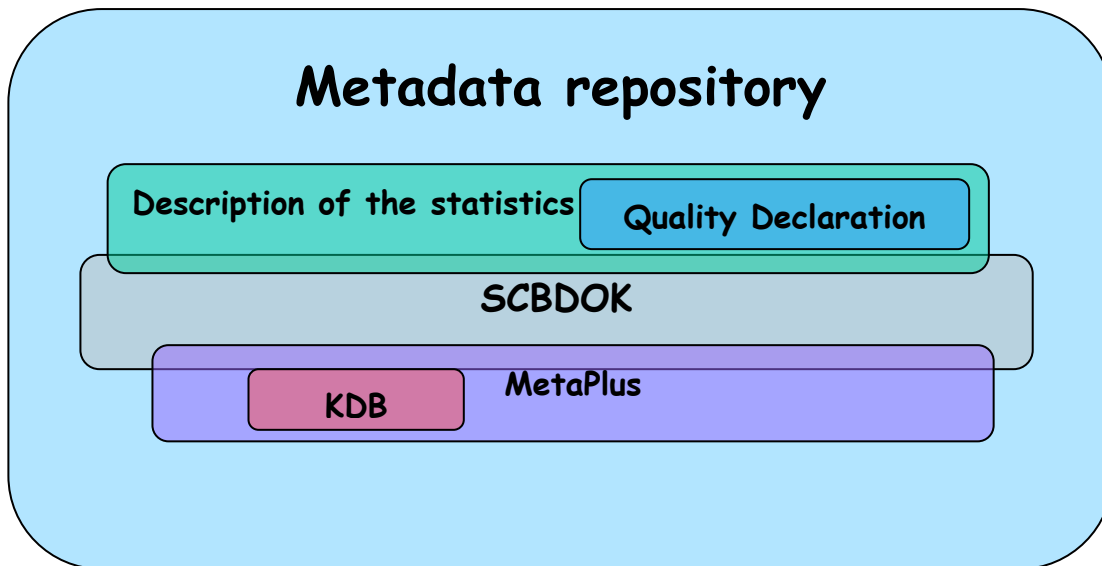
Figure 2, Metadata and MetaPlus in the statistical life cycle



B. Other parts of the metadata system

7. The other major parts of the central metadata system at Statistics Sweden are the Description of the statistics, The Quality Declaration and SCBDOK³.

Figure 3, metadata systems



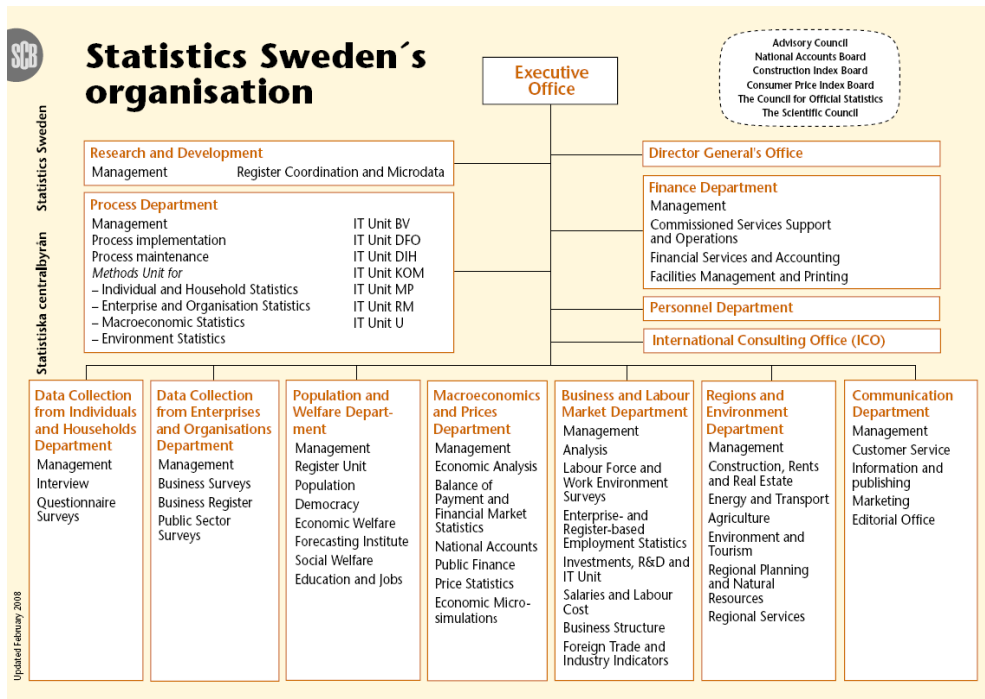
³ For more information about the Swedish metadata system see "METAPLUS" by Blomqvist, Lundell, Karling and Svensson.

