Statistical databases in theory and practice

Part IV: Metadata, quality, and documentation

Bo Sundgren

2010

Summary of needs for metadata

- Users of statistics need good metadata in order to identify, locate, retrieve, interpret, and analyse statistical data of relevance for their primary tasks.
- Producers/operators of statistical systems need metadata for the same purposes, but also for executing and monitoring the production processes properly, and for training new staff members; process data or paradata, data about processes.
- Producers/planners of statistical systems need metadata for designing, constructing, and implementing statistical systems.
- Respondents need metadata in order to understand why their participation in a survey is needed and for interpreting the meaning of the questions to be answered.
- Managers need metadata in order to evaluate different aspects of statistics production, including aspects of production efficiency, user satisfaction, and acceptance by respondents.
- Funders have similar needs as managers but on a more global level. They also need meta-data that help them to balance the needs for statistical information against other needs that they have.
- Researchers in statistical systems need well organised and systematised metadata (including process data) about how statistical systems are designed and how they perform, to be combined with existing knowledge about relevant theories and methods.
- Software products and applications need metadata in order to function properly. They should also generate process data which will help to regulate the performance of human and computerised processes.

What should statistical metadata inform about?

- Metadata about data and associated concepts
- Metadata about processes and associated procedures and software
- Metadata about instrumental resources, process enablers, including the metadata resources themselves

Metadata objects and metadata variables

For each process, analyse…

1. Which are the metadata needs of the process?
2. Are there any sources from which the process could get these metadata?
3. Which metadata could the process provide to other processes?

Quality of statistical data

- quality = property = metadata
- quality = good quality (for intended use or usages)
- quality = absence of errors and uncertainties
- quality = discrepancies between
  - statistical characteristics, ”true reality”
  - statistics, ”estimated reality”

Statistical characteristic

- a statistical measure (m) applied on
- the (true) values of a variable (V); V may be a vector
- for the objects in a population (O)

- O.V.m = statistical characteristic
- O.V = object characteristic
- V.m = parameter

Examples of statistical characteristics

- number of persons living in Sweden at the end of 2001
- average income of persons living in Sweden at the end of 2001
- correlation between sex and income for persons living in Sweden at the end of 2001

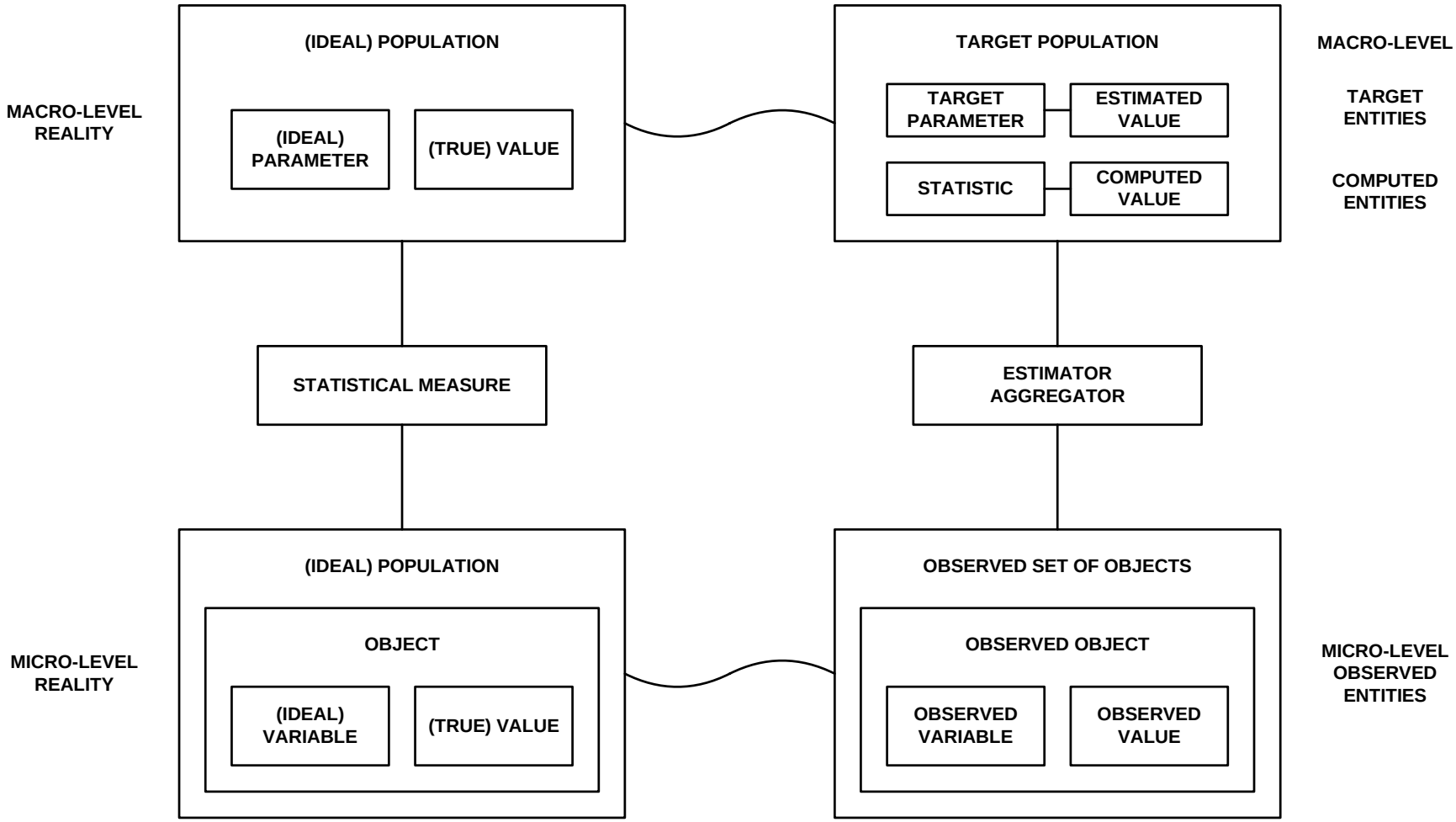
Statistic

- an estimator (e) applied on
- *observed* values of an *observed* variable (V');
- for a set of *observed* objects (O') *allegedly* belonging to a population (O)

