

**Metadata Working Group 2004**

**(Luxembourg, 17-18 June 2004)**

**BECH Building - Room AMPERE**

**DOCUMENTATION TEMPLATES AND METADATA MODELS  
AT STATISTICS SWEDEN**

**MWG/2004/7.1  
(Item 7 on the agenda)**

# Documentation templates and metadata models at Statistics Sweden

*Bo Sundgren*

*2004-06-07*

## Purpose of this paper

The purpose of this paper is to give a short overview of the documentation templates and metadata models used by Statistics Sweden and of the concepts underlying these. The reader is referred to the documents listed in the bibliography at the end of this paper for more details.

## The SCBDOK documentation template

Figure 1 gives an overview of the SCBDOK documentation template, version 3.0.

SCBDOK 3.0	
<b>0 General information</b> 0.1 Subject matter area 0.2 Statistics area 0.3 Official statistics? 0.4 Responsibility 0.5 Producer 0.6 Mandatory response? 0.7 Secrecy 0.8 Destruction rules 0.9 EU regulation 0.10 Purpose and history 0.11 Users and usage 0.12 General approach to implementation 0.13 Planned changes	<b>1 Contents overview</b> 1.1 Observation characteristics 1.2 Statistical target characteristics 1.3 Outputs: microdata and statistics 1.4 Documentation and metadata  <b>2 Data collection</b> 2.1 Frame and frame procedure 2.2 Sampling procedure (if applicable) 2.3 Measurement instruments 2.4 Data collection procedure 2.5 Data preparation
<b>3 Final observation registers</b> 3.1 Production versions 3.2 Archive versions 3.3 Experiences from the latest collection round	<b>4 Statistical processing and presentation</b> 4.1 Estimations: assumptions and formulas 4.2 Presentation and dissemination procedures
<b>5 Data processing system</b>	<b>6 Logbook</b>

*Figure 1. The SCBDOK documentation template, version 3.0.*

The major contents of an SCBDOK documentation could be described as follows (cf also figure 2 below):

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 in terms of population, classification variables, summation variables, and statistical measures. 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 (cf figure 2 below) 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 3 below shows the quality declaration template.

Figure 3 and figure 4 illustrate some fundamental concepts used in SCBDOK.

The theory underlying the SCBDOK documentation template is explained in detail in Rosèn&Sundgren (1991).

