Abstract

Documentation describes something, e.g. a physical site, a process, or a set of data. In connection with official statistics, documentation describes a statistical information system, a statistical process, or a set of statistical data (macrodata or microdata).

There is a triple relationship between documentation and quality. Firstly, documentation is an element of quality. If we want to discuss the quality of something, we need some kind of description of that “something” to start from. Secondly, a documentation could describe the quality of something, e.g. the quality of some official statistics, or some statistical data. Thirdly, a documentation is itself associated with quality. For example, a documentation may be more or less relevant, complete, accurate, and easy to understand. Thus a documentation has quality.

Both aggregated statistics (so-called macrodata) and observation data (so-called microdata) are data. Documentations of statistical data (macrodata and microdata) “data on data”, so-called metadata. Descriptions of processes may also be metadata, since the description of a process may often indirectly describe data emanating from the process as well. For example, a sampling procedure determines the precision of estimates that are based upon data from the sampling procedure.

There are four major purposes of documentation in connection with official statistics:

- Communication of meaning and quality of aggregated statistics. What is the meaning of these statistics? How uncertain are these figures? Are these statistics from country A comparable with these from country B, and if not, how can they be made more comparable? Which methods of statistical analysis can be applied to these statistics, and how should the results be interpreted?

- Communication of meaning and quality of statistical microdata. Can these microdata, originally collected for purpose A, be reused for purpose B? Can these data be combined with those? Is it possible to perform a longitudinal study on available yearly generations of these microdata?

- Communication of specific knowledge about a survey or a survey system, e.g. for training purposes in connection with new recruitments, or in connection with auditing activities. How is the labour force survey conducted in Sweden? How do you calculate uncertainty measures in this survey?

- Communication of general knowledge about a general statistical process or a general statistical information system. Which are the current best methods and practices for (i) frame construction; (ii) data editing; (iii) confidentiality control; (iv) coding; (v) non-response compensation? Which is the current best method for designing a data warehouse for the statistical system of a country?

The communication in connection with documentations is usually indirect, since the sender and the receiver are typically separated both in time and space. This requires a documentation to be formulated in a relatively well defined language, which is well understood by both the sender and the receiver.

The paper discusses conceptual issues as well as more practical aspects of how to structure and organise the documentation work and metadata work of a statistical organisation. The systems used by Statistics Sweden for documentation and quality declaration are presented in some detail. The problem of finding a feasible implementation strategy for documentation and metadata systems is addressed, as well as the topic of process data. The management of process data may hopefully soon become a natural part of the integrated data/metadata management systems that we now at last see emerging in practice.
1 Documentation, metadata, and quality

Documentation describes something, the documentation object. The object of documentation may be, for example, a physical site, a process, or a set of data. A documentation contains information about the documentation object. If the documentation object is itself information or something closely related to information, e.g. an information process, the documentation contains metainformation. Thus for example, a documentation of a statistical survey contains metainformation, whereas a documentation of a city’s water supply system contains information. A documentation of a metainformation system contains meta-metainformation, etc.

A documentation is physically represented in some way, e.g. by means of a printed or an electronic document, or by means of data in a database. A documentation may be more or less structured and more or less formalised. A documentation represented by an unstructured Word document may be an example of an unstructured an informal documentation, and a documentation represented by a set of relational tables in a database may be an example of a structured and formalised documentation. There are many semi-structured and semi-formalised variants in between. For example, documentation templates and checklists are often used to give some structure to a documentation.

If the documentation contains metainformation, the representation is called metadata. Documentations and metadata are aimed at different categories of users. There is a basic distinction between two categories of users: human users and computer software. Computer software requires metadata to be relatively structured and formalised, whereas human users, e.g. producers of statistics, often prefer a less structured and more informal documentation. If the operations of a software product are completely controlled by metadata, it is said to be metadata-driven.

There are obvious but rather complex relationships between documentation and quality. Firstly, according to most definitions of quality, e.g. ISO standards, a documentation is a necessary prerequisite for quality. We cannot even discuss the quality of something, if we do not have a documentation of that something. Thus documentation is an element of quality.

Secondly, a documentation is needed in order to describe the quality of something, e.g. a statistical survey. Thus a documentation describes quality.

Thirdly, a documentation is itself associated with quality. For example, a documentation may be more or less relevant, complete, accurate, and easy to understand. Thus a documentation has quality.

The three relationships between documentation and quality are visualised in figure 1.

**Figure 1.** Three different relationships between documentation and quality.

What has here been said about documentation and quality can be easily transferred to the topic of metadata and quality.
Finally it should be noted that there are two basic interpretations of the term “quality”, one which is value-neutral, and one which is not. According to the former “quality” is more or less a synonym of “property” and “characteristic”, and according to the latter “quality” means “good quality”. Often the two aspects are mixed in one and the same context. For example, a quality declaration of statistical data may contain different parts that describe the contents, the accuracy, and the availability of the data. The accuracy of statistical data may usually be measured independently of the intended use of the data. Moreover, at least some components of the accuracy, e.g. the sampling error, may even be measured in an objective and quantitative way. Thus it may be possible to talk about “good” or “bad” accuracy of statistical data, even if we do not know what the data will be used for. The same is true for many aspects of the quality component “availability”. However, the situation is different for the quality component “contents”, or “relevance”. The producer of statistical data can at best describe the contents of a set of statistical data in such a way that a user can himself/herself determine whether the statistical data can be used for a particular purpose. The same set of data may be relevant for one purpose and irrelevant, or at least less relevant, for another purpose. Only the user can make the final judgment.

2 A general model of statistics production

Figure 2 visualises a statistics production process, or a set of statistics production processes that make up a statistical system.

There is a major loop in the visualisation model, consisting of three main phases in statistics production:

- design and planning
- operation
- evaluation

During the evaluation phase the experiences of the previous design and operation are evaluated and fed back into the next round of (re)design and operation.

Each one of the major phases is internally structured into steps that reflect the inherent structure of a classical statistical survey:

- objectives and contents
- frame and frame procedure
- sampling procedure (if applicable)
- measurement, data collection, and data preparation
- aggregation and estimation
- presentation and dissemination

Traditionally these steps have been executed more or less sequentially. With today’s technology it is often feasible and efficient to integrate several consequtive steps, e.g. measurement, data collection, and data preparation, and it is not always necessary to finalise one step before starting the next. For example, it may be efficient to “take a quick look ahead” by aggregating and estimating on the basis of incomplete and possibly erroneous data, in order to make intelligent decisions on further actions as regards follow-ups on non-response and suspected errors.

