

# Metadata: A Quality Element in Official Statistics - the Swedish Approach

Bo Sundgren and Pat Dean

R&D Report  
Research - Methods - Development  
1996:5

## INLEDNING

### TILL

**R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.**

**Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.**

#### **Föregångare:**

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

#### **Efterföljare:**

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 1996:5. Metadata: A quality element in official statistics – the Swedish approach / Pat Dean, Bo Sundgren.  
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1996-X1010P9605



Statistics Sweden

Statistiska centralbyrån

# Metadata: A Quality Element in Official Statistics - the Swedish Approach

Bo Sundgren and Pat Dean

R&D Report  
Research - Methods - Development  
1996:5

**Från trycket** Januari 1997  
**Producent** Statistiska centralbyrån, utvecklingsavdelningen  
**Ansvarig utgivare** Lars Lyberg

**Förfrågningar** Bo Sundgren  
telefon 08-783 41 48  
telefax 08-783 45 99

Pat Dean  
telefon 08-783 47 27  
telefax 08-783 45 99

**© 1997, Statistiska centralbyrån Box 24300, 104 51 STOCKHOLM**  
**ISSN 0283-8680**

**Printed** January, 1997  
**Producer** Statistics Sweden, Department of Research  
and Development  
**Publisher** Lars Lyberg

**Inquiries** Bo Sundgren  
telephone +46 8 783 41 48

Pat Dean  
Telephone +46 8 783 47 27

Paper presented at The American Statistical Association's Joint Statistical Meetings,  
Chicago, IL, August 1996

**© 1997, Statistics Sweden, Box 24300, S-104 51 STOCKHOLM, Sweden**

## Metadata: A Quality Element in Official Statistics - the Swedish Approach

Bo Sundgren and Pat Dean, Statistics Sweden

Bo Sundgren, U/SIB, Statistics Sweden, Box 24300, S-10451 Stockholm, Sweden, BoSu@scb.se

**Key Words:** statistical databases, documentation templates (SCBDOK), users' needs, quality, official statistics.

### **1. Introduction**

Statistical databases are becoming one of the most important dissemination media for statistical data from government agencies. Electronic publishing in general has replaced many of the uses and functions previously fulfilled by paper reports. Furthermore, electronic publishing promises to meet new needs and requirements, some of which have emanated from the availability of the databases themselves. At this moment, statistical agencies are rallying to find ways of making their products available over the World Wide Web.

Electronic databases per se are nothing new, and have been around in various forms since the 1970s, but their content, flexibility to be used in diverse and novel ways, and their general use by government, academia, researchers, and the general public has grown exponentially since 1990. For example, daily users of Statistics Sweden's databases are: ministries, scientists, public sector analysts, private sector analysts, financiers, international organizations and organs, journalists, politicians and campaign managers, teachers and pupils, interested citizens. These users are as diverse in their needs as they are in their technical and statistical sophistication. Analysts and governmental organs require data at a level of reliability and precision that far surpasses what pupils or journalists require or are even interested in. And yet, databases must be constructed in such a way that all levels of needs and requirements are met simultaneously. This is the challenge of building and operating a database service: fulfilling a variety of needs, needs that can even be conflicting, with a single system.

Statistics Sweden has followed three guiding principles or philosophies when navigating through the quagmire of users' needs. These are: The legal obligations to provide public data as mandated by

the Swedish government, Statistics Sweden's own policy on quality matters, and the legal restraints imposed by the Data Inspection Board on the protection of confidentiality.

These obligations are spelled out in SFS 1992:889, a legal act which defines and stipulates Swedish Official Statistics (SOS). By definition, these are the statistics necessary for societal planning, research, general information, and international reporting that a government agency produces in accordance with its mandate from the government. According to this law, SOS must be objective and readily available, produced and disseminated while protecting the individual respondent's identity and confidentiality (SOU 1994:1). Users of SOS should be able to count on data quality and regularity of presentation.

There are also laws governing the dissemination of SOS. In the main, they can be summarised as follows:

Official statistics shall be published (presented and disseminated) in the series Sweden's Official Statistics (SOS) in printed form or electronic media. Printed publications are the traditional dissemination medium, but dissemination can also occur electronically. The presentation of official statistics electronically concerns data matrices, series, or ready-made tables. A matrix in a database can be used to produce a wide range of tables.