Ideally the value of a statistic O'.V'.e should be "close to the true value of the statistical characteristic O.V.m that it aims at estimating

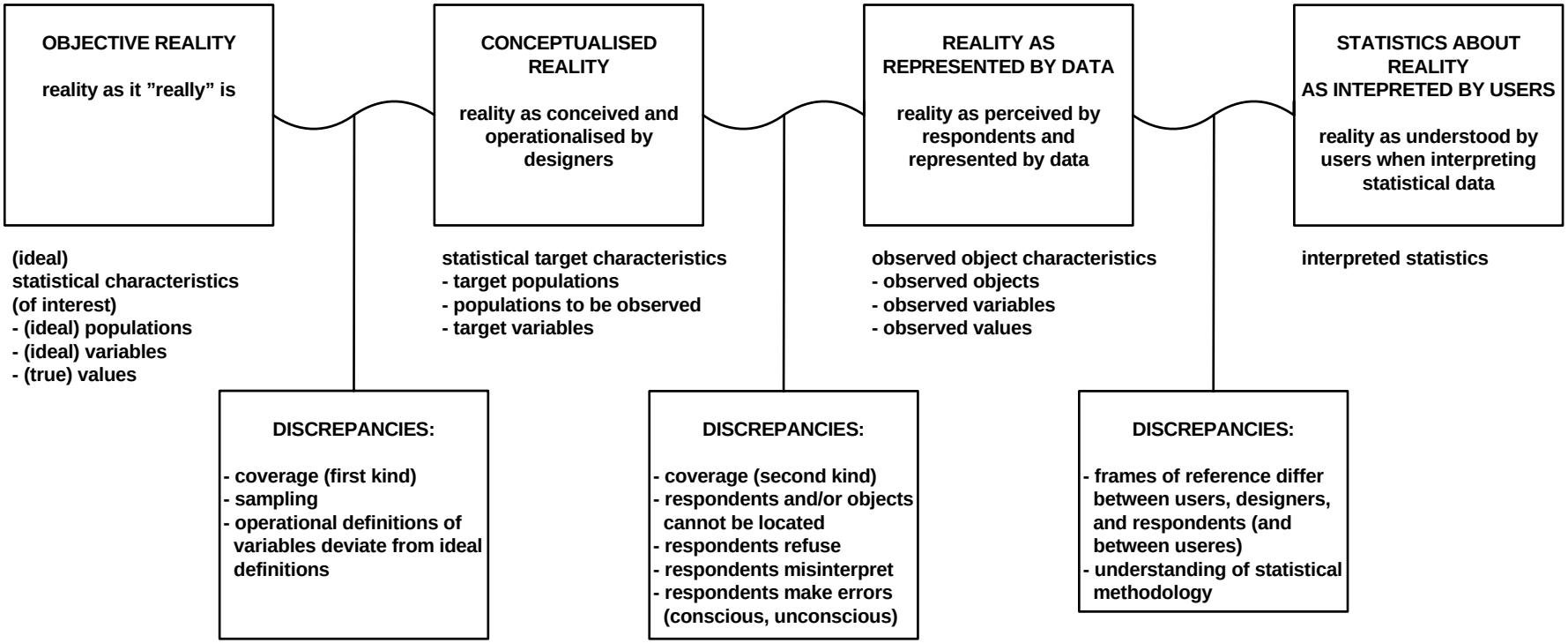
Examples

- the estimated number of persons living in Sweden at the end of 2001
- the estimated average income of persons living in Sweden at the end of 2001
- the estimated correlation between sex and income for persons living in Sweden at the end of 2001



Statistics = estimated statistical characteristics

Discrepancies between reality ”as it is” and as it is reflected by statistics: which they are, and why they occur



”Errors”: uncertainties, discrepancies

- caused by design decisions
- occurring during operation processes
- occurring during use processes

In summary, there are differences between …

- *reality as it "really" is*, the true or objective reality
- *reality as it is conceptualised and operationalised* during the design of a statistical system
- *reality as it is represented by data* during the statistical production processes
- *reality as it is interpreted* by users of statistics

Template for documentation of processes and microdata

SCBDOK 3.0	
0 General information 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	**1 Contents overview** 1.1 Observation characteristics 1.2 Statistical target characteristics 1.3 Outputs: microdata and statistics 1.4 Documentation and metadata **2 Data collection** 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
3 Final observation registers 3.1 Production versions 3.2 Archive versions 3.3 Experiences from the latest collection round	**4 Statistical processing and presentation** 4.1 Estimations: assumptions and formulas 4.2 Presentation and dissemination procedures
5 Data processing system	**6 Logbook**

The Quality Declaration Template of Statistics Sweden

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

Descriptive and prescriptive models

- system specification (prescriptive model)
- system documentation (descriptive model)

- system/data specification +
- process data and documented experiences =
- system/data documentation

Log-book

- immediate, dated reports, by identified persons, about any kind of exceptional events or changes in plans and design that occur during system execution

Part IV: Extra material

Bo Sundgren

2010

Users and usages of statistical metadata

- Users of statistical outputs
  - use cases
- Producers of statistics operating and monitoring statistical systems
- Software supporting statistical processes
  - metadata-driven systems
- Respondents and data providers
- Planners and evaluators of statistical systems
- Highlevel managers and funders of statistical systems
- Researchers on statistical systems

Usage process

Metadata needs of this process

- Which statistical data and services are available? Are they relevant for my task?
- Contents and qualities of available statistical data (macrodata and microdata)?
  - statistical microdata and their definitions
  - relevance, accuracy, timeliness, availability, comparability, coherence
- Where to find, and how to retrieve, chosen statistical data?
- How to interpret retrieved statistical data?

Potential sources of metadata needed

- Planning, design, and construction processes → e.g. definitions
- Operation and monitoring processes → e.g. non-response rates (→ precision)

This process is a potential provider of

- Data about usage and requests (both satisfied and not satisfied) for data and services
- Data about user satisfaction with products and services

Use case 1: The user has a problem but does not know which relevant statistical data are available

1. The user tries to find statistical data that are possibly relevant for the problem by using available search tools and metadata.
2. The user identifies some data sets that may possibly contain some relevant data and asks for more detailed information about these data: contents, storage, and availability.
3. The user retrieves certain statistical data.
4. The user interprets and analyses the data, guided by metadata and documentation accompanying the data.
5. The user may return to one of the earlier steps.