---

1 Data preparation includes data entry, coding, editing, correction, and imputation.
Figure 2. Major phases and steps of a statistical production process or a statistical system.
Today we have many surveys\(^2\) that are more or less continuously on-going, in contrast to the traditional surveys that were either executed just once, one-time surveys, or surveys that were repeated at regular intervals, repetitive surveys. A register is an example of an on-going survey, where there is a more or less continuous stream of input transactions reporting about births and deaths and other important status changes of the objects in the register. The input transactions may either be accumulated over time, or they may only be used for updating the current status of the register. In the former case we have the possibility to make retrospective and longitudinal studies, in the latter case we are limited to cross-sectional analyses. Often register systems are designed to combine the two aspects: the historical aspect and the updated-status aspect.

Each phase and step in a statistics production system uses and produces data and metadata. In the production model suggested by figure 2, the data and metadata are assumed to be organised in databases. Some of the databases are global in the sense that they are available to the whole statistical system under consideration, others are local in the sense that they are only available to a particular production system, or a small set of related production systems.

### 3 Major purposes of documentation in connection with statistics

There are three major categories of users of documentation and metadata in connection with statistics production:

- users of end-products from statistics production
- producers of statistics
- computer software used for statistics production

There are basically two different kinds of end-products from a statistical production process:

- statistics, that is aggregated statistical data – macrodata
- final observation registers from which the statistics have been produced – microdata

There are many different categories of users of statistical macrodata and microdata. Some important categories are indicated in figure 3. Different users have different, and sometimes even contradictory needs as regards data and metadata. Some users are future users, living in a future society without any possibilities to ask the original producers of archived statistical data about advice, living in an environment that may be completely different from our present society. Such users must have access to very rich documentations if they are to be able to interpret the data that we have left for them.

Whereas researchers usually have plenty of time, motivation, and knowledge to use good documentations, other categories of users have completely different needs and competence profiles. A journalist has little time and is pressed to present interesting and clear messages to the audience. A broker on the stock market must interpret new statistical figures within a fraction of a second when making buy and sell decisions. A politician must be able to interpret the meaning of statistical figures in terms of how it affects the different groups of citizens, different regions, and different categories of business. The man in the street should be able to understand statistics when they are used in the public debate, as part of democratic processes.

---

\(^2\) It may be debated whether an ongoing survey is really a survey, or if it should be called something else. However, from a theoretical point of view it is attractive to define “survey” as a generic concept covering one-time surveys, repetitive surveys, registers, on-going surveys, longitudinal surveys, secondary surveys (e.g. national accounts), etc.
The message of figure 3 is that it may not be the best idea to try to satisfy all these different and contradictory user needs by means of a compromise, that is, by means of a single data/metadata system that satisfies everybody – and nobody. Instead the solution may be to build an infrastructure that may be used as a basis for a large number of different applications and user-interfaces that are tailored to the needs of different users and user categories.

Schematically speaking, a user of statistical data needs documentation/metadata in order to

- identify potentially relevant data for his/her question/problem
- determine whether potentially relevant data are actually of interest
- locate interesting data
- retrieve the data
- interpret the data “correctly”
- process and analyse the data further

Important tools for finding relevant data are thesauruses and classification databases with references (links) to surveys and data sets with their documentations. When it comes to “correct” interpretations, that is, interpretations that are consistent with the *de facto* definitions that were actually used when the observation data were collected, and with the rules and procedures that *de facto* controlled the statistical processes leading to the outputs, detailed documentations are indispensable for serious users like social researchers. The documentations must not only cover the output data as such; they must also give detailed accounts of the processes and input data behind the outputs. For example a user must know which assumptions were made by the statistics producer as regards non-respondents. An advanced user of statistical data must also know about coding and editing rules and practices.

For more casual and inexperienced users there are other requirements as regards documentation and metadata. These requirements are not necessarily easier to satisfy than the needs of advanced users. For these users the accompanying documentation and metadata must be designed and presented on the assumption that the user will not be particularly aware of the need to consult human or computerised expertise and knowledge before making far-reaching interpretations and
analyses of statistics. Comments, footnotes, and outright warnings are tools that are often used by the producers in this connection, but there is a risk that such tools are more efficient in silencing the producer’s own conscience than in actually improving the user’s proper understanding of the statistical outputs. More innovative methods are needed to encourage journalists and other users, who are in a hurry, to dig a little deeper into the meaning of the figures presented by statistics producers. One possibility is to make everything “clickable”, figures as well as text labels in statistical tables. Another possibility is to prepare some analyses that a user can use directly as they are or at least use as patterns for their own analyses, hopefully leading the users in the right direction and helping them to avoid the worst pitfalls. Wizards guiding the user during interpretation and analysis is yet another possibility. All these tools require good documentation and well organised metadata in order to work properly.

Producers of statistics need documentation and metadata for the following purposes:

- production control
- training of new staff
- evaluation of existing processes (requires process data, a particular kind of metadata)
- maintenance and redesign of processes, both manual and computerised
- designing new surveys, using components, tools, and experiences from existing ones, as well as general knowledge from CBM documents, text-books, and expert systems

Computer software require metadata for tasks like

- automatic control and transfer of control on the basis of user requests
- automatic presentation, restructuring, and reformatting (e.g. pivot functions)
- automatic optimisation and reorganisation of the technical system

At least conceptually the knowledge, documentation, and metadata needed by users, producers, and software make up the knowledge base of a statistical system as visualised by figure 2 above.

4 The SCBDOK documentation model

SCBDOK is a documentation model that specifies the documentation basis needed for the different purposes mentioned in the previous chapter of this paper. The contents of an SCBDOK documentation should be sufficient for all purposes mentioned there, but the presentation form may have to be further developed, in some cases much further developed, in order to meet the needs of such user categories as journalists and “the man in the street”.

The SCBDOK model is based on a general description of the statistics production process; see Rosén & Sundgren (1991). This description is basically a verbal text that introduces and defines all statistical concepts needed in their natural context. A recommended term is given for each concept, but many synonyms are also mentioned and allowed in order to avoid destructive quarrelling on terms; it is the concepts that are important and that have to be well understood and agreed upon.