Official statistics must be made public in a timely fashion, i.e., as soon as they are produced. Publication should follow a schedule and be conducted in such a way that all users receive equal treatment. Level of detail and publication medium should take into consideration the shelf-life of the information. Official statistics should be available through country libraries, university libraries, public institutions of higher education, and Statistics Sweden's library without cost to the user. In other cases, marginal

pricing may be applied to the point of cost recovery for the publishing operations; nothing is charged for official statistics as such. The publication of official statistics in SOS should include explanations that allow the users to understand the statistics. Official statistics must, without exception, include a quality declaration and a product description. Titles and headings should be translated into English. Publication of official statistic in electronic medium should use a standardised format and include metadata.

## ***2. Production databases and dissemination databases***

When we look at a statistical system as a whole, for example, the statistical system of a country, we may distinguish between production databases and dissemination databases. A production database is typically confined to a single production system, for example, the production system associated with a particular statistical survey, or a particular register (population register, business register, etc.). A production database is a tool for organizing the input data and for maintaining the data throughout the different steps of the statistics production process: data collection, data preparation (coding, editing, etc.), aggregation, estimation, analysis and presentation. A production database is essentially a tool for the statistics producer.

In contrast, a dissemination database aims at satisfying the needs of statistics consumers. Since most users will need statistics from several surveys, a dissemination database should typically be fed with statistical data from several production systems. In this paper we are mainly discussing dissemination databases.

### ***2.1. Statistics Sweden's statistical databases and production environment***

Statistics Sweden is now engaged in two major re-engineering projects. One project, the so-called platform project, is down-sizing all production systems from the old mainframe environment to a completely PC-based, networked client/server environment. The other major project, the database project, is developing a new system for flexible data dissemination. This new database system is in itself one of the PC-based client/server systems in the network designed by the platform project, and the databases will be based upon the same data and metadata standards as the new, down-sized production systems. Data will be stored in relational databases, accessed through SQL, in the

dissemination databases as well as in the production databases. Structured metadata will be stored in relational databases, too, and for unstructured, textual metadata, a free-text management system will be used.

It has not been possible to obtain a perfect synchronization of the two re-engineering projects. The down-sizing of the some 200 production systems will not be finalized until late 1999. However, because of very strong pressure from statistics users, supported by the Ministry of Finance, which is itself a major user, the new system for data dissemination must be in operation already by the January 1, 1997. Thus, during a transition period Statistics Sweden will have to maintain both the "old" and the "new" dissemination databases - and an appropriate flow of data and metadata between them.

### ***2.2. Services to be provided by the new databases***

"Flexibility", "user-friendliness", "good documentation", and "low marginal costs for basic services" are key words in statements describing the goals of the new dissemination databases that are now being developed at Statistics Sweden.

The flexibility should allow a wide range of users to make productive use of the databases for a wide range of purposes. Although different user categories have quite different needs and preferences, and although they are differently equipped as regards hardware, software, and knowledge at their possession, and as regards money and time that they have available, they should all be able to find suitable services. Regardless of whether they are casual users (seldom-users) or professional users (frequent users), they should be able to find their preferred way of interacting with the database. It must be kept in mind that what is user-friendly to one user may indeed be very user-hostile to another one.

The most common complaints from users concerning the services offered by Statistics Sweden have been related to documentation (metadata) and costs. Because of insufficient or non-existing documentation, users have found it difficult to determine whether Statistics Sweden has relevant statistical data available that could help them with some problem or other. Furthermore, if they manage to find out that there are relevant data available, they sometimes find that even experienced staff members of Statistics Sweden may have difficulties to

retrieve and process the data, and that such operations take too long and cost far too much.

In response to these complaints the new databases are designed to provide a wide range of metadata that will facilitate search and retrieval as well as interpretation and analysis of statistical data. For example, the new databases will contain

- product descriptions
- publication plans and publication catalogues
- quality declarations
- observation register documentations
- production system documentations
- a classification database.