III. The Organizational impact on documentation and registering metadata

C. Re-organising Statistics Sweden

8. Like many other national statistical organisations Statistics Sweden is currently going through a major reorganisation towards a process oriented production process. The long term goal is to achieve a more efficient production process than the traditional “stove pipe” model. The main means to reach this goal are standardisation and reduction of the number of IT tools used in the production. In order to realise these plans a new Process Department has been launched.

Figure 4, the new organization



D. The Process Department

9. A new department, called the Process Department, has been established in order to control the different parts of the production process. The Process Department has the overall responsibility to support the production process by methods and tools. The department is currently under construction and during the period of forming the new organization all Methodology and IT personnel have been transferred to the Process Department in order to centralize the control of the development resources. The Process Department consists of sub-process managers and assistant managers (responsible for a sub-process respectively), implementers, IT and Methodology personnel. The new organisation focuses on reducing costs, reducing the respondent burden and increasing quality in relation to investment.

10. The Process Department is responsible for providing the statistical production process with appropriate functionality by recommending and supplying methods and tools such as documentation along with support and training for the processes by evaluating and continuously improving them. The department is responsible for governance and operations of all sub-processes. This means that the development of the processes shall be carried through in a unified way. The department supplies support for implementing harmonized processes that are the outcome of development projects and makes sure that these processes will have access to the IT tools and statistical methods needed to carry out the processes. A governance model has been developed and a common test model is to be developed in order to provide fully operational and tested tools. These models will be used to support and control the sub-process functionality.

E. The Research and Development Department

11. The Research and Development Department, previously having varied roles in the organisation, has been given a new role, dedicated to high level architecture issues and long-term strategies.

F. The Communication Department

12. The Communication Department (formerly the Publishing Department) is responsible for presenting metadata and is also responsible for the Swedish Statistical Databases (SSD) and the metadata belonging to them.

G. Classifications, metadata and content harmonisation (KMI)

13. Metadata has been pointed out as a prioritised development area, since it is of great value as a driving force in a process oriented production environment and serves to promote standardisation and harmonisation. During 2007 the Research and Development Department established a new working group, called the KMI Group to deal with classifications, metadata and content harmonisation. In the new organisation this group is an organisational body within the Process Department, given an overall responsibility for classifications and metadata. It is organisationally parallel to the different parts of the production process, but affiliated to the sub-process "Support and infrastructure" (process no 9 in figure 2). The most important responsibilities in the KMI area include the central support for and management of the tools for registering metadata (MetaPlus, SCBDOK and Description of the statistics).

14. KMI promotes the view of having one central metadata system and work towards it. Today many internal users see metadata fragments that are not working together, but in the future this should be avoided. KMI wants to make the connections between the different parts of the metadata system clearer. The metadata system should be seen as one coherent unit that supports the whole production process. One of the group's most important goals is to make the benefits of using metadata visible in the production. Much work within the KMI Group so far has been focused on finding ways of cooperating within the group, the department and with other departments. Much effort has also been put into the implementation of MetaPlus.

15. The most important aspects of the work in the KMI area are:

- Responsibility for governance and development of Statistics Sweden's metadata system (this includes implementing MetaPlus, providing training, running a helpdesk, giving support to users, creating bug-fixes, presenting metadata on the web, etc.)
- Harmonisation of classifications and classification related methods
- Content harmonisation within the metadata systems

IV. Registering metadata with MetaPlus

16. The responsibilities in the documentation and metadata domain has for a long time been divided between three parties:

- The Research and Development Department was in charge of maintenance and development of the various parts of the metadata system
- The Publishing Department was responsible for presenting metadata, including metadata for the statistical databases (SSD)
- The subject matter departments were responsible for the contents and the registration

17. Due to the decentralised organisation an emphasis has been put on maintaining networks where those working with documentation and metadata can meet to exchange experiences and ideas. In the previous organisation these groups were chaired by the R&D Department.