The SCBDOK description model follows basically the process structure and contents of figure 2 above. The major documentation items are organised in the form of a numbered list making up the SCBDOK documentation template. Version 3.0 of this template is shown in figure 4. The numbered list consists of two levels. Below the second level there are unnumbered checklists of more detailed information to be included on that level.

3 Current Best Methods
The major contents of an SCBDOK documentation could be described as follows; cf figure 2:

Chapter 0 contains administrative information and is aimed at managers and others who need rather superficial information about the survey. Chapter 0 is also used separately from the complete SCBDOK documentation, and it is then called “product description”.

Chapter 1 gives an overview of the contents of the documented survey in a structured way. The statistical observation characteristics and their relations to each other are specified by means of an object graph, and the statistical target characteristics are specified by means of \( \alpha \beta \gamma \tau \)-structures (see section 5.2.4 and figure 12 later in this paper). The outputs from the survey are specified, both microdata (final observation registers) and macrodata (statistics). Finally, references are made to other relevant sources of documentation and metadata, e.g. methodological reports.

Chapter 2 gives a detailed description of the data collection process, including frame and frame procedure, sampling procedure (if applicable), measurement instruments, the data collection procedure proper, and data preparation procedures (data entry, coding, editing and correction, imputation, production of derived objects and variables). The detailed information should include all rules, instructions, and practices that have an impact on the meaning and quality of the data.

Chapter 3 describes both the contents and the storage and other technical aspects of the final observation registers. There are often several versions of the final observation registers, possibly with different contents and managed by different software. It is important to distinguish between production versions, e.g. versions managed by a database management system, and those versions that are submitted to an archive for reuse in the near or (very) distant future. The latter versions of the final observation registers must be stored in such a way that they will last for a long time and can be reused without access to the hardware and software that we use today.
Chapter 3 also contains an item for so-called process data, that is, metadata that describe circumstances and events that are unique for each repetition of the survey and its processes, e.g. data about response and non-response.

Major parts of chapter 3 are supported by a software tool called METADOK that ensures that the metadata that are captured in this part of the documentation are structured and formalised in such a way that they can easily be used by software used in statistics production. Among other things this facilitates the implementation of metadata-driven statistics production systems.

Chapter 4 describes the estimation procedures, including mathematical formulas and assumptions made, as well as presentation and dissemination procedures.

 Chapters 1, 2, and 4 contain the documentation basis needed for the production of so-called quality declarations that (together with the product descriptions mentioned above; chapter 0 in SCBDOK) are mandatory for published official statistics in Sweden, regardless of whether they are published electronically or by means of traditional paper publications. Figure 5 shows the quality declaration template.

<table>
<thead>
<tr>
<th>Quality Declaration Template</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Contents</td>
</tr>
<tr>
<td>1.1 Statistical target characteristics</td>
</tr>
<tr>
<td>1.1.1 Objects and population</td>
</tr>
<tr>
<td>1.1.2 Variables</td>
</tr>
<tr>
<td>1.1.3 Statistical measures</td>
</tr>
<tr>
<td>1.1.4 Study domains</td>
</tr>
<tr>
<td>1.1.5 Reference time</td>
</tr>
<tr>
<td>1.2 Comprehensiveness</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3 Timeliness</td>
</tr>
<tr>
<td>3.1 Frequency</td>
</tr>
<tr>
<td>3.2 Production time</td>
</tr>
<tr>
<td>3.3 Punctuality</td>
</tr>
<tr>
<td>5 Availability and clarity</td>
</tr>
<tr>
<td>5.1 Forms of dissemination</td>
</tr>
<tr>
<td>5.2 Presentation</td>
</tr>
<tr>
<td>5.3 Documentation</td>
</tr>
<tr>
<td>5.4 Access to microdata</td>
</tr>
<tr>
<td>5.5 Information services</td>
</tr>
</tbody>
</table>

**Figure 5. The quality declaration template used in the Swedish statistical system.**

Chapter 5 of the SCBDOK template gives a detailed technical description of the production process. It aims only at the needs of the producers; all documentation that is needed by users should be included in the other chapters. Chapter 5 is at present under redesign by the IT unit of Statistics Sweden. The purpose of the redesign is to make it more harmonised with the hardware/software infrastructure and new work procedures that have been introduced in connection

---

⁴ Actually the present version of the Quality Declaration Template uses the term "unit" here. In order not to confuse the reader of this paper and many other papers concerning SCBDOK, we will use the standard term "object" here. In SCBDOK the term "unit" stands for "measurement unit", e.g. US dollars, tons.

⁵ Also called "sources of error".

⁶ The term used elsewhere in this paper and in other papers concerning SCBDOK is "data preparation".

9
with the so-called platform shift (moving from mainframe to PCs in network) that was finalised by the end of 1999 at Statistics Sweden.

Chapter 6 is called the logbook and is intended to help the users to quickly identify what changes have been made to the survey since last time it was executed. Each time a change is made to the design of a survey, the documentation should be updated, but an entry informing about the change should also be made in the logbook. This facilitates comparisons over time.

5 An object-oriented approach to SCBDOK

SCBDOK is a relatively informal documentation model. When it was designed, priority was given to capturing the necessary information about surveys and their outputs in one form or another, rather than enforcing very structured and formalised description methods. Whereas IT people are used to, and appreciate, structured and formalised descriptions, statisticians and subject matter specialists often prefer to describe things by means of plain text, using an ordinary word processor.

The only structure imposed by SCBDOK is the two-level hierarchy of documentation items prescribed by the documentation template in figure 4. As was already mentioned there are checklists of topics that must be covered under each documentation item. However, it is not prescribed in detail how these topics should be covered, or even in which order they should be covered.

With growing documentation experience among the producers, a demand for computerised tools has arisen. One reason for this is that producers begin to see possibilities to use the metadata contained in the documentations for additional purposes. For example, formalised metadata concerning the final observation registers can also be used for driving software products, provided that the software can obtain metadata from the documentations through some kind of metadata interface. Another example is that producers want to make use of metadata that already exist, when they prepare the documentation of a survey. For example, the value set of a variable in a survey may be based on a standard classification, available from a classification database.