Part of the cost problem will be solved by moving production and dissemination systems from the old-fashioned, non-standardized mainframe environment to the modern, PC-based environments with its inexpensive technology and widely accepted de facto standards. For example, standard interfaces like SQL, Windows, and Internet will make it possible for users to access our data and metadata through numerous, inexpensive software products of their own choice.

### ***2.3. Basic services and value-added services***

The new database service will offer the users a number of low-cost (or sometimes even free-of-charge) services. These services will be based upon self-service and standard products delivered "off-the-shelf". The customers will not be charged for the data and metadata as such, but only for costs directly associated with the data delivery, for example media costs.

Some examples of basic services are:

- The user comes to the library of Statistics Sweden and retrieves statistics from the SQL-databases through the PC-AXIS retrieval software and user-interface. PC-AXIS is a Windows product developed by Statistics Sweden. It supports retrieval and manipulation of multi-dimensional statistical data from relational tables, and it facilitates data transformations to many different storage formats used

by other popular software products. The user downloads the selected statistical data and metadata to a suitable storage medium and takes them home.

- The user locates, studies, and prints statistical tables of interest through the Internet home page of Statistics Sweden.

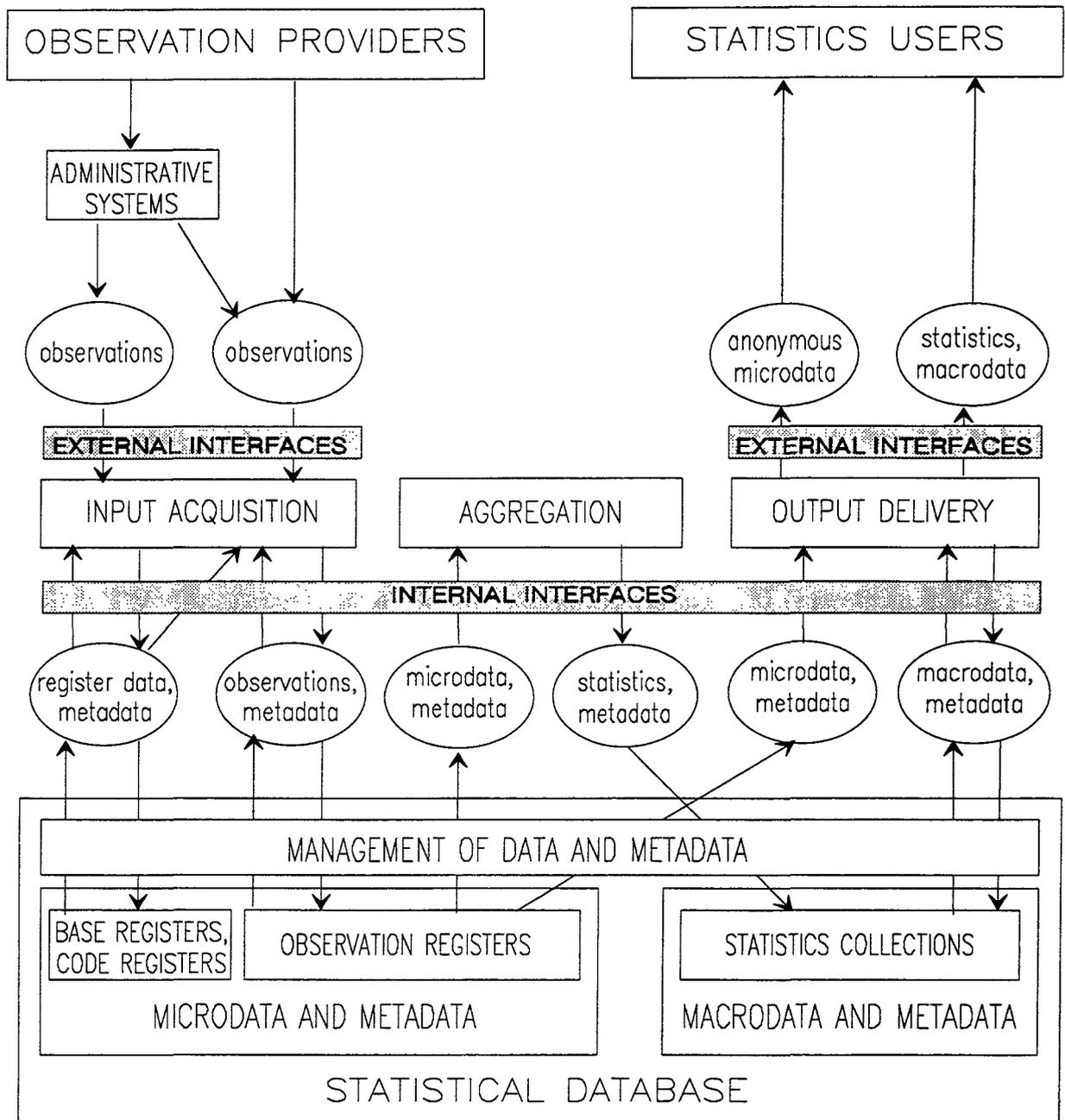
- The user locates, studies and downloads statistical data sets through Internet, using a free version of PC-AXIS.