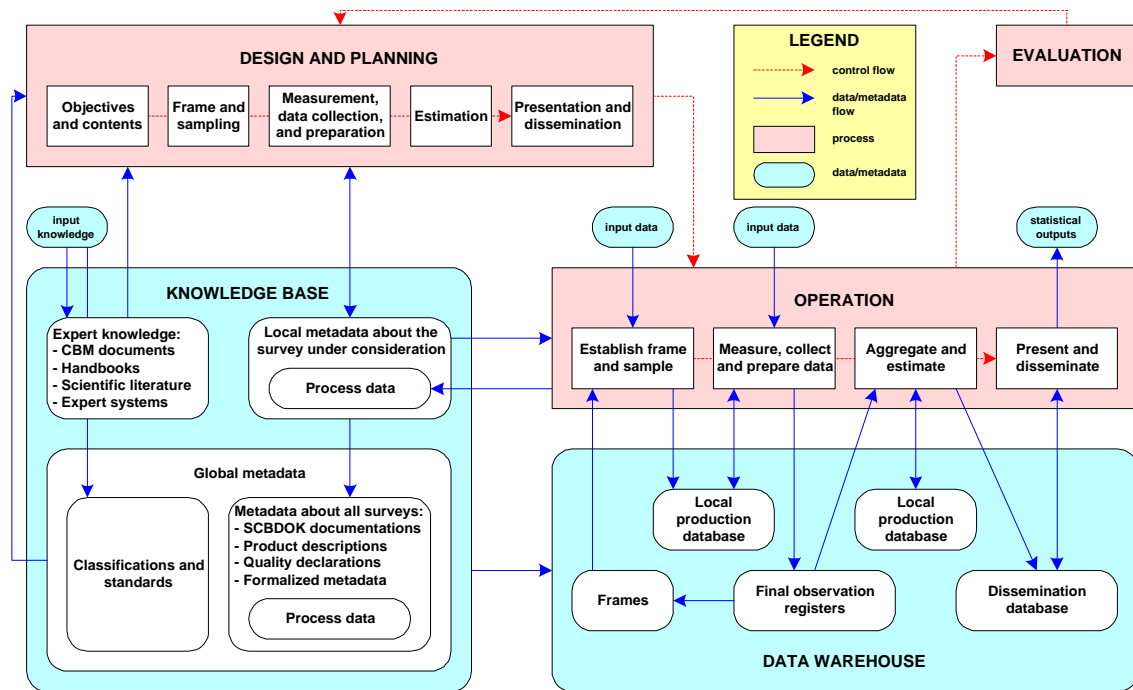


Figure 2. Processes and data/metadata sets in statistics production.

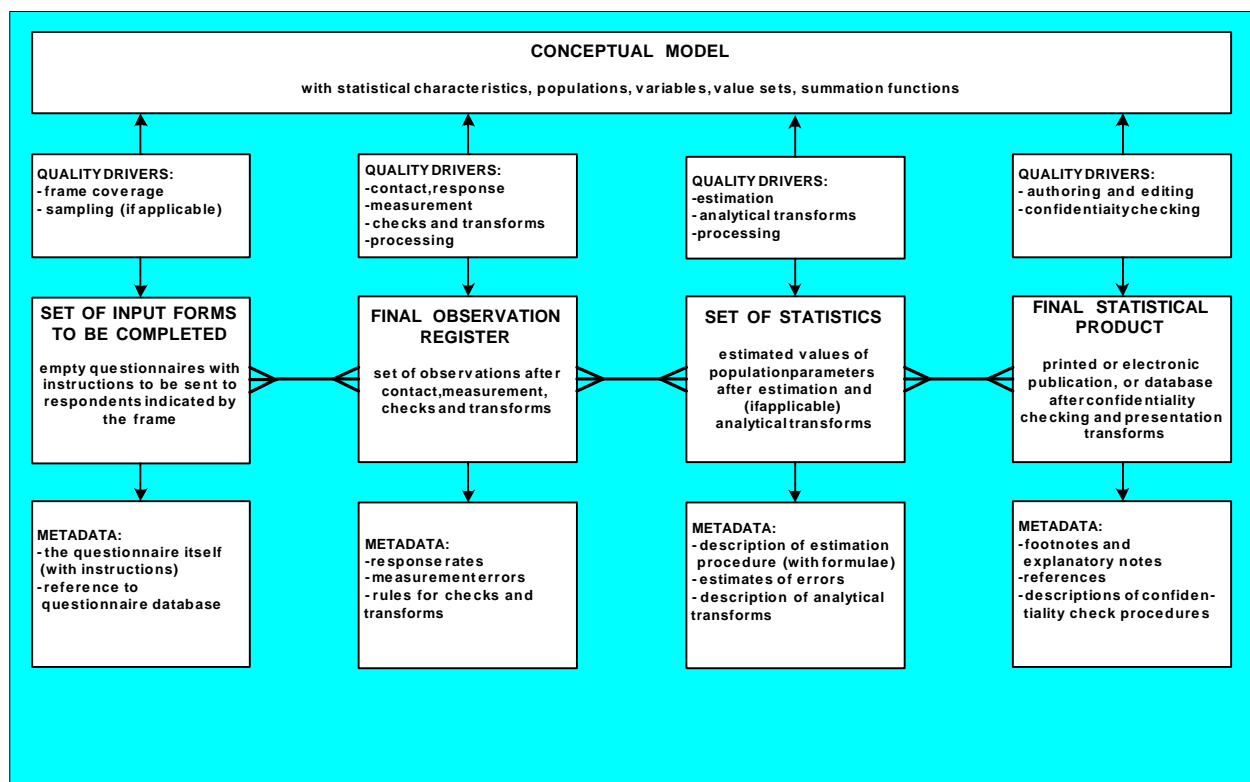


Figure 3. Conceptual model, quality drivers, and required metadata in statistics production.