Use case 2: The user has a recurrent problem and knows which relevant statistical data are available.

1. The user wants fast and simple access to the data of interest. Preferably the statistical system should already know about the user's regular requests, and the user will select one of these request types, e.g. from a list, and maybe add some parameters concerning time version etc.
2. The user receives the requested data with (links to) associated metadata.
3. The user interprets and analyses the data.
4. The user may return to step 1 (for another regular request) or move to *Use case 1* for some more unique request, triggered by the analysis of the regular request.

Use case 3: The user is engaged in an on-going process (like doing business on the stock market) and wants to be alerted when new statistical data of a certain kind become available.

1. The statistical system signals to the user, when new data of the requested kind becomes available. Deviations from the expected are clearly indicated.
2. The user acts on the statistical data received, possibly after some further analysis (assisted by metadata and tools) and some further request for other statistical data.

Production process: operation and monitoring

Metadata needs of this process

- Instructional metadata: specifications and instructions for inputs, outputs, procedures
- Motivational and instructional metadata for respondents and data providers: why is participation in a survey important, how is confidentiality ensured, how to interpret the meaning of the questions to be answered, how to respond
- Information about methods, tools, and resources available
- Process data feed-back about process performance in terms of errors, timeliness, resource consumption, etc

Potential sources of metadata needed

- Planning, design, and construction processes → instructions, specifications, descriptions
- Operation and monitoring processes → process data

This process is a potential provider of

- Process data

Software supporting statistical processes

- highly structured and formalised metadata
- instructional metadata
- process data
- software/metadata independence, metadata-driven software

Metadata-driven systems

A well designed computerised statistical production process will

- be dynamically driven by up-to-date metadata, managed by separate processes
- automatically generate process data about its own performance
- feed back the process data into itself, regulating its own behaviour, when needed
- feed back adequate process data to human beings monitoring the process

Planning and (re)design process

Metadata needs of this process

- Information about user needs and other stakeholder requirements
- Detailed, up-to-date documentation of the present system (if it exists)
- Experiences from this system (if it exists) and similar systems here and elsewhere: *ad hoc* information as well as well planned feed-back information, both formal and informal, concerning both the production parts and the usage parts of the statistical system
- Special evaluation studies performed on an *ad hoc* basis
- Information about methods, tools, available data sources, available software components, and other resources

Potential sources of metadata satisfying these needs

- Knowledge bases: physical and electronic libraries, websites, etc
- Business intelligence processes: using search engines, intelligent agents as well as more traditional research methods

This process is a potential provider of

- Specifications of processes (inputs, procedures, outputs), descriptions of resources, instructions

Management and evaluation (including funders)

Metadata needs of this process

- Information about user needs and other stakeholder requirements
- Costs and benefits?
- To which extent do users actually use the statistical outputs, and are they satisfied with the qualities of the data as regards contents, accuracy, timeliness, availability, comparability, coherence, etc? Information about the performance of production processes and usage processes
- Are there complaints from respondents, and how much unpaid work will they have to do?
- Knowledge about methods, tools, available data sources, and other resources
- Experiences from comparable systems

Potential sources of metadata satisfying these needs

- Knowledge bases and business intelligence processes
- Data generated by or requested from production and usage processes

This process is a potential provider of

- Experiences
- Revised user needs and other stakeholder requirements

Research and general development on statistical systems

Metadata needs of this process

- General knowledge about statistical systems and statistics production, e.g. recognised theories and methods, standards like “current best methods” and “current best practices”
- Specific knowledge, *experiences*, about “how things are done” in different statistical organisations
- Experiences, *process data* about costs and different aspects of quality in statistical processes performed by statistical organisations

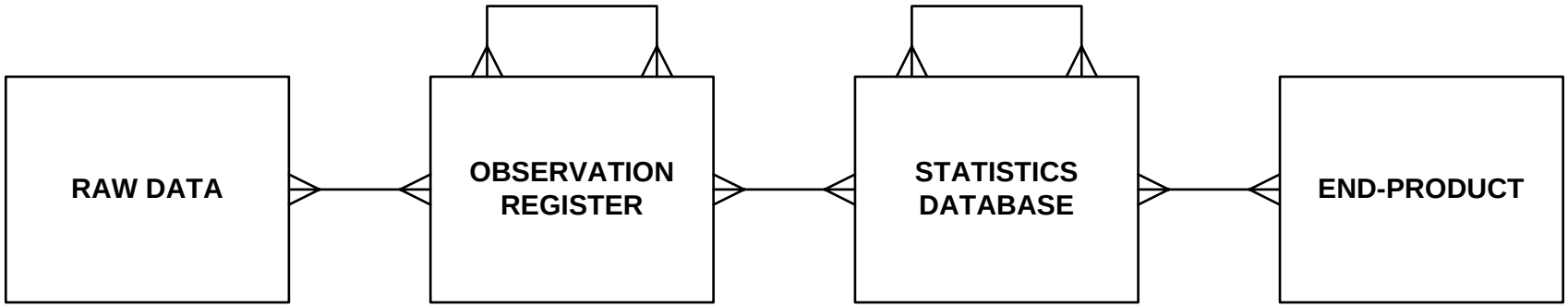
Potential sources of metadata satisfying these needs

- Knowledge bases and business intelligence processes
- Experiences from production processes in many systems
- Experiences from usage processes in many systems