- Some standard CD-ROMs with frequently demanded statistics, will be produced and updated monthly, quarterly, or yearly.

The value-added services are more tailored to the needs of specific users or user groups, and they are developed and priced on the assumption that all costs associated with the service should be recovered (except such costs that have already been covered over the Government budget of Statistics Sweden). Some examples of value-added services are: data shooting services; database replication; special CD-ROMs; high-speed on-line services.

### ***2.4. Database contents***

The macrodatabase will contain statistics on: the labor market, population, dwellings and construction, energy, business and industry, internal trade, foreign trade transports, agriculture and forestry, capital market, environment, national accounts, public finances, prices, justice, social welfare, illness, death causes, and health care, living conditions, incomes, education, elections, international statistics. The microdatabase will contain the following observation registers: the register of the total population, the register of population changes, population and housing censuses, the taxation register of real estate, the register of the education of the population, labor force surveys, the register of payments that are subject to taxation, the register of value added tax, the register of foreign trade, the register of motor vehicles, the register of incomes and wealth, the register of occupation, the register for education analyses, the register of criminality, the register of living conditions.



**Figure 1.** A functionally oriented model of a database-oriented statistical information system

### 3. SCBDOK - a description model for statistical surveys

In the early 1990s, the Director General of Statistics Sweden, assigned top priority to a documentation project. The project should develop a system for documenting the final data sets from statistical surveys, that is, the so-called observation registers that are archived for future use. The quality of the

documentation should be so good that future generations of researchers, living maybe 200 years from now, and who would have no first-hand knowledge of the data collections as such or the society in which they were collected, would be able to interpret and analyze the archived data.

The task was soon found to require a rather ambitious approach. In order to achieve the quality goal stipulated, conventional, rather technical documentation would not be sufficient. Technical documentation would certainly be needed in order to ensure that the data sets could be physically retrieved and processed by future researchers. But in order to ensure that the full information content of the data - and not only the data themselves - would be correctly transferred to future generations, many more aspects would have to be captured by the documentation, including subject matter aspects and statistical methodology. Thus Statistics Sweden decided to develop a documentation system, called SCBDOK, that would integrate aspects of subject matter contents, statistical methodology, and information systems methodology and technology.

Furthermore, it soon became evident that not even within a single organization such as Statistics Sweden there existed any consensus concerning concepts and terms to be used when describing statistical surveys. Thus, the first step in developing a documentation system for statistical surveys must be to establish a common conceptual framework. Such a framework was developed using a general description of a statistical survey, where key concepts were defined in their natural context. This is in contrast to another possible approach, where concepts are defined, one by one, out of their context. It was also decided to focus on concepts rather than terms. Certain terms would be recommended as preferred terms, but synonyms would be generously permitted.

The general description model was developed as a framework for standard definitions of concepts and terms that were precise and operational - yet as general as possible. For example, the concept of a survey was given a very general interpretation, covering surveys from different subject matter areas, sample surveys as well as total enumerations, one-time surveys as well as surveys that are repeated at regular (or irregular) intervals, indirect surveys, obtaining their data from other surveys or administrative information systems, as well as direct surveys with their own data collection processes, or combinations of direct and indirect surveys. Both registers, like the central population register and the central business register, and a statistical information system, like the system of national accounts, would qualify as surveys according to this very general description model.

