4. Statistical Metadata Systems (Statistical Office of the Republic of Slovenia)

back to case study home page

4.1 Metadata system(s)

A centralised, corporate metadata repository is implemented at SORS. It includes metadata about surveys, publications, statistical terminology, classifications and nomenclatures, advance release calendar. For each statistical survey methodological explanations have been developed and are available on SORS's website (in English as well).
Metadata repositories were developed over the years and today they are dislocated:
- Klasje - classification server,
- METIS for the annual programme of statistical surveys, survey instances, activities, working plan with activities, publications (and) release calendar,
- ISIS is a system for variables, questionnaires, address lists, process metadatada.

Integrated Statistical Information system (ISIS) is integrating all those systems.
We expect ISIS will become the central metadata system in SORS. The connection to classification server is very important especially when we prepare new variables. We define variable in ISIS. If the variable has the classification behind, than that classification has to be defined in classification server. From the ISIS there is the possibility to search through classification server and to find the proper classification for a new variable. If we do not find it, we have to insert it for the first and then find it from ISIS. In the next step the application shows us all versions of the chosen classification. That allow us to choose the version we need. And in the third step we can chose one of the levels in hierarchy of the chosen version. With confirming one of the levels of the version of the classification, we connect it to the new variable. We do not load all the categories to the ISIS database from classification server, but there is written only the link to those categories that the new variable needs.

4.2 Costs and Benefits

The current metadata system was gradually developed, starting in 1997.

4.2.1
It started with a "Modernisation and development of the statistical information system in Slovenia", Feasibility Study on the Architecture of Information Systems and Related Equipment Issues.
The study was carried out in the period of February - September 1997. A number of short term missions to SORS by experts from Statistics Sweden took place. Among the main conclusions of the study were:
- SORS has an excellent potential for developing a modern, register-based statistical system, based upon administrative sources in combination with sample surveys (which need to be designed in a more optimal way than at present). However, there are many demanding tasks to be tackled within a relatively short time period and with quite limited resources. In this situation there is an urgent need for focus and systematic planning in the development work.
- One way to obtain better focus in the development of the systems for statistics production in Slovenia is to specify very precise and concrete target architecture for the development, and to formulate a strategy for implementing this architecture step by step, in a systematic way. As a matter of fact, this approach has been proposed by SORS itself. Both top management of SORS and other staff members have expressed their sincere interest in establishing an "ideal" target architecture and a systematic implementation plan.
- The information system architecture for a statistical office should cover a number of different information systems types and their relations to each other: registers, survey processing systems (primary systems), analytical systems (secondary systems), and metainformation systems.
- It was further recommended, that the first step in the proposed architecture should be the building of a classification database. The prototype was presented at the board of director general as of 1 March 2000 and put in production in November 2000.

4.2.2
Within the StatCop98 project, the component 4.1: Development of conceptual, technical and software solutions of common (infrastructure) importance had the following goals:
- creating the concept of a statistical data warehouse with special emphasis on common functions and metadata as well as its testing on a pilot project;
- specification, development and introduction of EDI tools and procedures;
- classification database - upgrading the existing functionality, developing software for managing the concordances;
- developing software for browsing classifications via internet.

Another component - 4.3: Development of databases and software solutions - aimed at an integrated process of aggregation and dissemination of data from the Census of agriculture, horticulture and viticulture 2000 (AC2000) and other agricultural statistics (AGRISTAT). Within these two components, the basic common functions in the context of statistical data warehouse were defined according to Sundgren (Sundgren 1997): "Statistical metadata are descriptive information or documentation about statistical data, i.e. microdata, macrodata, or other metadata. Statistical metadata facilitates sharing, querying, and understanding of statistical data over the lifetime of the data".

4.2.3

The next project, STAT 2000 - focused on dissemination procedures. At that time, we believed that the electronic dissemination procedures at SORS are not well adapted to user needs and EU requirements. Most of the data collected by SORS are to the users available disperse, i.e. not in an integrated and comprehensive form and in addition they mostly lack the metadata. All official statistics have to be available in a uniform and user-friendly way via the Internet. Providing users with ample amounts of high-quality metadata encourages them to process and analyse official statistics in their own computer environments and to give feed back to the statistical office. The interaction between production and dissemination systems has to be organised on metadata based concept, so that the production system automatically feeds the output database, used by dissemination systems, with data and metadata.

An integral and contemporary dissemination system should be designed to cover the national and international user's needs. The data used in the compilation of SORS's dissemination system will cover all fields of statistics as well as series of regional data indicators for which the metadata system has already been initiated in the scope of the COP 98 project.

Metadata play a major role in dissemination of statistics, including helping users to find, understand and assess statistics in the context of their specific objectives. As standard tools and approaches to creating and managing, the metadata used for dissemination can also be used for survey design and statistical production activities.

Sharing information and metadata requires standard solutions both for the technology and the content.

Within the STAT 2000 project it was therefore essential to analyse and establish the underlying principles for specification and modelling of the dissemination database.