This process is a potential provider of

- Knowledge, methods, and tools

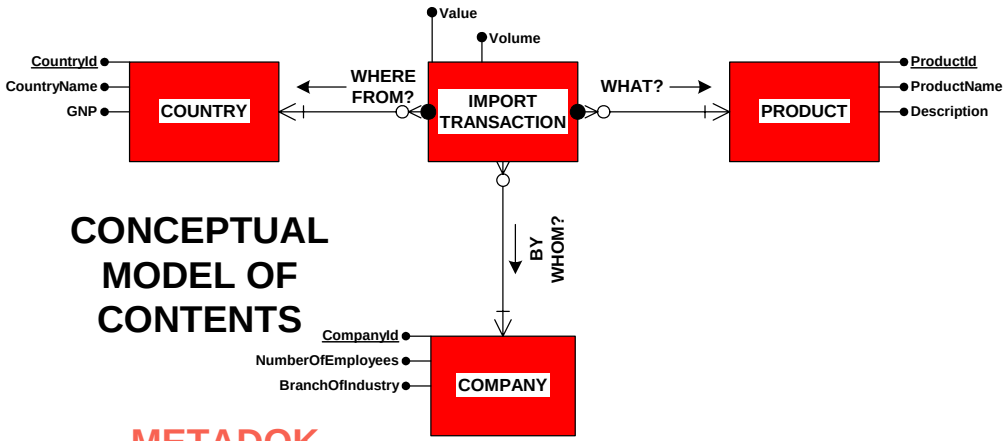
Main stages during the lifecycle of statistical data and metadata



- The raw data stage
- The observation register stage
- The statistics stage
- The end-product stage

Fundamental interfaces and documentation objects

SCBDOK
QualityDeclaration
Classification
Database (KDB)



CONCEPTUAL MODEL OF CONTENTS

METADOK MicroMeta

SCBDOK

Quality Declaration

MacroMeta

INPUT DATA COLLECTION

e.g. set of completed questionnaires from one data collection process

SCBDOK

FINAL OBSERVATION REGISTER

e.g. a relational database, a SAS dataset, or a set of flat files, organised as one set of microdata

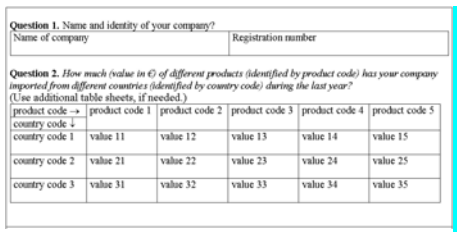
SCBDOK

FINAL SET OF STATISTICS

e.g. a relational database, or a set of multidimensional cubes (stars), organised as one set of macrodata

OUTPUT DATA COLLECTION

e.g. a publication or a set of tables on a website, organised as one integrated set of statistical outputs

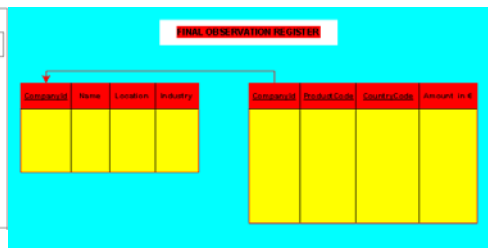


Question 1. Name and identity of your company?

Name of company	Registration number

Question 2. *How much (value in €) of different products (identified by product code) has your company imported from different countries (identified by country code) during the last year?*
(Use additional table sheets, if needed.)

product code → / country code ↓	product code 1	product code 2	product code 3	product code 4	product code 5
country code 1	value 11	value 12	value 13	value 14	value 15
country code 2	value 21	value 22	value 23	value 24	value 25
country code 3	value 31	value 32	value 33	value 34	value 35



METADOK

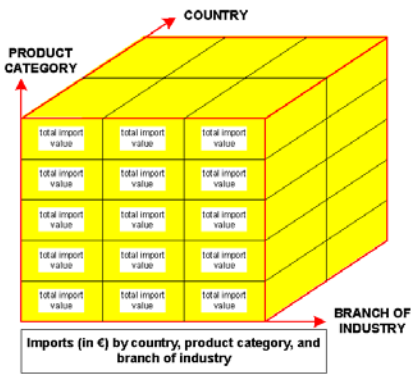
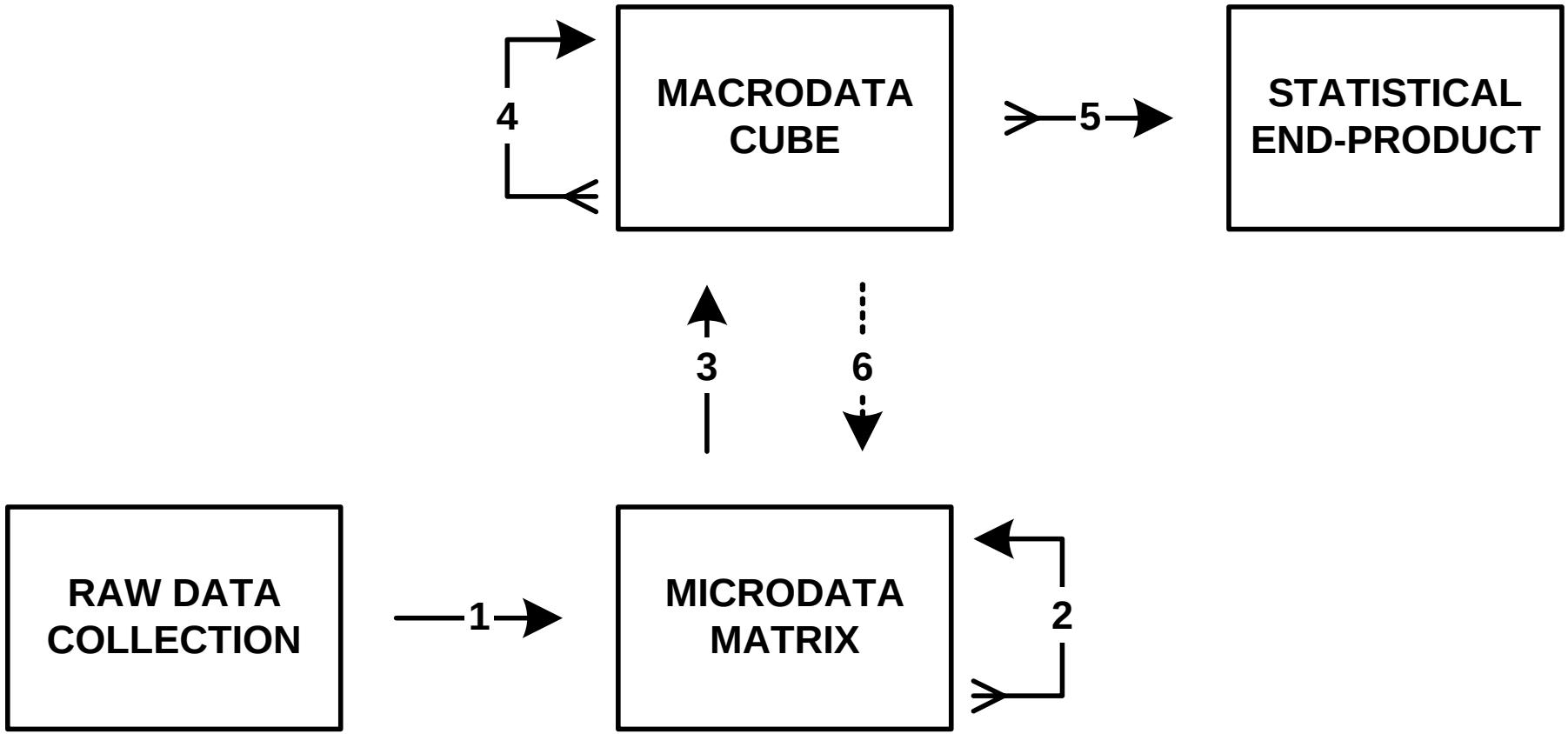