Both the general description model and the documentation system based upon the description model were given the name SCBDOK.

### *3.1. Observation register documentation*

As already mentioned, the documentation task, which had been given top priority, was the task of documenting observation registers with such quality that they could be used by future generations of researchers without first-hand knowledge of neither the data collections, nor the society in which they had been collected.

Naturally, an observation register documentation must focus on the observation register as such and the variables (data items) that it contains as major documentation objects. Such an "object register declaration" could be described as purely "object-oriented" in the sense known from contemporary theories of information systems and software design. However, such a documentation would not be sufficient for the purposes stated. In order to fully understand the meaning and quality of archived data, and to appreciate the possibilities and limitations of the data, one must often know quite a lot about the processes behind the data: How were the observations made? Were there any unusual circumstances around the observation processes? What were the non-response rates, and what do we know about non-response distributions? How were the data edited and (presumably) corrected? What were the coding rules that were applied?

Thus an observation register documentation must contain both declarative (non-procedural) object descriptions, as well as rule-oriented and procedural process descriptions.

### *3.2. Production system documentation*

At first it was a somewhat surprising conclusion that a high-quality observation register documentation would have to contain quite detailed information about the processes behind the observation register. Once this insight had been gained, it had the good side effect that it was then relatively easy to specify a documentation template for a full-fledged production system documentation. Such a documentation would have the purpose to serve the production staff responsible for the operation and maintenance of a statistical survey and its production system.

### 3.3. Quality declarations

Not long after the SCBDOK project had been started, the Director General of Statistics Sweden gave top priority to yet another documentation project. This time the task was to develop a system for quality declarations to be associated with statistics published by Statistics Sweden. Whereas the SCBDOK project focused on observation registers, that is, microdata, the Quality Declaration project focused on aggregated statistics, that is, macrodata. At the time when the project was started, macrodata were still typically published in traditional, printed publications, but today quality declarations of statistics published through dissemination databases or other electronic media, like diskettes and CD-ROMs, are equally important.

The purpose of a quality declaration is to assist a user of published statistics to judge the quality of the statistics. It should be noted that the quality concept used here is not an absolute one, but it must be related to the intended use. Thus it is only when a concrete user is considering using certain published statistics for a concrete purpose that this use can determine the quality of those particular statistics for that particular purpose. A quality declaration should inform a potential user of statistics about facts that are relevant for the user when he or she is going to make such a judgment. It must be stressed that the user must take the responsibility for the final judgment of the relevance and usefulness of the published statistics for a certain purpose, even if the statistics producer is, of course, fully responsible for the factual contents of the quality declaration. It should also be noted that the very same published statistics may be of satisfactory quality for one purpose, although they are not at all useful for another purpose.

In MIS 1994:3, Statistic Sweden's quality policy is described as both a formulation of a notion of quality for official statistics and guidelines for quality declarations of official statistics. The guidelines are intended to serve both the producers and the users of official statistics. Quality is a fluid concept which changes over time. Recently, quality has come to include a holistic aspect, a focus on "total" quality. Fundamental precepts in this new quality thinking encompass (1) that the user and his/her needs are central. The user's perspective defines the quality of a product (goods and services) and its utility. Product development should also assume the user's perspective. To operationalize this

notion of quality, (2) that quality expresses itself in all aspects of a product and should strive to satisfy all aspects of the user's needs and expectations. The user's "needs and expectations" should be interpreted as both the product's performance and the service required to procure and use the product. For the user, quality is relative to intended use. A single product could have very high quality for one use and very poor quality for another use.

### 3.4. Product overviews

In addition to the documentation needs of different types of users of observation registers and published statistics, and in addition to further documentations needs of producers of statistics, there are yet some other documentation needs, originating from managers of the statistical system as a whole. For example, the government officials of the Swedish ministry of finance, who are preparing budget proposals and other instructions for Statistics Sweden, have their particular needs for documentation. They need a relatively short and informative overview of each survey, or "product", that is financed by the government budget. Such a product overview, or "product description", should focus on certain administratively important facts, such as purpose and main users of the product, contents in terms of major statistical characteristics covered by the product, frequency, and costs.