18. In the new organisation registering metadata is still decentralised, but the division of responsibilities has been reduced to two parties: the Process Department and the subject matter departments. The subject matter

departments are still in charge of the actual registration, including editing and approval of the documentations. The Process Department has delegated the metadata and documentation responsibilities to the KMI Group.

19. The management is responsible for following up and monitoring that the documentation work is actually executed.

H. The documentation network

20. Due to the decentralised organisation groups and networks play an important part. The KMI Group provides a helpdesk and central support for the subject matter statisticians that register the metadata. Due to the KMI Group's limited knowledge of the specific statistical products the group member often finds it difficult, or at least very time consuming, to provide good advice in the subject matter specific issues that often arise in the registering process. Thus a link is needed, someone with more knowledge of the specific statistics who can provide support. The documentation network was established to create this missing link. Each subject matter department contributed at least one participant to the network, responsible for coordination of the department's needs for training and also for approving or failing the finished documentation. Due to the nature of this work many of the representatives chosen were methodologists. This fact, coupled to their relatively good insight into the statistical products made the documentation network a bridge between the specialised knowledge of the KMI Group and the needs of the users, but also a means of gathering a critical mass for these issues. This is a valuable recourse of ideas for future development and exchange of experiences.

21. In the new organisation all methodologists have been moved to the Process Department. In combination with a much higher pressure from management to document this means that the support organisation has been somewhat reshaped. A new group has been formed within the Process Department to deal with documentation issues. It provides methodological support to the subject matter departments in documentation related matters. Each department is now obliged to have at least one contact person who is responsible for documentation issues and a member of the documentation network. When needed support can be given from methodologists in the Process Department or from the KMI Group.

I. Content harmonisation

22. The KMI Group's governance of MetaPlus includes a responsibility for harmonising its contents. This is a difficult and time consuming task due to several reasons. One being that it must be possible to migrate metadata from the old system for registering metadata for microdata, Metadok, to MetaPlus. In Metadok there was almost no content harmonisation, a variable was almost synonymous to a database column. This means that after migration from Metadok there are many duplicate entities in MetaPlus, especially among the variables. The subject matter statisticians that register metadata also contribute to the duplicates since they do not always understand the principles for organising variables. This means that harmonisation becomes a cumbersome task that in principle will always be ongoing. In order to get started and be able to move forward a stepwise approach to harmonisation work has been adopted. The first step being to classify all registered variables with "finalised" status in the system according to the three categories below (see appendix 1 for more details on the structure).

- Statistical variables
- Administrative variables
- General variables

23. Each category has at least one subcategory where the variables are placed. When a "new" variable is registered that is not classified according to this structure it will be classified as a "temporary variable" and when its status becomes "final" it will be classified according to the three categories above. The second step in harmonising is to check for duplicates. The third is to make sure that the variables categorized in the structure will fit the requirements of the subject matter statisticians. If needed more subgroups are added.

J. Statistics and evaluation

24. Statistics Sweden's top management has defined MetaPlus as the organisation's microdata documentation tool. Since the subject matter statisticians handle the actual registering of metadata there is little central control of how much is actually being registered. To make sure that documentation is not forgotten or given low priority and to force the subject matter departments to take an active part in the process of registering metadata the documentation status is regularly monitored by the management. Statistics on the current status is extracted on department level from the MetaPlus database and reported to the management as well as being published on the intranet.

25. The management evaluates the status reports and is responsible for following up and taking relevant actions to make sure that all departments and units do their duties to register metadata. The KMI Group is responsible for following up and evaluating the detailed contents of the MetaPlus database to make sure that the quality of the metadata is at a reasonable level.

V. Experiences from working with MetaPlus

26. Statistics Sweden's long history of decentralisation, generally as well as regarding metadata, has made a continuation of this mode of operations the first-hand option. Changing to centralised registration of metadata would have called for a changed organisation having a relatively large central documentation group with a mandate to demand local cooperation. Motivating such a change would have been a difficult task.