When using a documentation system like SCBDOK in a software environment, it becomes desirable to formalise the documentation model according to object-oriented principles. As regards SCBDOK such an attempt was made as part of the EU-funded research project Integrated MetaInformation Management (IMIM), and another one is underway in the new EU-funded research project METAWARE.

5.1 Overview of the object model of SCBDOK: object graph

This section will outline the object structure for the description model underlying SCBDOK. The main documentation object of SCBDOK is the documented survey itself. “Survey” is a generic concept in SCBDOK. As is shown by figure 6, it covers a large variety:

- the traditional one-time survey, as treated in text-books on statistics
- the repetitive survey, or survey series, so common in official statistics
- on-going surveys, to which events (births, deaths, changes of status) are reported more or less continuously, e.g. registers
- longitudinal surveys, maintaining the history of individual objects
- secondary surveys that integrate data from other surveys, e.g. consumer price index, national accounts.
The so-called integration registers represent a new type of survey that has become very popular recently, especially among researchers. Like secondary surveys they are based upon other surveys, and like longitudinal studies they maintain life histories of individual objects.

Figure 7 visualises the major steps in statistics production, as reflected in the model in figure 2 earlier in this paper, as well as in the description model underlying SCBDOK itself; cf the major chapters in the SCBDOK documentation template in figure 4 above.

Figure 8 combines the models in figure 6 and figure 7. It should be interpreted in the following way. Regardless of which type of survey that is actually the documentation object at hand, it is possible to identify the same seven production steps. However, the interpretations may differ between survey types. For example, what is the frame of a register, which may itself be a frame? And what is the final observation register of a continuously on-going survey?

In some situations it makes sense to regard a register as its own frame. This is the case, when a register is established “from scratch” by means of direct observations and is then used both as a frame for other surveys and for the production of some basic statistics about the registered objects. In other situations, the register is created from another register, e.g. an administrative register. In such a case, the administrative register is the frame of the statistical register.

In a way, an on-going survey is its own final observation register. However, it usually makes sense to “freeze” the observation register of an on-going survey at regular intervals, e.g. monthly or yearly, and to regard the frozen copies as the final observation registers to be archived etc.

The survey as a whole is an example of a documentation object that is itself a process. We have already encountered a number of subprocesses of the survey process that are also documentation objects: frame management and sampling, data collection, estimation, and dissemination. These processes may be further subdivided into procedures that are documentation objects as well: frame procedure, sampling procedure, measurement procedure, data preparation procedure, etc.

We have also identified a number of information and data sets that are documentation objects: contents, final observation register, and final statistics. These “crude” objects may be broken down into finer ones: observation characteristic, statistical target characteristic, population, object type, variable, value set, value, frame, frame link, information source, observation object (in sample), target object (in sample), questionnaire, question, response, dictionary, thesaurus, classification database, register version, database, data matrix, etc.

The most important documentation objects in connection with a statistical survey, and their most important relations to each other, are visualised in figure 10.

Above the level of statistical surveys we find different levels of statistical systems, e.g. groups or programs of related statistical surveys, the system of official statistics of a country, a region, or a group of countries, e.g. EU. Statistical systems are also important documentation objects, especially when quality aspects like comparability and coherence are discussed.
Figure 6. The statistical survey – a generic concept in SCBDOK.

Figure 7. Major steps in statistics production.

Figure 8. A combination of figure 6 and figure 7.
Figure 10. Statistical surveys: important documentation objects and their relations to each other.
5.2 Some key concepts

Some of the concepts introduced in figure 10 may require some explanation. The following key concepts will be explained and illustrated in this section:

- statistical characteristics
- estimates of statistical characteristics
- uncertainties of estimates of statistical characteristics
- estimates of the uncertainties of statistical characteristics
- structured sets of statistical characteristics and statistics
- frames and frame procedures

5.2.1 Statistical characteristics

A statistical characteristic $S$ is a statistical measure, $f$, applied to the values of a variable, $V$, of the object instances in a set of objects, $O$, in order to summarize some aspect of those values:

$$ S = O.V.f $$

**Example 1:** “Average yearly income during 1999 for the persons who were registered in Stockholm at the end of the year 2000.”

Here the set of objects, $O$, usually called the population, is “the persons who were registered in Stockholm at the end of the year 2000”, the variable, $V$, is “yearly income during 1999”, and the statistical measure, $f$, is “average”.

**Example 2:** “Number of persons registered in Stockholm at the end of the year 2000.”

Here the statistical measure is a function, “count”, that summarizes “the frequency aspect” of the objects directly, not via any particular variable. Thus “count” is a statistical measure with zero arguments, whereas “average” in the previous example had one argument.

**Example 3:** “The correlation between age at the end of the year 2000 and yearly income during 1999 for persons registered in Stockholm at the end of the year 2000.”

Here the statistical measure is “correlation”, a function with two arguments. Thus we have to generalise the concept of a statistical characteristic by allowing the “$V$” in “$S = O.V.f$” to be a vector of variables.

5.2.2 Estimates of statistical characteristics

If it were possible to make perfectly correct observations of exactly those objects that are in the population aimed for, the target population, then it would be possible to obtain perfectly correct values of statistical characteristics for this population; one would just have to apply the appropriate statistical measures correctly. In practice this “ideal procedure” for computing statistical characteristics is hardly ever possible to implement. Some important reasons for this are:

1. One cannot identify and localise exactly those objects that are in the target population.

Typically one uses some kind of list or register, called the frame, in order to find the objects

---

7 Complete explanations of all concepts used in SCBDOK can be found in Rosen & Sundgren (1991).
concerned. The set of objects that the frame leads to is called the frame population. The frame population may differ from the target population by containing objects that are not part of the target population, over-coverage, and by not containing objects that are part of the target population, under-coverage.

2. One cannot afford to investigate/observe all objects in the target population. This can lead to a sample survey instead of a complete survey.

3. Regardless of whether one goes for a sample survey or a complete survey, one will usually not succeed in observing all objects and all variables aimed for; total or partial non-response.

4. The observations that are actually made will be subject to errors and uncertainties of different kinds, measurement errors, processing errors, etc, that is, one will not always be able to obtain the true values of the variables to be summarized by the statistical measures.