Table 1. *Imports (in €) by country, product category, and branch of industry.*

		Industry code 1	Industry code 2	Industry code 3	…
Country 1					
	Product code 1				
	Product code 2				
	…				
Country 2					
	Product code 1				
	Product code 2				
	…				
Country 3					
…					

Deriving statistical end-products from raw data collections



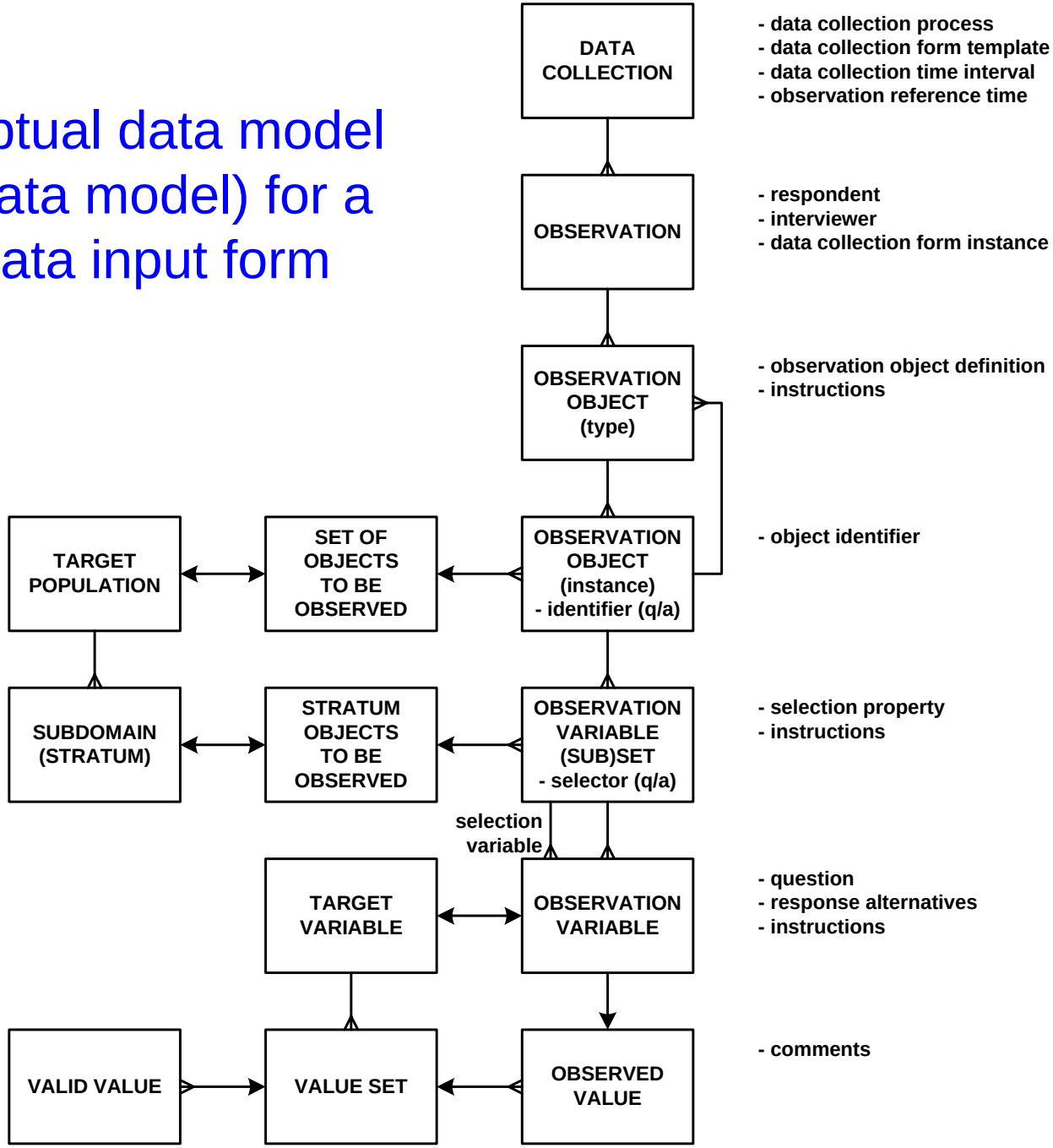
Example of an observation matrix (observation data, microdata)

PERSON	Identifier	HouseholdId	Sex	Age	Education	Occupation	Income
Person 1							
Person 2							
Person 3							
...							
Person m							