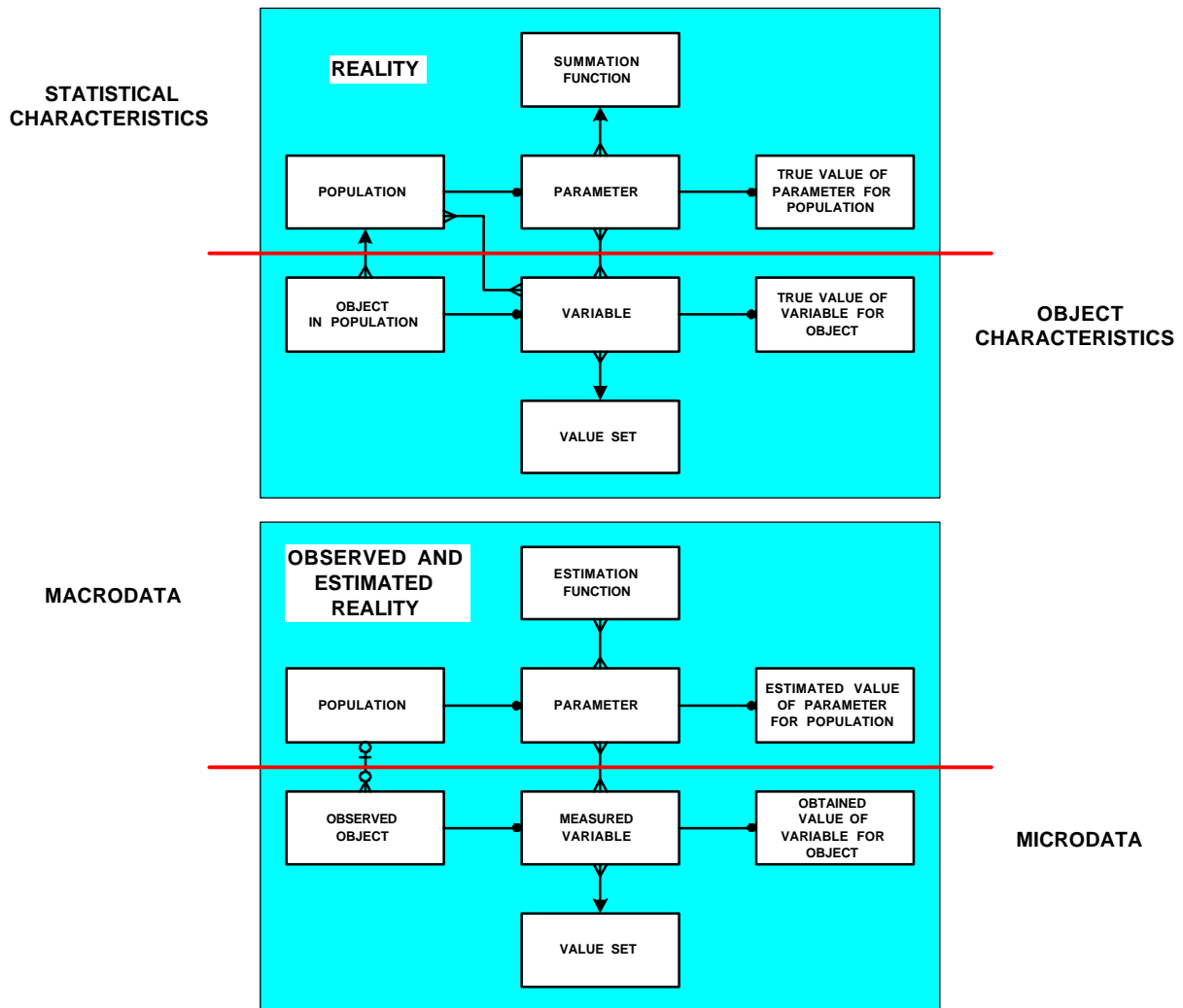


Figure 4. Fundamental concepts in statistics production.