5. Sometimes it is not possible to observe the target objects and/or the target variables directly. Instead one may observe them indirectly, either by using other sources, like other surveys or administrative registers, or by observing related objects and/or related variables and deriving the objects and variables aimed for. This procedure may also lead to over-coverage, under-coverage, non-response, and measurement errors.

Thus instead of the target population O and the target variables V, aimed for, one will in practice obtain an actually observed set of objects O’, differing from O because of over-coverage, under-coverage, sampling, and total non-response, and an actually observed variable V’, differing from V because of total and partial non-response, measuring errors, and processing errors.

Hence one has to be satisfied with approximations, estimates, of the true values of the statistical characteristics aimed for. The estimations have to be based on the incomplete, erroneous, and uncertain observations that one has been able to make, directly or indirectly.

When estimating the true value of a statistical characteristic, $S = O \cdot V \cdot f$, on the basis of an actually observed set of objects, O’, and actually obtained, processed, and registered values of a variable, or variable vector, V’, one has to apply a function $f'$, the estimator, which is somehow related to, but usually not identical with, the statistical measure $f$. The difference between $f'$ and $f$ is the result of an attempt to compensate for the deviations of O’ and V’ from the ideal O and V.

In summary, the basic idea of a statistical survey, in a broad sense (including surveys based on administrative registers), implemented by means of a statistical production system, is to

- estimate the true values of statistical characteristics, $O \cdot V \cdot f$,
- on the basis of observed values of one or more observation characteristics, $O' \cdot V'$,
- by applying an estimator, $f'$, on the observed values,
- thus computing $O' \cdot V' \cdot f'$.

See also figure 11.

5.2.3 Estimates of the uncertainties of estimates

It follows from what has been said, that estimates of statistical characteristics, and hence statistics as such, are subject to uncertainties of different kinds. One can try to decrease these uncertainties by improving the statistical production processes, including the estimation procedures, but some uncertainties will always remain. At best the uncertainties can themselves be estimated, once again with some uncertainty, of course. The estimates and even more verbal descriptions that one can produce in order to describe and quantify the uncertainties of estimates of statistical characteristics may form the basis for quality declarations of the produced statistics.
A function that, on the basis of the true values of the incomes during 1999 for the persons who were actually registered in Stockholm at the end of 2000, computes...

...the true value of the statistical characteristic "Average income during 1999 for the persons who were actually registered in Stockholm at the end of 2000"

A function that, on the basis of the registered and "cleaned" values (and metadata concerning these values) computes...

...the estimated value of the statistical characteristic "Average income during 1999 for the persons who were actually registered in Stockholm at the end of 2000"

...the uncertainty in the estimate, that is, the deviation of the estimated value from the true value of the statistical characteristic

Figure 11. The concepts of “statistical characteristic” and “estimation of a statistical characteristic”.
5.2.4 Structured sets of statistical characteristics and statistics: the \( \alpha\beta\gamma\tau \)-model

A typical statistical table contains statistics concerning a family of statistical characteristics, where the family members are usually related in a typical way.

**Example 4.** Consider the following statistics: “Average yearly income during 1999 for the persons who were registered in Stockholm at the end of the year 2000: by sex and age group.”

Suppose that there are two sexes, male and female, and three age groups, young, middle-aged, old. Then Example 4 specifies at least six statistics in addition to the statistic in Example 1 earlier in this paper. This kind of structuring of statistics, and underlying statistical characteristics, corresponds to the \( \gamma \)-dimension of the \( \alpha\beta\gamma\tau \)-model.

Figure 12 illustrates the following discussion of the four major dimensions in the \( \alpha\beta\gamma\tau \)-model. It contains a matrix, a so-called \( \alpha\beta\gamma\tau \)-matrix, with different columns for the major dimensions, as well as for the major components of the concept of a statistical characteristic.

*The \( \alpha \)-dimension*

The \( \alpha \)-dimension of the matrix contains the populations of the statistical characteristics. In figure 12 all three statistical characteristics (S1, S2, S3) refer to one and the same population:

\( \alpha1: \) persons registered in Sweden at the end of \( t \)

Note that “O” in “O.V.f” denotes alternatively the whole \( \alpha \)-population, \( O^\alpha \), or each one of the subdomains of interest, \( O^\alpha \) by \( V^\gamma \), or \( O^\alpha \) by \( V^\gamma \) defined by the classification \( O^\alpha \) by \( V^\gamma \) where \( V^\gamma \), is a crossclassification of \( n \) variables, the \( \gamma \)-variables: \( V^\gamma = V^{\gamma1} \times V^{\gamma2} \times ... \times V^{\gammam} \). See also the discussion of the \( \gamma \)-dimension below.

*The \( \beta \)-dimension*

The \( \beta \)-dimension contains the summarizing functions of the statistical characteristics, that is, statistical measures that are applied to zero, one, or more variables, the \( \beta \)-variables. Thus a statistical measure may have zero, one, or more arguments. Some examples:

- **count** counts the number of object instances in \( O \), a function with zero arguments
- **sum** summarizes the values of a variable \( V \), a function with one argument
- **average** averages the values of a variable \( V \), a function with one argument
- **correlation** computes the correlation between two variables, \( V1 \) and \( V2 \), thus two arguments
- **percentage** computes the percentage of object instances in \( O \) satisfying a Boolean variable \( V \), a function with one variable

The average function can alternatively be expressed as a sum divided by a count, and a count can alternatively be expressed as a sum of a variable that takes the value 1 for all objects in \( O \). Correlations and percentages can also be expressed in terms of other functions. The following \( \beta \)-variables appear in figure 12:

- \( \beta1: \) income\((t-1)\): a person’s income during \( t-1 \) according to taxation performed during \( t \)
- \( \beta2: \) age\((t)\): a person’s age in whole years at the end of \( t \)
### Table: STATISTICAL CHARACTERISTIC, REFERENCE TIME t, SET OF OBJECTS O, SUMMARIZING FUNCTION