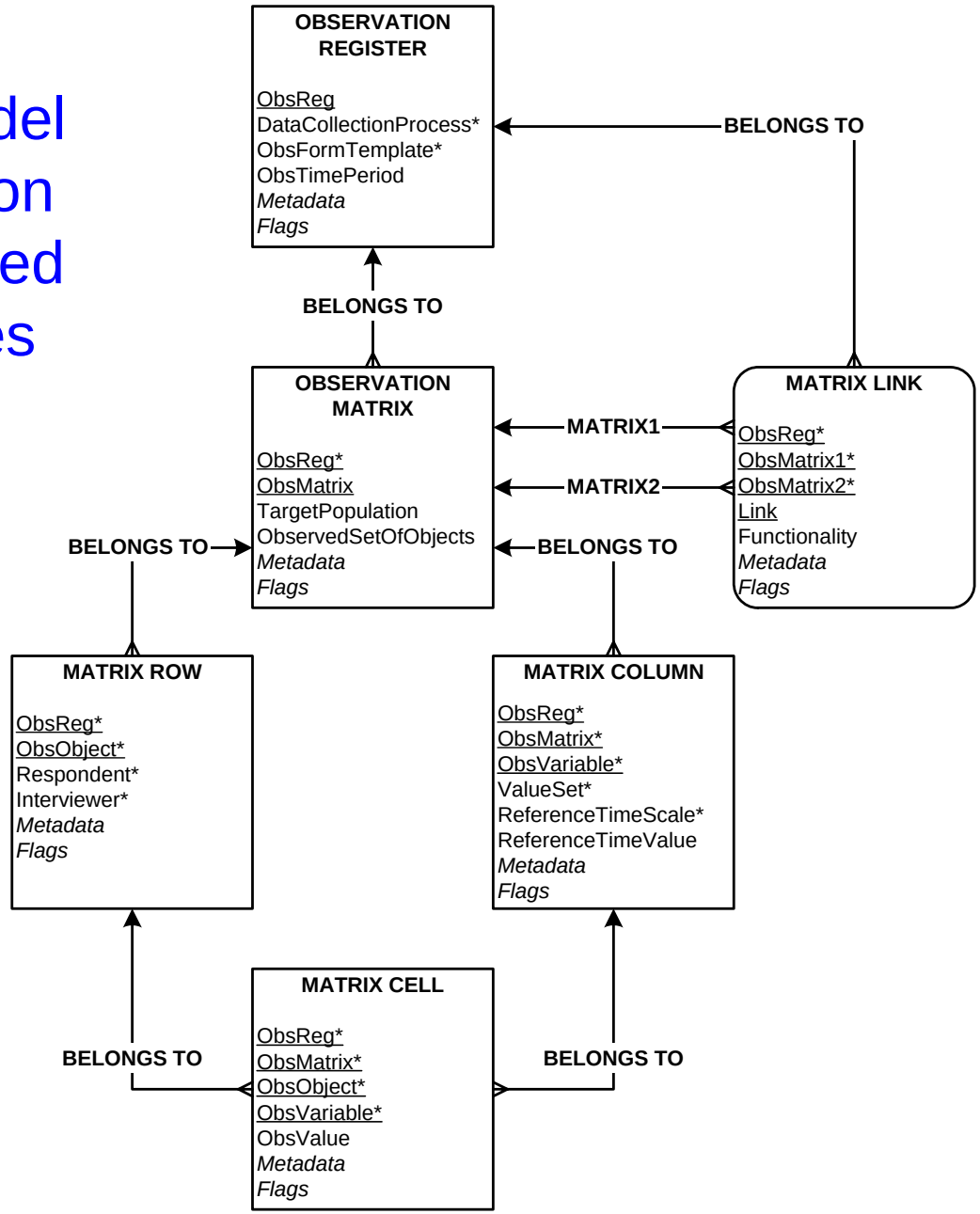
Structure of an observation matrix

MATRIX	**Variable 1**	**Variable 2**	**Variable 3**	**...**	**Variable n**
Object 1	Value 11	Value 12	Value 13	...	Value 1n
Object 2	Value 21	Value 22	Value 23	...	Value 2n
Object 3	Value 31	Value 32	Value 33	...	Value 3n
...	...	...	...	...	...
Object m	Value m1	Value m2	Value m3	...	Value mn

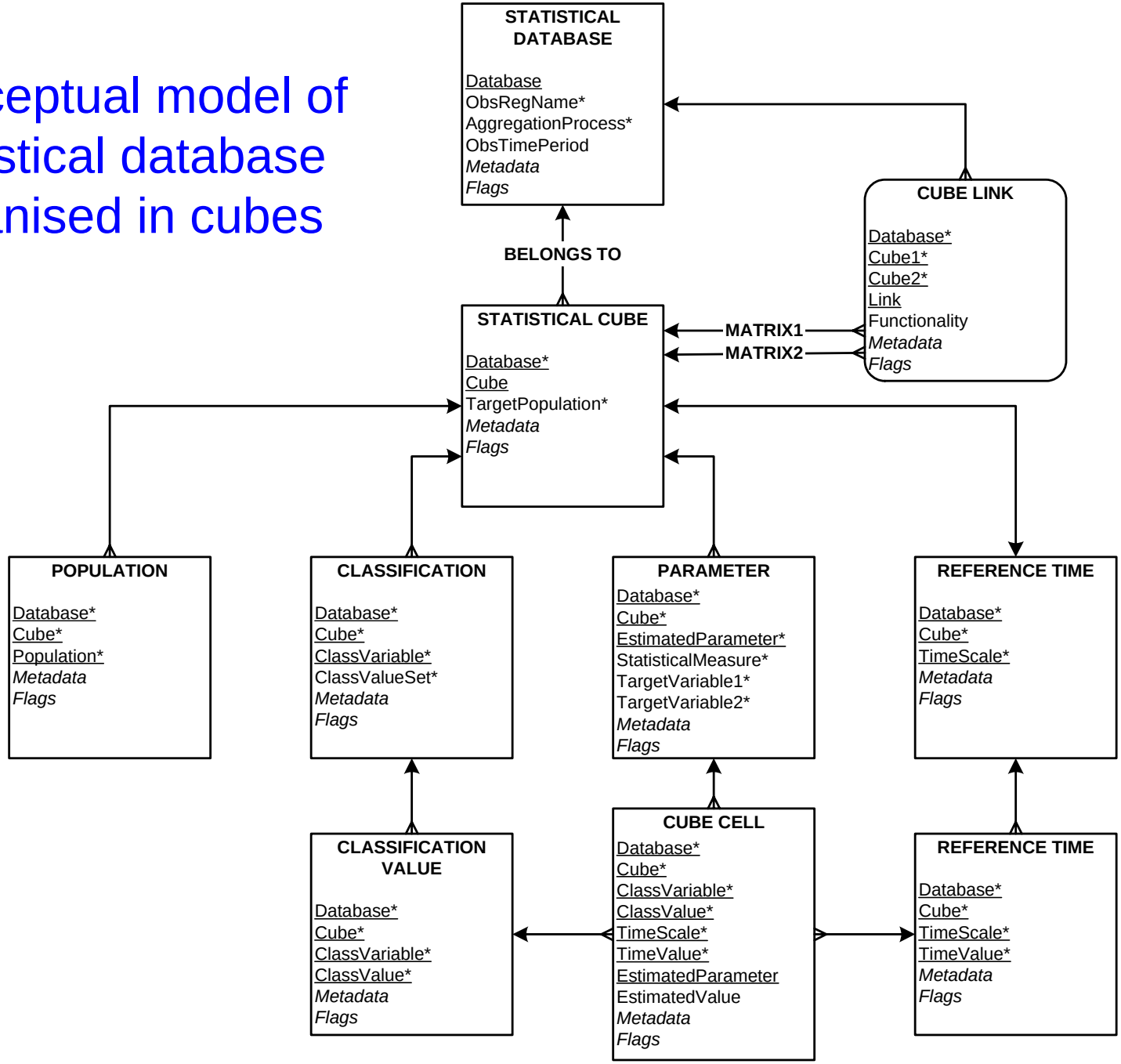
Conceptual data model (metadata model) for a raw data input form