## MicroMeta: metadata model for final observation registers, used in METADOK

Figure 5 gives an overview of the metadata model used by METADOK when describing final observation registers. A detailed description of this model is given in Abelin&Andersson (2004).

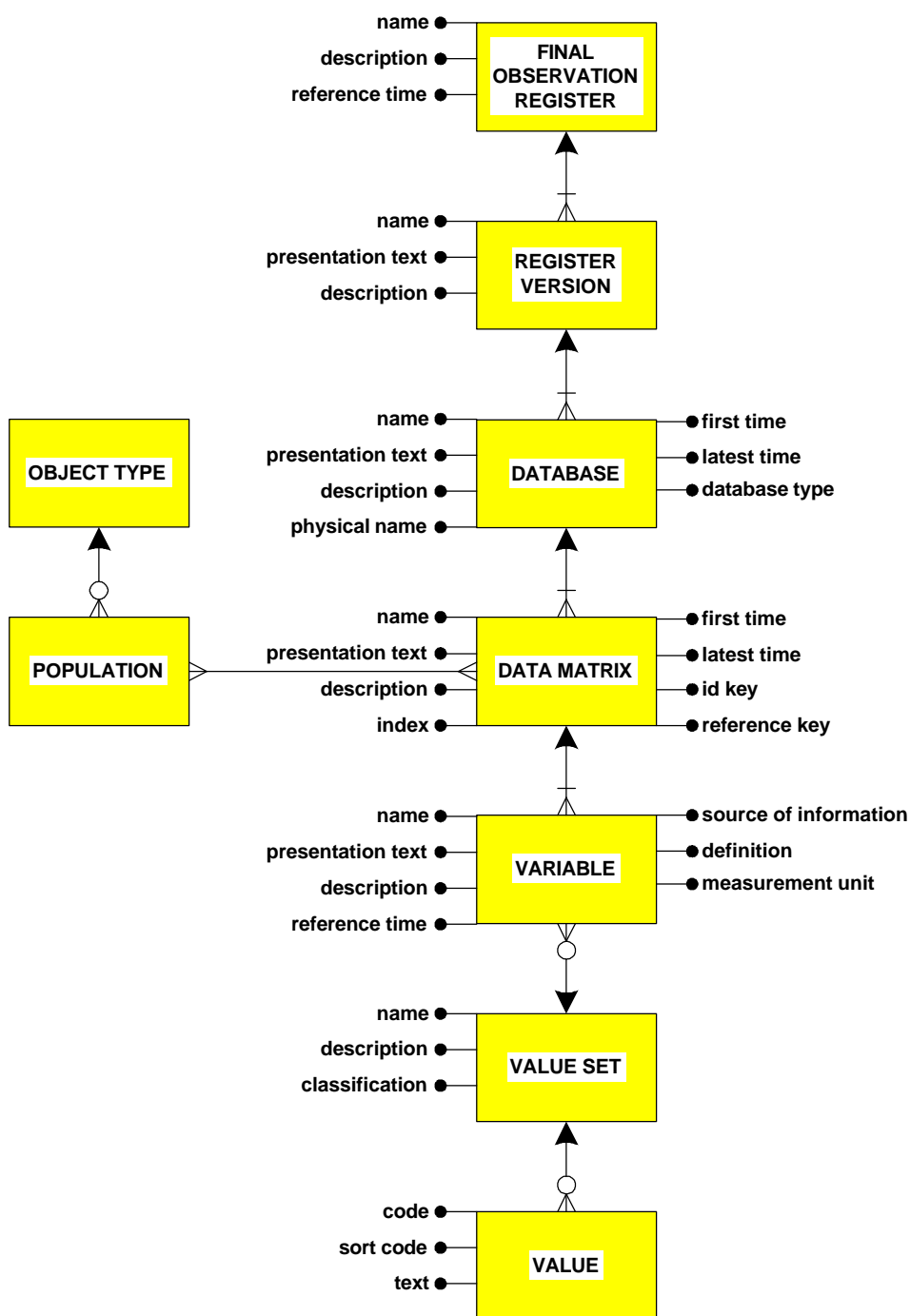


Figure 5. Conceptual model of basic metadata for a final observation register. This is a simplified version of the metadata model used in the Swedish SCBDOK/METADOK system for describing final observation registers, stored as, for example, relational databases, SAS data sets, or flat files. This model corresponds to section 3 of the SCBDOK documentation template, version 3.0.