<table>
<thead>
<tr>
<th>STATISTICAL CHARACTERISTIC S = O.V.f by variables G</th>
<th>REFERENCE TIME t</th>
<th>SET OF OBJECTS O</th>
<th>SUMMARIZING FUNCTION β-dimension</th>
<th>VARIABLES V</th>
<th>STATISTICAL MEASURE f</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1: “Average income during the year t-1 for those persons who were registered in Sweden at the end of the year t: by commune, sex, and age.”</td>
<td>Year t = 1995, 1996, ...</td>
<td>persons registered in Sweden at the end of t</td>
<td>• commune(t): the commune where the person was registered at the end of t&lt;br&gt;• sex(t): the person’s sex at the end of t&lt;br&gt;• age(t): the person’s age in whole years at the end of t</td>
<td>• income(t-1): the person’s income during t-1 according to taxation performed during t</td>
<td>average</td>
</tr>
<tr>
<td>S2: “Number of persons registered in Sweden at the end of the year t: by sex, age, and income bracket.”</td>
<td>Year t = 1995, 1996, ...</td>
<td>persons registered in Sweden at the end of t</td>
<td>• sex(t): see above&lt;br&gt;• age(t): see above&lt;br&gt;• income bracket: the person’s income bracket according to classification xxx, based on the person’s income during t-1</td>
<td></td>
<td>count</td>
</tr>
<tr>
<td>S3: “The correlation between age at the end of the year t and income during the year t-1 for persons registered in Sweden at the end of the year t: by commune and sex.”</td>
<td>Year t = 1995, 1996, ...</td>
<td>persons registered in Sweden at the end of t</td>
<td>• commute(t): see above&lt;br&gt;• sex(t): see above</td>
<td>• age(t): the person’s age in whole years at the end of t&lt;br&gt;• income(t): the person’s income during the year t-1</td>
<td>correlation</td>
</tr>
</tbody>
</table>

**Figure 12. αβγτ-matrix.**

The following summarizing functions are formed by applying statistical measures to the β-variables:

S1: \text{average}(β_1) or, with dot notation, β_1.average

S2: \text{count}

S3: \text{correlation}(β_1, β_2) or (β_1, β_2).correlation

**The γ-dimension**

The γ-dimension contains variables that crossclassifies the population into domains of interest\(^8\), to which the statistical measures are applied in the same way as they are applied to the cross-classified population itself. In figure 2 the following cross-classifications occur:

S1: The population α_1 is crossclassified by

γ_1: commute(t): commune where a person was registered at the end of t
γ_2: sex(t): a person’s sex at the end of t
γ_3: age(t): a person’s age in whole years at the end of t, that is, =β_2

S2: The population α_1 is crossclassified by γ_2, γ_3, and

---

\(^8\) In the Quality Declaration templet (see figure 5 above) domains of interest are called study domains.
$\gamma_4$: income_bracket(t-1): a person’s income bracket according to a classification xxx, based upon the person’s income during t-1

S3: The population $\alpha_1$ is crossclassified by $\gamma_1$ and $\gamma_2$.

Note that the $\gamma$-variables form subdimensions of the $\gamma$-dimension.

The $\tau$-dimension

The $\tau$-dimension specifies reference times for the statistical characteristics. Time can be explicitly specified for all populations, variables, etc, but this is often unpractical. Instead a time parameter $t$ is used, and all times are expressed as functions of $t$. The $\tau$-dimension also states the value set of $t$. In figure 12 the $\tau$-dimension specifies the parameter $t$ with a value set consisting of the years 1995, 1996, and onwards.

5.2.5 Frames and frame procedures

When we are about to make observations and collect data to form the basis for our estimations of the values of certain statistical characteristics, we are first confronted with two questions:

1. What, or whom, should we collect data about? (Which are the observation objects?)
2. Whom should we collect data from, and how can we get in contact with them? (Which are the information sources, be they human respondents or other types of data sources, e.g. registers, and how do we find them?)

The tools for finding the answers to these questions are a frame and a frame procedure. In the most simple situation the frame is a list (register) of elements that are one-to-one-related to the objects in the target population. Information about these objects are assumed to be obtainable directly from the target objects themselves. In the case of a sample survey, the sample is first generated by letting a subset of frame elements be selected at random, and by then letting the corresponding target objects be the random sample of target objects to be observed.

In this simple situation we have a one-to-one-correspondence between the four concepts: frame element, target object, observation object, and source of information.

In a more general situation, some or all of the correspondences just mentioned may be more complex, and sometimes even much more complex. In such cases a relatively complex procedure has to be designed that leads from frame elements, via frame links, to sources of information, observation objects, and target objects. For example, the source of information may be an official at a court of justice, the observation object may be a conviction, and the target object may be the convicted criminal. A statistical characteristic to be estimated may be “Number of juveniles who have been convicted for a certain type of crime during a certain year, distributed by sex, age, home commune, birth country of parents”.

The frame procedure is seldom perfect. It will lead to some target objects, which are not part of the target population, over-coverage, and on the other hand it will not lead to some objects that actually are part of the target population, under-coverage. The set of objects “of the same kind” as the objects in the target population that the frame procedure leads to is called the frame population. The intersection between the frame population and the target population is called the attainable part of the target population. See also figure 13.
Figure 13. The concepts of “frame procedure”, “undercoverage”, and “overcoverage”.

Figure 14. Object system structure of the METADOK part of SCBDOK.
The concepts of over-coverage, under-coverage, and attainable part may also be applied to information sources and observation objects in analogy with the description for target objects.

5.3 Structured and formalised metadata

For most parts of an SCBDOK documentation relatively unstructured and informal verbal descriptions are sufficient. Generally speaking such descriptions are often sufficient for human users. For those parts of the documentation that are to be used for generating metadata driving software, a higher degree of structuring and formalisation is necessary. An idea of how this can be done is given by figure 14, which gives an overview of the metadata model used by the above-mentioned METADOK tool for describing the contents and physical representation of a final observation register.

6 Implementation strategy

When a documentation effort starts in a statistical office, there is usually a huge back-log of insufficiently documented surveys and data sets, and the documentations that exist reflect many different terminologies, documentation structures, and focuses of interest. Very much of the documentation needed by advanced users, such as researchers, is still in the heads of the producers, if they still happen to be available. The published statistics, be it by electronic or printed media, are not always accompanied by adequate explanations and quality declarations.