Conceptual model of an observation register organised in data matrixes



Conceptual model of statistical database organised in cubes



”Errors” caused by design decisions

- The *“ideal” populations* may be replaced by *target populations*, the objects of which can be identified and located by means of some existing registers. As a result, some objects belonging to the ideal population may be missed (*undercoverage*), and some others, which do not belong to the ideal population may be included in the target population (*overcoverage*).
- The observations may be limited to samples of objects. This leads to *sampling errors*.
- The *“ideal” variables* may be replaced by operationalised *target variables*, which are easier measure, but which may be less relevant. For example, the income reported by a person in her income statement to the tax authorities will be easier to measure than the “true” income, but it will exclude income components that are not subject to taxation, and those which have to be reported may be systematically underestimated.

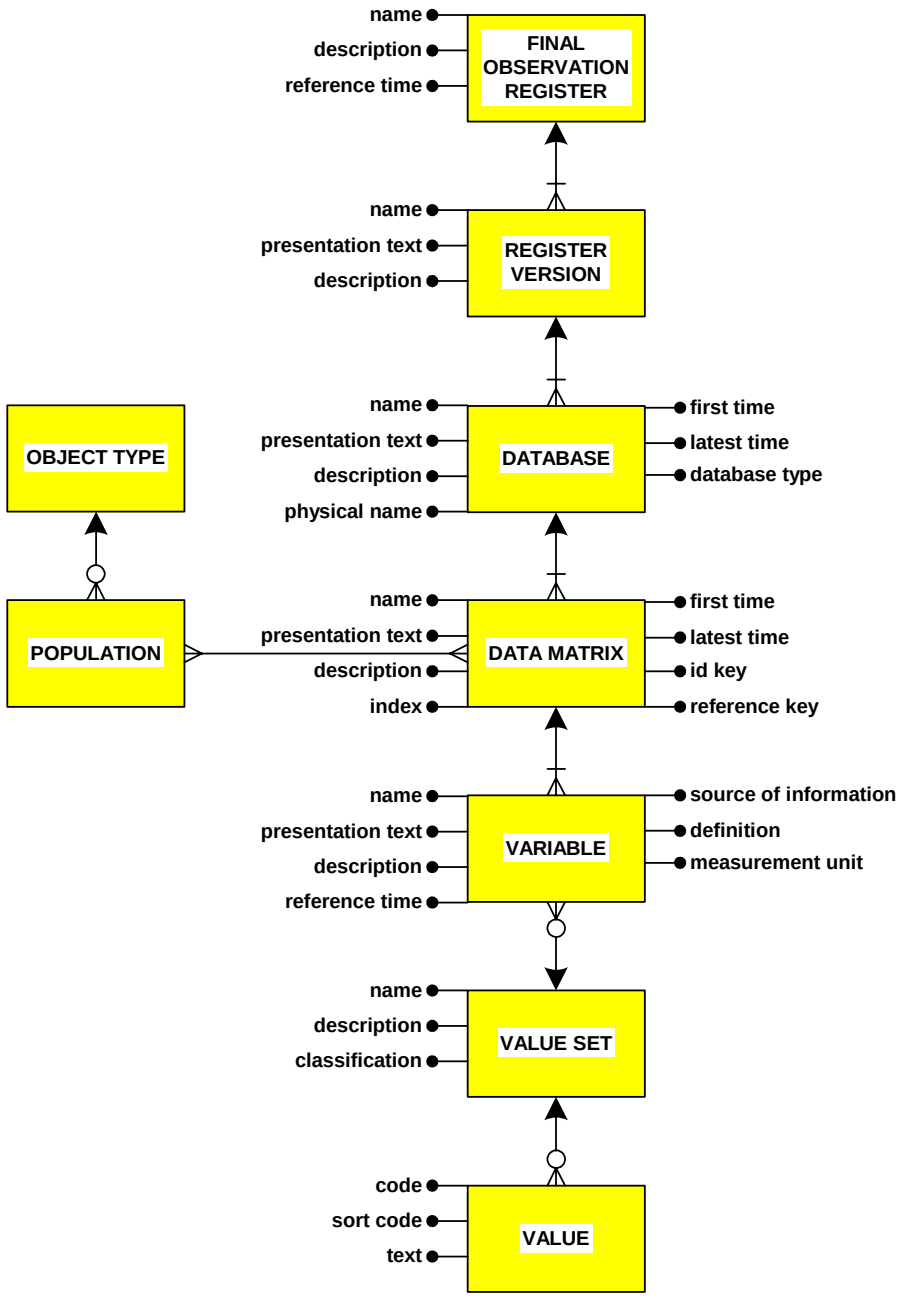
”Errors” occurring during operations

- A *second form of coverage error* may occur, if it has to be determined during the observation process, whether a certain object belongs to the target population or not. For example, an interviewer may have to ask a respondent whether he or she has a certain *inclusion property* or not, and of course the response to such a question are subject to the same kind errors as responses to other questions (see next items in this list).
- It may *not* be *possible to locate* some respondents and/or objects to be observed.
- Respondents *refuse* to answer questions.
- Respondents *misinterpret* questions in the sense that they do not interpret them as intended by the designers.
- Respondents give *wrong answers*, intentionally or by mistake.
- Further *processing errors* may occur during the processing of collected data, e.g. because of human mistakes or technical dysfunctions.