## The Quality Declaration Template

Figure 6 gives an overview of the Quality Declaration Template used by Statistics Sweden.

Quality Declaration Template	
<b>1 Contents</b> 1.1 Statistical target characteristics 1.1.1 Objects <sup>1</sup> and population 1.1.2 Variables 1.1.3 Statistical measures 1.1.4 Study domains 1.1.5 Reference time 1.2 Comprehensiveness	<b>2 Accuracy</b> 2.1 Overall accuracy 2.2 Sources of inaccuracy <sup>2</sup> 2.2.1 Sampling 2.2.2 Coverage 2.2.3 Measurement 2.2.4 Non-response 2.2.5 Data processing <sup>3</sup> 2.2.6 Model assumptions 2.3 Presentation of accuracy measures
<b>3 Timeliness</b> 3.1 Frequency 3.2 Production time 3.3 Punctuality	<b>4 Coherence especially comparability</b> 4.1 Comparability over time 4.2 Comparability over space 4.3 Coherence in general
<b>5 Availability and clarity</b> 5.1 Forms of dissemination 5.2 Presentation 5.3 Documentation 5.4 Access to microdata 5.5 Information services	

*Figure 6. The quality declaration template used in the Swedish statistical system. The model is similar to the model used by the so-called European Quality Concept.*

The Quality Declaration Template, and the concepts used, there are discussed in detail in Statistics Sweden (2001). The conceptual foundation of the Quality Declaration Template is the same as that of SCBDOK.

Figure 7 illustrates the origins and effects of different types of errors in a statistical production process. The mechanisms behind these errors have to be in SCBDOK and summarised in the Quality Declaration for the needs of end-users of statistics (macrodata).

<sup>1</sup> 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.

<sup>2</sup> Also called "sources of error".

<sup>3</sup> The term used elsewhere in this paper and in other papers concerning SCBDOK is "data preparation".





## MacroMeta: metadata model for macrodata used by the Statistics Sweden's output database

Figure 8 gives a simplified overview of the metadata model used by the software system supporting the dissemination or output databases of Statistics Sweden, which are available on the Internet. The complete MacroMeta model, containing among other things support for footnotes and bilinguality, is shown in figure 9.