When the responsibility for some parts of the official statistics in Sweden was decentralized, it was decided that such a product summary would be able to play an important role not only for the ministry of finance, but also for Statistics Sweden in its role as coordinator of the Swedish statistical system. In order to serve appropriately for that purpose as well, the entirely administrative focus would have to be broadened in order to cover important contents and quality issues.

In 1995, the product overview template was used for the first time to collect the basic facts needed for management and coordination purposes. The questionnaire corresponding to the template has to be completed by everyone, who is responsible for a survey belonging to the Swedish system of official statistics, that is, some 200 statistical surveys, or "products". The product overview will be updated yearly.

Product overviews also play an important role as a metadata entry to the new dissemination databases.

### 3.5. *Unified and simplified SCBDOK template (SCBDOK 2.0)*

In 1994 the Director General of Statistics Sweden took the decision that the SCBDOK documentation templates should be used for creating documentations for observation registers and statistical surveys produced by Statistics Sweden. During the implementation of this decision there were certain reactions concerning the documentation templates from those who were responsible for creating the documentations.

One reaction was that it was confusing that the observation register documentation template and the production system documentation template were so similar - and yet not identical. Naturally the instructions accompanying the two documentation templates explained very carefully how the two types of documentations were related to each other, and how the feared duplication of work could very easily be avoided. Nevertheless, these reactions indicated that the two documentation templates needed to be further coordinated, or even unified.

Another reaction was that the documentation templates were far too detailed. Once again the accompanying instructions very clearly explained that the templates need not be "slavishly" obeyed. Obviously, what is most important is that all important aspects of the observation register and the survey are covered somewhere in the documentation. The main purpose of the templates is to serve as checklists, so that nothing important is forgotten. Nevertheless, the complaints indicated that there would be certain pedagogical advantages to be gained by simplifying the documentation templates.

As a result of these reactions a new, unified and simplified SCBDOK documentation template was developed in close cooperation with some of the most outspoken critics of the old templates. The new documentation template, called SCBDOK 2.0, was formally established by the Director General in 1995.

### 3.6. *The PCDOK documentation tool*

The creation of SCBDOK documentations is facilitated by a simple, computerized tool, called PCDOK. A number of different versions have been developed, reflecting on the one hand changes in the underlying SCBDOK templates, and on the other hand reactions from the users of the tool.

Already at an early stage it was decided that the tool should be technically simple and embedded in Microsoft Word for Windows, the word processing system used by all staff members of Statistics Sweden. By choosing this strategy, the developers hoped to keep the "threshold" to documentation work as low as possible. Since everyone was familiar with the word processing system, the users of PCDOK would automatically be familiar with most technical features of the documentation tool as well. Some macros were written to further facilitate certain parts of the documentation work. For example, the consistency between certain parts of the documentation was automatically maintained, and certain graphical tools enabling the documenter to introduce object graphs (in the U.S. often called Entity Relationship diagrams) and system flows.

PCDOK was designed to support the formalization of variable description tables, so that users could automatically be transferred to SQL databases. Some other parts of the documentation were formalized in similar ways, but the bulk of any documentation would be free text. This was in line with the preferences expressed by subject matter experts and statistical methodologist. Many information system experts would have preferred the documentation to be more structured and formalized. However, at this stage we regarded the acceptance by subject matter experts and statistical methodologists to be more critical than the acceptance by information system experts, which we could more or less count on from the beginning, since such people are long acquainted with the necessity (at least in principle) of documentation work.