At Statistics Sweden the documentation model SCBDOK was presented in 1991. It was enthusiastically welcomed by users, especially qualified users represented in the scientific council (and somewhat later the database council) of Statistics Sweden. It was not equally easy to gain acceptance among the producers internally in the organisation. After lengthy discussions with representatives of the producers a slightly revised proposal was adopted by the Director General in 1994. The decision by the Director General contains the following main items:

1. Systems for production of official statistics must be documented. The documentation should contain the information specified by the SCBDOK documentation template, and it is recommended that the information is structured according to this template.

2. Observation registers (microdata) emanating from production systems for official statistics, and which are to be preserved for future use, must be documented. The documentation must contain the information specified by the SCBDOK documentation template, and it must be structured according to this template.

An exception from the rules above was stated for production systems developed before 1994. The documentation of those systems was allowed to be postponed until some major revision of the system concerned, e.g. the planned change of platform from mainframe to PC.

At about the same time the Director General also decided that all official statistics should be accompanied by yearly updated product descriptions (according to chapter 0 in SCBDOK) and quality declarations (according to the Quality Declaration Template).

Some years later, in connection with the introduction of web-based electronic publishing, it was decided that all regular reports published by statistics Sweden (in whatever form) should be accompanied by an easily understandable description called “Facts about the statistics”, based upon the quality descriptions but edited for more casual users.
How have these formal decisions worked out in practice?

The decision concerning SCBDOM has not been followed in practice. Responsible producers have claimed that they have not had sufficient time for documentation due to the strict time tables for (a) the above-mentioned change of technical platform that had to be completed by the end of 1999 because of the Y2K problem; (b) the integration of their production systems with Sweden’s Statistical Databases. During the year 2001 there is another attempt to document at least some 20 production systems and final observation registers of top priority, including

- Population
- Population changes
- Population censuses
- Real estates
- Education
- Labour force surveys
- Incomes
- Central business register
- Foreign trade
- Motor vehicles
- Agriculture
- Income and wealth
- Employment
- Education analyses
- Criminality
- Living conditions
- Value added taxes

The decision concerning yearly updated product descriptions and quality declarations has been followed, at least *pro forma*, since they are enforced by law. However, these documentations have not been properly checked by statistical methodologists, and the quality of the documentations is sometimes not acceptable.

Thus we have identified two major problems in connection with the implementation of a documentation standard. One is to motivate the producers to co-operate, and the other is to monitor the quality of the documentations.

The motivation problem calls for both sticks and carrots. An efficient stick, which can be used selectively, is for top management to set operational goals for the documentation work – and to follow up these goals regularly.

Naturally carrots more attractive and probably also more efficient than sticks. However, it is not so easy to find good carrots for the producers. They do not usually feel that they will gain very much themselves from having better documentations. Actually they may loose part of an “information monopoly” that they have as long as people have to come and ask them – in the absence of proper documentations. They also fear that the documentation work will take a lot of time.

During the last few years some producers have themselves proposed that certain parts of the SCBDOM documentation, especially part 3 on final observation registers, should become more structured and formalised and more efficiently supported by software tools. This demand has resulted in the development of the METADOK system that supports the capturing of metadata.
corresponding to this part of the documentation. To the extent that the METADOK system can interface other software products, the metadata can be used by these products, thus reducing the work for the producers to manually feed different software products with metadata. In other words, statistics production may become more and more metadata-driven, an attractive feature for both producers and users.

Another carrot would be to minimise the pure documentation work by integrating it with other tasks, preferably such tasks that the producers experience as being more attractive from their point of view. This can be achieved by making as much as possible of the documentation work a natural spin-off from the survey design and planning process. Each time a design decision is taken, or changed, this decision should also become part of an incrementally growing SCBDOK documentation and an incrementally growing Quality Declaration.

Figures 16-18 give an idea about how this can be done. Figure 16 visualises the SCBMOD model for survey design and based upon the SCBDOK description model for statistical surveys. The picture shows, how the design decisions made during the different design phases step by step produce the contents of different entries in the SCBDOK and QD templates. After completion of the design and planning process, almost all entries in the documentation templates should be filled. The only contents missing should be what is dependent upon experiences from the actual operation of the production process, e.g. non-response rates and other so-called process data.

The integration between survey design and documentation may be brought one step further by introducing a feedback loop from documentation to design as indicated by figure 2 earlier in this paper. This should facilitate the work of the producers/designers. In order to achieve this, one should systematically accumulate and organise the knowledge emanating from

- design and documentation processes in one’s own statistical office as well as in other statistical organisations all over the world
- feedback from users, respondents, interviewers, data preparation staff, etc
- developments of useful tools, wherever they take place: catalogues of variables and statistical characteristics, classification databases, standardisation documents, thesauruses, tools for automatic coding, generalised software packages and components, expert systems
- handbooks and reports about current best practices and current best methods
- scientific literature
- references to human experts all over the world

This knowledge should be made easily available to all producers/designers of statistical surveys and survey systems. With today’s technology intranets and the Internet seem to provide the most efficient and user-friendly possibilities in this respect.

Figure 19 shows an extended version of figure 16, where the feedback loop has been introduced in the sense that the survey design and planning model shows not only the documentation and metadata spin-offs of each design phase, but also the documentation and metadata inputs that can be used by the same design phases, if these inputs are properly organised and disseminated.
Figure 16. The SCBMOD survey planning and development model.
Figure 17. How the contents of an SCBDOK documentation is generated step by step by the design decisions in the design phases with the corresponding colours in figure 16.

Figure 18. How the contents of a Quality Declaration (QD) is generated step by step by the design decisions in the design phases with the corresponding colours in figure 16.
Figure 19. Extended version of the SCBMOD survey design and planning model in figure 16, covering both inputs and outputs.
Finally it is of utmost importance for the producers that they feel that different documentation and metadata systems are well co-ordinated so as to minimise their work and secure that the results become maximally useful.

It is inevitable that several types of documentation and metadata have to be created and maintained in a statistical organisation. There are so many different users and usages of documentation and metadata, and the users have quite different needs and competence profiles. It is important and also feasible to design the different documentation and metadata systems in such a way that each one of them can make use of what is already available in the others, and vice versa. Some systems are more fundamental than others. For example, a complete SCBDOK documentation contains most of the information needed by other systems, e.g. metadata systems supporting user-friendly presentations on the Internet especially intended for casual users with little time and with little statistical competence. However tempting it may be to start by designing the user-friendly metadata system on the Internet, it makes more sense to start by producing good SCBDOK documentations, since they can then be used as “raw material” for the user-friendly presentations, thus creating a more or less automatic metadata transformation and feeding procedure from the information richer but not so elegant SCBDOK system to the elegant but not so rich presentation system. To go the opposite way would be much more complicated.