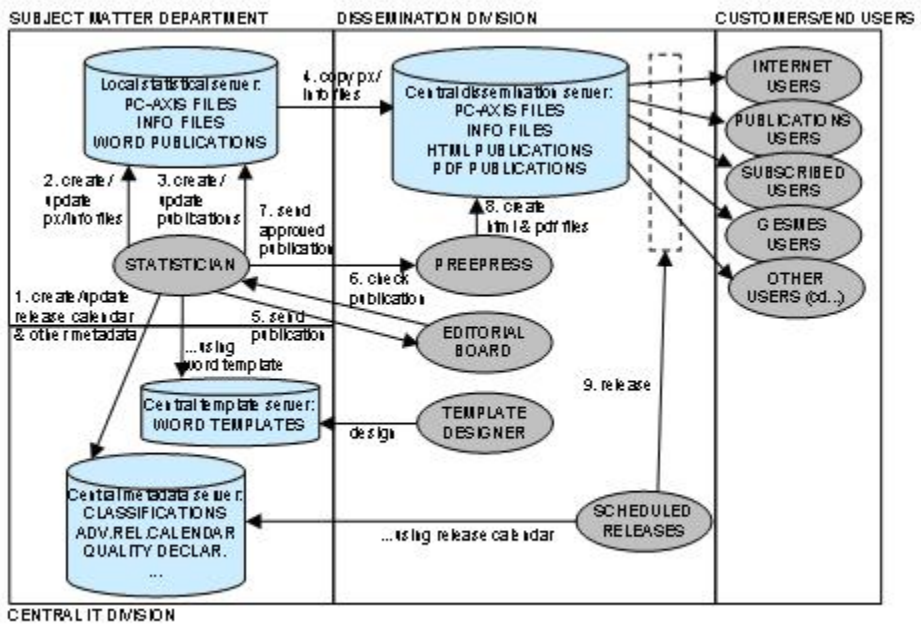

Figure 4: Conceptual scheme of functions to be supported within the "output" process

It was essential to study the entire data and metadata flow as far as possible all the way from the microdatabase to allow effective and efficient process, based on guidelines and rules that facilitate both: first production of statistical data and any other future use and reuse of the data and metadata.

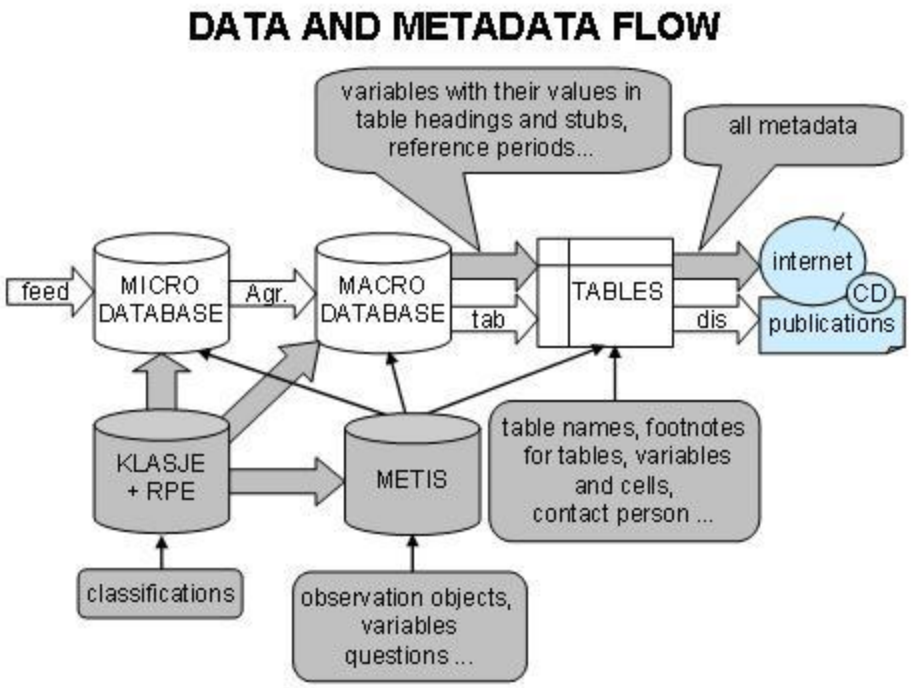

Figure 5: Conceptual scheme of the data and metadata flow from the microdatabase to the end user

4.3 Implementation strategy

Despite the fact that a large part of the development tasks in the area of information services for the support of statistical production should be finished by the end of 2008, the introduction of solutions into regular production is a particular challenge, and will continue at least until 2010. The introduction of new applications and tools into general use will increase SORS's need for constant improvement of the level of IT services (providing functioning, user support, infrastructure management, etc.) in order for the second main goal of the strategy to be efficient, and that internal and external users of information services are satisfied.

The following actions plans are elaborated in *Priorities of the national statistics in 2009 (Annex 6):

- Action plan for implementing the STRATEGY FOR FURTHER DEVELOPMENT OF NATIONAL STATISTICS IN SLOVENIA in 2009;
- Action plan for implementing the STRATEGY FOR FURTHER REDUCTION OF ADMINISTRATIVE BURDENS in 2009;
- Action plan for implementing the STRATEGY OF QUALITY in 2009;
- Action plan for implementing the STRATEGY OF DATA PROTECTION in 2009;
- Action plan for implementing the STRATEGY OF COOPERATION WITH REPORTING UNITS AND REDUCTION OF BURDENS in 2009;
- Action plan for implementing the STRATEGY FOR DISSEMINATION AND COMMUNICATION WITH USERS in 2009;
- Action plan for implementing the STRATEGY OF INFORMATION TECHNOLOGY UPDATE in 2009;
- Action plan for implementing the STRATEGY OF HUMAN RESOURCE MANAGEMENT in 2009;
- Action plan for implementing the STRATEGY OF FINANCIAL RESOURCES MANAGEMENT in 2009.