**DOCUMENTATION TEMPLATE FOR A STATISTICAL SURVEY**

<p><b>0 Administrative information</b></p> <p>0.1 Survey name  0.2 Branch of statistics  0.3 Responsible organisation, person, etc.  0.4 Approximate cost of the survey  0.5 Purpose and history of the survey  0.6 Users and usages  0.7 Voluntary/mandatory response  0.8 Confidentiality and data annihilation  0.9 EU regulations and requirements</p>	<p><b>1 Summary</b></p> <p>1.1 The survey plan  1.2 Contents: statistical characteristics  1.3 Output: statistics and microdata  1.4 Time frame  1.5 Documentation</p>
<p><b>2 Data collection</b></p> <p>2.1 Frame and frame procedures  2.2 Sampling procedures  2.3 Measurement instrument  2.4 Data collection procedures  2.5 Data preparation (coding, editing, etc.)</p>	<p><b>3 Observation register</b></p> <p>3.1 Target and observation objects  3.2 Variable lists  3.3 Physical organisation  3.4 Experiences from last survey cycle</p>
<p><b>4 Statistical processing and presentation</b></p> <p>4.1 Estimates: assumptions and computation formulas  4.2 Presentation and dissemination procedures</p>	<p><b>5 Data processing system</b></p> <p>5.1 System summary and system flow  5.2 Processing  5.3 Database models  5.4 Database tables  5.5 Database accessories  5.6 Reports  5.7 Other data sets</p>
<p><b>6 The log book</b></p>	

*Figure 2. Documentation template for a statistical survey and its production system.*

### **3.7. Metadata for the dissemination databases**

We have introduced a number of documentation types: observation register documentations, production system documentations, quality declarations, and product overviews. By creating these kinds of documentations by means of the proposed documentation templates we will create semi-structured sets of metadata that can be used by computerised information systems in general, and by dissemination database systems in particular.

The dissemination databases that are now being developed at Statistics Sweden will use the above-mentioned documentations (with the possible exception of production system documentations) as

important metadata sources. Other metadata sources will be the classification database, containing detailed information about classifications used by Statistics Sweden, as well as international standard classifications, and yearly produced publication plans and publication catalogues.

Certain parts of the metadata are highly structured and will be stored in SQL databases. Less structured parts of the metadata will be treated like free text, managed by commercial software products available for that purpose, like BRS.

In a dissemination system for statistical data (macrodata as well as microdata), metadata are needed for several purposes:

- to identify statistical data of potential relevance for a certain user problem;
- to make a more detailed analysis of the usefulness of particular data for a particular purpose;
- to locate statistical data, that is, to find out where it is available, and how it can be retrieved;
- to actually access and retrieve statistical data and metadata;
- to interpret what has been retrieved;
- to process and analyse retrieved data and metadata;
- to evaluate the results of analyses;
- to obtain references to other analyses of the same or related data, to retrieve presentations of the results from such analysis, and to establish contacts with experts behind them.

For some of these tasks, highly structured metadata are needed, for others free-text metadata will serve as well or even better.

A new class of problems in connection with flexible dissemination databases concerns the rights and responsibilities of the statistical agency behind the databases. The problems may be illustrated by the following questions:

1. Who has the copyright of official statistics and of statistical data and metadata available and derivable from statistical dissemination databases managed by statistical agencies?
2. When should a statistical agency request to be quoted as being the source of data/metadata retrieved and/or derived from a dissemination database managed by the agency, and when should it rather request not to be quoted? When should it even request to be explicitly stated as not responsible for data/metadata retrieved and/or derived from the dissemination database?

As regards question 1, Statistics Sweden has taken the policy decision that it should not demand any copyright whatsoever concerning official statistics. Official statistics have been paid for over the Government budget on behalf of the people, and there should not be any further charges for such data/metadata as such. However, if Statistics Sweden adds value to the basic, official products, and if it obtains financing for such added value from other sources than the Government budget, then it should be entitled to the copyright.

Question 2 is much more complicated. Quotation rules and immaterial author's rights are relatively straightforward as long as traditional, printed publications are concerned. If a part of publication has been exactly quoted, and if the quotation does not change its meaning by being taken out from its context, the source of the quotation could and should be stated, and the agency responsible for the quoted part may rightfully be held responsible for it.

In connection with electronic sources, there is typically not such a relatively simple, one-to-one correspondence between the source, the quotation, and what is finally published by the user of the electronic source. For example, in a macrodatabase statistics are usually not stored as printable tables but rather as multidimensional data structures associated with metadata that may be stored elsewhere. The agency responsible for such a database cannot control exactly what data/metadata combinations that the user will actually retrieve, and even less the agency will be able to control the interpretation, editing, and analyses that the user will do before exposing results to others.

There are no ready solutions yet to these problems. By and large, the most serious problem facing a statistical agency seems to be to protect itself against being quoted as the source, in situations where it would not like to be implied as responsible for results more or less directly derived from data/metadata in a dissemination database under its control and responsibility.