27. The project that developed MetaPlus had from an early stage the ambition to make MetaPlus "more than just a documentation tool", i.e. to give MetaPlus a central role in the production process – from system design, through data editing, cross tabulation to publishing results and archiving files. This ambition favoured giving all kinds of personnel relevant access to MetaPlus and its database, thus requiring decentralised registration. The project's ambition also included making MetaPlus an open, flexible, and modular system, where in the future new components can be plugged in, which are currently planned, assumed or even unknown.

28. Somewhat more than one year's experience of working with MetaPlus is perhaps not enough to draw extensive conclusions, but we feel that we can pick out some results.

29. Some advantages of using a decentralised mode:

- Relatively sparse and efficient use of central resources
- Enforces knowledgeable participation from the subject matter specialists
- Possibility to join the metadata system with other systems

30. Some disadvantages of the decentralised mode:

- Central control is more difficult
- Coordination and harmonisation of contents is more difficult
- Local decisions may not always put registration of documentation as a top priority

31. We find that the advantages of working decentralised dominate the overall picture from Statistics Sweden's perspective.

K. MetaPlus and archiving

32. Much of Statistics Sweden's data is eventually sent to the National Archives for long-term archiving. Statistics Sweden has developed a dedicated archiving system that copies the database tables or files and transforms them into the required flat file archiving format, or vice versa. This system uses the information in MetaPlus to perform its tasks. The metadata from MetaPlus and SCBDOK accompanies the data file to the archives.

L. MetaPlus and the Personal Data Act

33. Statistics Sweden (and all other organisations keeping registers on persons) is required by the Personal Data Act to provide any person who so demands a print-out of the data on himself/herself. The lists printed must contain the codes and explanations (metadata). Since the number of registers containing personal data is high and the number of demands for print-outs may from time to time be high Statistics Sweden has been forced to develop a custom system to handle this task. This system is controlled by and uses information from MetaPlus.

VI. Future plans

34. Ever since MetaPlus was launched there have been plans for extensions and improvements in order to utilise the potential of its metadata store. The first of these plans to be realised is a revised presentation of metadata on Statistics Sweden's website. This presentation will be based on MetaPlus's database and will initially offer basic possibilities to search and download documentations and classifications. It will be generally available, but is designed to primarily serve the needs of researchers. The first version is scheduled for launching during April, but the exact date will depend on the number of finalised documentations available in the MetaPlus database.

35. Later additions to the web presentation will include links between metadata on micro and macro levels making it possible to trace the original variables behind a certain cross-table or to find the tables to which a certain (micro) variable makes a contribution.

36. A pilot project showed that it is possible and relatively easy to use the variable definitions and value domains documented in MetaPlus to build cross-tables in a standardised way. A planned addition to MetaPlus will provide an end-user tool for metadata driven production of tables.

37. The classification database will be expanded to make better use of correspondence tables and indexes according to the Neuchâtel terminology model.

VII. References

Bergdahl and Elvers "Reducing costs and raising quality through standardised processes and tools at Statistics Sweden"

Blomqvist, Lundell, Karling and Svensson "MetaPlus"

Blomqvist and Kristiansson "Developing a system for description of microdata at Statistics Sweden"

Sundgren "PROCESS REENGINEERING AT STATISTICS SWEDEN"

Sundgren "Reality as a statistical construction – Helping users find statistics relevant for them"

Sundgren "Manual to SCBDOK, Version 3.0"

Sundgren "DOCUMENTATION TEMPLATES AND METADATA MODELS AT STATISTICS SWEDEN"

Appendix: Classification of variables

Administrative variables

- Identifiers
- Communication variables
- Process variables

General variables

- Geographical variables
- Time

Statistical variables

- Agriculture, forestry and fishery
- Business activities
- Citizen influence
- Culture and leisure
- Education and research
- Energy
- Environment
- Financial markets
- Health and medical care
- Household finances
- Housing and construction
- Information technology
- Judicial system
- Labour market
- Living conditions
- National accounts
- Population
- Prices and Consumption
- Public finances
- Social insurance
- Social services
- Trade in goods and services
- Transport and communications
- Other