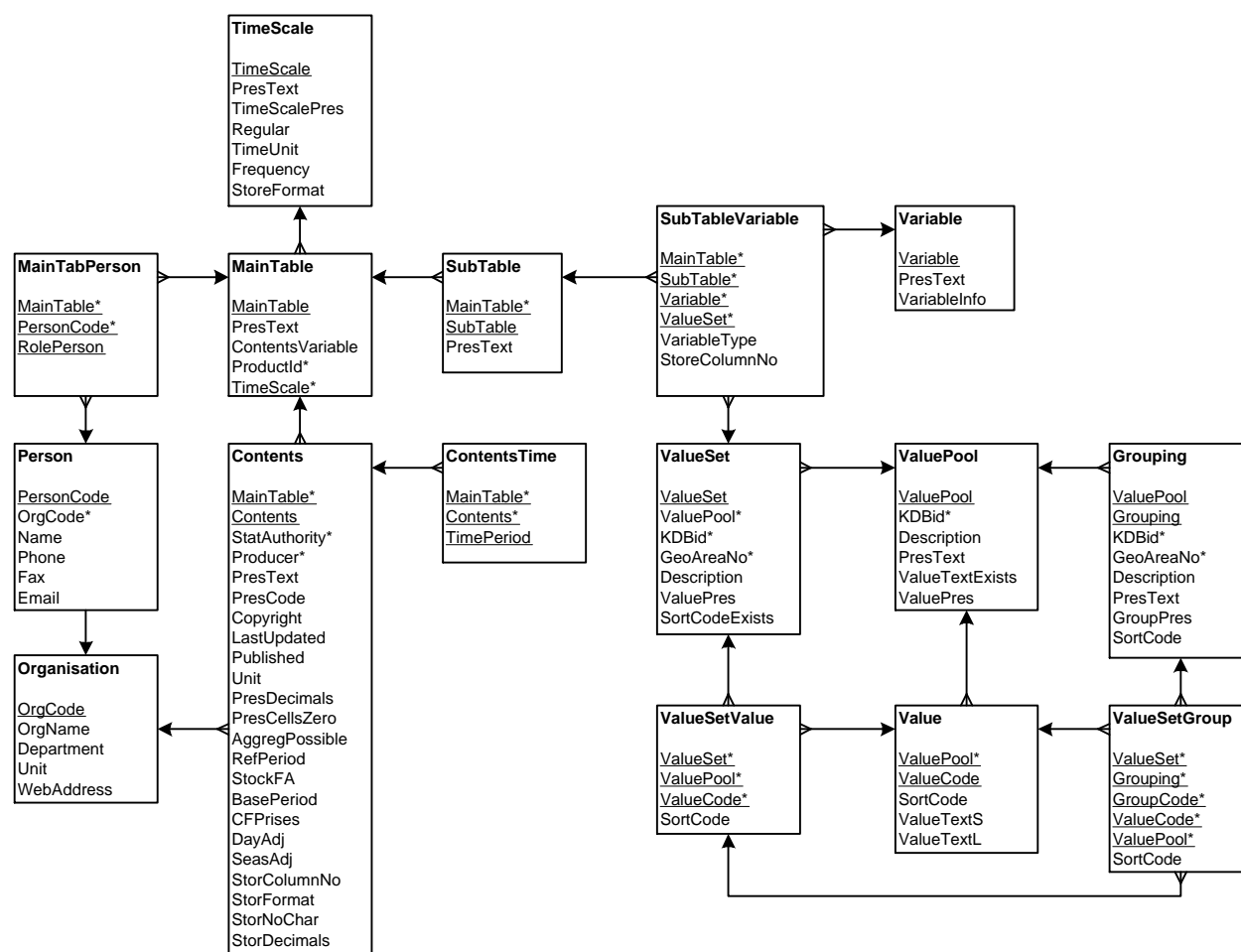


Figure 8. Relational data model for basic metadata for final statistics stored in a relational database. This is a simplified version of the metadata model underlying the statistical output database of Statistics Sweden, which is available on the Internet at [www.scb.se](http://www.scb.se). The complete model is shown in Figure 9 below.



## Bibliography

**Rosén, B. & Sundgren, B. (1991)** *Documentation for reuse of microdata from the surveys carried out by Statistics Sweden*. Statistics Sweden 1991. (This paper is the conceptual foundation for the SCBDOK documentation template.)

**Sundgren, B. (2000)** *The Swedish Statistical Metadata System*. Statistics Sweden 2000.

**Statistics Sweden (2001)** *Quality definition and recommendations for quality declarations of official statistics*. Reports on Statistical Co-ordination for the Official Statistics of Sweden, MIS 2001:1.

**Sundgren, B. (2001)** *Documentation and Quality in Official Statistics*. Paper presented at the International Conference on Quality in Official Statistics (Q2001), Stockholm, May 14-15, 2001.

**The Neuchâtel Group (2002)** *Neuchâtel Terminology. Classification database object types and their attributes. Version 2.0*. (The members of the informal Neuchâtel Group have varied over time, but there has been a core group of representatives from Statistics Denmark, Statistics Norway, the Swiss Federal Statistical Office, and un Software-Werkstatt.)

**Abelin, M. & Andersson, E. (2004)** *Metadata in the microdatabase, version 1.30*. Statistics Sweden 2003-10-15.

**Statistics Sweden (2004)** *A description of the central metadatabase for macrodata. Version 2.00*. Statistics Sweden 2004-01-19.