### *3.8. Towards a viable data/metadata infrastructure in statistical systems*

A system of data/metadata databases can only be viably operated and maintained if it is supported by a well integrated infrastructure, consisting of production systems that more or less automatically feed the dissemination databases with data and metadata.

More concretely, Statistics Sweden has decided that the observation registers that are produced by the local, survey-based production systems should be physically stored as SQL-databases that are technically available via the internal network to authorised users elsewhere in the organisation, including authorised users of the dissemination databases. Thus observation register microdata need not necessarily be physically replicated in the dissemination databases.

In contrast, it is foreseen that most of the macrodata contents of the dissemination databases will actually be physically stored in the dissemination databases. However, the "product" which is responsible for producing certain official statistics, to be made available through the dissemination databases, will also be responsible for delivering data and metadata to the dissemination databases in accordance with agreed standardised procedures and formats. Thus the regular updating of the dissemination databases is anticipated to be an automatic process. A central unit will check that agreed standards are adhered to.

#### **4. References**

Ad hoc Committee on Databases (1994). Statistics Sweden's Future Database Service - Final Report. Stockholm: Statistics Sweden.

Dean, P. and Sundgren, B. (1996). Quality Aspects of a Modern Database Service. In Proceeding of Ninth Intn Conf. on Scientific and Statistical Database Management.

Dippo, C. (in press) Survey Measurement and Process Improvement: Concepts and Integration. In Survey Measurement and Process Quality, (eds) L.Lyberg, P. Biemer, M. Collins, E.DeLeeuw, C. Dippo, N.Schwarz, and D. Trewin. New York: John Wiley.

Malmberg, E. and Sundgren, B. (1994). Integration of Statistical Information Systems - Theory and Practice. Proc of the 7th Int Conf on Scientific and Statistical Database Management, Charlottesville, VA, USA.

Shoshani, A. (1982). Statistical Databases: Characteristics, Problems, and Some Solutions. Proc of the 8th Int Conf on Very Large Databases.

Statistics Sweden (1994). Quality Definitions and Recommendations for Quality Declarations of Official Statistics. Stockholm: (MIS 1994:3).

Sundgren, B. (1993). Guidelines on the Design and Implementation of Statistical Metainformation Systems. Statistics Sweden R&D Report 1993:4. ECE Work Session on Statistical Metadata, November 1993, revised versions 1994 and 1995.

Sundgren, B. (1995). Making Statistical Data More Available. Proceedings of the 50th Session of the International Statistical Institute, Beijing.

# Förteckning över utkomna R&D Reports

R&D Reports är en för U/ADB och U/STM gemensam publikationsserie, som 1988-01-01 ersatte de tidigare "gula" och "gröna" serierna. I serien ingick fram till årsskiftet 1992-93 även **Abstracts** (sammanfattning av metodrapporter från SCB).

## Reports published during 1996:

- 1996:1      On Sampling with Probability Proportional to Size (*Bengt Rosén*)  
(grön)
- 1996:2      Bortfallsbarometern nr 11 (*Antti Ahtiainen, Stefan Berg, Margareta Eriksson, Åsa Greijer, Dan Hedlin, Monica Rennermalm, Anita Ullberg*)  
(grön)
- 1996:3      Regression Estimators in Theory and in Practice (*Tomas Garås*)  
(grön)
- 1996:4      Quality Aspects of a Modern Database Service (*Pat Dean and Bo Sundgren*)  
(gul)

Tidigare utgivna *R & D Reports* kan beställas genom Ingvar Andersson, SCB, U/SIB, Box 24300, 115 81 STOCKHOLM (telefon 08-783 41 47, fax 08-783 45 99, e-post [ingvar.andersson@scb.se](mailto:ingvar.andersson@scb.se)).

*R & D Reports* listed above as well as issues from 1988-1994 can - in case they are still in stock - be ordered from Statistics Sweden, att. Ingvar Andersson U/SIB, Box 24300, S-115 81 STOCKHOLM (telephone +46 8 783 41 47, fax +46 8 783 45 99, e-mail [ingvar.andersson@scb.se](mailto:ingvar.andersson@scb.se)).