If a certain part of a documentation can be made in both a formalised and less formalised way, and if it is envisaged that the formalised version is needed, one should start with the formalised version, preferably with the support of efficient and user-friendly tools, and, if necessary, then generate a less formalised, and possibly more user-friendly version. This is easier than to go the other way. For example, it is easier to generate plain text from a set of relational tables than to generate a set of relational tables from plain text, even if the two representations contain exactly the same information, and even if the text is relatively well structured. The relationship between SCBDOK and METADOK is a good example of this.

At Statistics Sweden there is an on-going process for streamlining and co-ordination of existing documentation and metadata systems. For example, figure 20 shows the relatively close correspondence between SCBDOK, aimed primarily at advanced users/researchers of microdata, and the Quality Declarations (QD), aimed primarily at relatively knowledgeable users of officially published statistics (macrodata). On the basis of the quality declarations popularised “Facts about the statistics” are produced to meet the needs of casual users of the statistical reports and press releases that are published regularly, both electronically via the Internet, and in printed form. However, the “Facts about the Statistics” contains a subset of the same information as the quality declarations presented in a way better suited for journalists and other casual users. For those who need the full quality declarations and maybe even the full SCBDOK documentations, references are made to these documentations, on the Internet in the form of dynamic links.

Figure 21 summarizes the Swedish system of documentation and metadata system based upon the SCBDOK conceptual model as described in Rosén & Sundgren (1991).
<table>
<thead>
<tr>
<th>SCBDOK vs QD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Contents overview</strong></td>
</tr>
<tr>
<td>1.1 Observation characteristics</td>
</tr>
<tr>
<td>1.2 Statistical target characteristics</td>
</tr>
<tr>
<td>1.2.1 Objects and population</td>
</tr>
<tr>
<td>1.2.2 Variables</td>
</tr>
<tr>
<td>1.2.3 Statistical measures</td>
</tr>
<tr>
<td>1.2.4 Study domains</td>
</tr>
<tr>
<td>1.2.5 Reference time</td>
</tr>
<tr>
<td>1.2 Comprehensiveness</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>2 Data collection</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Overall accuracy</td>
</tr>
<tr>
<td>2.2 Sources of inaccuracy</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>3 Timeliness</strong></td>
</tr>
<tr>
<td>3.1 Frequency</td>
</tr>
<tr>
<td>3.2 Production time</td>
</tr>
<tr>
<td>3.3 Punctuality</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>4 Coherence especially comparability</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Comparability over time</td>
</tr>
<tr>
<td>4.2 Comparability over space</td>
</tr>
<tr>
<td>4.3 Coherence in general</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>5 Availability and clarity</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Forms of dissemination</td>
</tr>
<tr>
<td>5.2 Presentation</td>
</tr>
<tr>
<td>5.4 Access to microdata</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>3 Final observation registers</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Production versions</td>
</tr>
<tr>
<td>3.2 Archive versions</td>
</tr>
<tr>
<td>3.3 Experiences from the latest collection round</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>4 Statistical processing and presentation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Estimations: assumptions and formulas</td>
</tr>
<tr>
<td>4.2 Presentation and dissemination procedures</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>5 Data processing system</strong></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>6 Logbook</strong></th>
</tr>
</thead>
</table>

Figure 20. An analysis of the correspondences between the SCBDOK and QD templates.
Figure 21. The documentation and metadata systems at Statistics Sweden based upon the conceptual model of the SCBDOK description model for statistical surveys.
7 Process data

Here and there in this paper explicit or implicit references have been made to a special kind of metadata, so-called process data. Process data inform about processes, either individual instances of processes or subprocesses, e.g. the editing process of a certain survey, or all processes or subprocesses of a certain type, e.g. the coding process in the statistical system of Sweden.

The most explicit reference to process data in the SCBOK template is item 3.3: “Experiences from the latest collection round”. Chapter 2, “Accuracy”, in the Quality Declaration template will also require process data.

Up to now, it has not been common for statistical offices to collect, organise, and analyse process data in a systematic way. With a growing interest in systematic quality work, such as Total Quality Management (TQM) the needs for such efforts have become more and more evident.

The same rules apply for collection of process data as for collection of other types of metadata. They must be collected and organised with as little efforts as possible on the part of the producers. Preferably the processes of a statistical survey or survey system should be designed in such a way that when implemented and put into operation, they will more or less automatically generate relevant, basic process data for the respective processes. Figure 22 gives the data model for a general system that would be able to collect basic process data for any type of data collection process: collection of responses to questionnaires and questions in questionnaires, follow-ups on non-response, coding, identification of suspected errors, follow-ups on suspected errors resulting or not resulting in updates9, etc.

---

Figure 22. Data model for a general system collecting basic process data about the data collection processes of a statistical survey or survey system.

---

9 Hopefully the updates are actually corrections, but this cannot be taken for granted.
In order to become useful the basic process data should be organised and analysed. This can actually be regarded as a statistical survey in itself, the design of which should use the same methods, rules, and tools as the design of any other statistical survey.

At present we can see a relatively rapid development, where the metadata generated by statistical surveys will ultimately become an integrated part of the computerised outputs (databases of microdata and macrodata) generated by the surveys. Thus we can hope for statistical data sets to become more and more self-explanatory, and for the statistical processes to become metadata-driven. Similarly, the statistical data sets should also be enriched with process data that may be useful or even necessary for the users’ proper understanding and evaluation of the quality of the statistical data. The statistical processes should become not only metadata-driven but also process data driven in the sense that process data is used for systematic, sometimes even automatic, improvements of the design and operation of the processes.

Process data is a relevant topic not only for input processes but also for output processes. When we are now more and more using electronic channels like the Internet for the dissemination of statistical data, we get excellent new opportunities for collecting feedback from the users. Part of this feedback can be obtained simply by following the traces of the users on our website; excellent software is already available for such tasks, which require no extra efforts on the part of the users. The users can also be given opportunities to give the producers more sophisticated feedback, such as information about statistical data and services that they are missing.

References