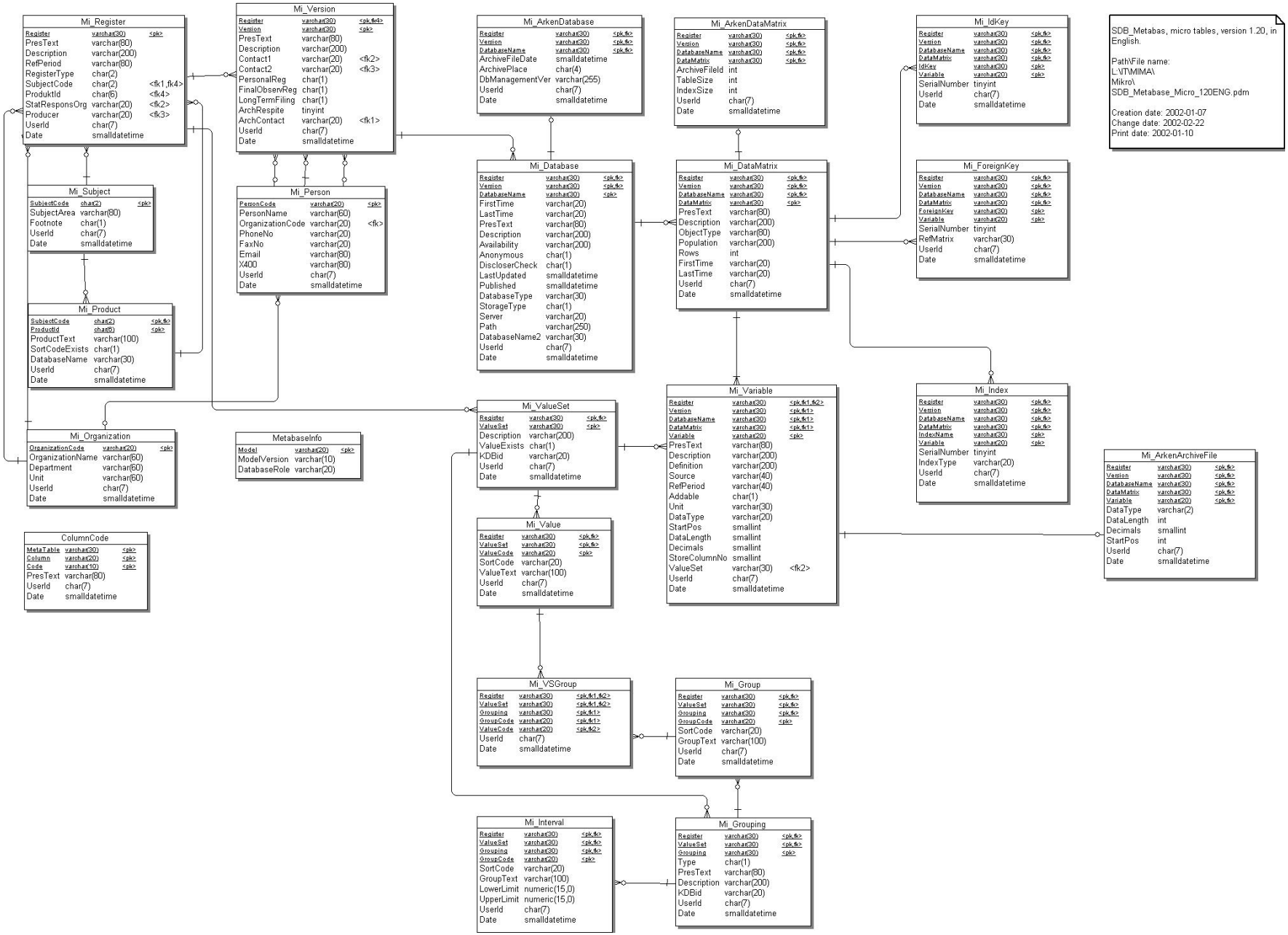
User (mis)interpretations

- Users will have other frames of reference than the designers and operators of the statistical system
- Explanations and documentation accompanying the statistical data may be insufficient or difficult to understand for the users
- Users may have insufficient understanding of statistical methodology

The MicroMeta Model (SCBDOK/METADOK)



The MicroMeta relational data model for final observation registers



Rules & Tools
(Rules embedded in tools)

- models
- mathematical expressions, formulae, and systems of equations
- graphs and flows
- structured programs
- decision tables and decision trees

R&T OBJECT	RULES	TOOL
Register	Rules controlling an object's belonging to a population, a subdomain of interest, and/or a statum Rules associating (a) objects in the target population with (b) objects to be observed, and (c) respondents to provide the observation data	The register as such with its pre-observed and pre-stored values of variables controlling the classification of objects The register and associated software
Questionnaire	Rules controlling valid values of variables Rules controlling which questions a certain user should answer in which order	Closed questions, e.g. multiple choice questions The so-called routing structure of the questionnaire
Classification	Rules determining an object's belonging to a certain class Rules determining how observation data and/or aggregated data based on one classification should be transformed into data based upon another classification (version)	A classification database in combination with registers associated with the classification, e.g. a business register containing main activity codes for the objects in the register So-called correspondence tables in a classification database and/or distribution keys with associated software for schematic recoding and recomputation of data
Coding R&T	Rules determining the classification of free-text answers to open questions into predefined alternatives	Classification databases, dictionaries/thesauri, and associated software for automatic or computer-assisted coding
Editing R&T Imputation R&T	Rules determining the validity and reasonability of certain (combinations of) answers to certain (combinations of) questions Rules suggesting more reasonable answers to certain questions than have been given (or not given) by the respondents	- Classification databases providing valid values of variables - Conceptual data models indicating valid and non-valid relationships between objects - Databases providing multidimensional distributions of values for combinations of variables - Supporting software
Estimation R&T	Models of assumptions made and estimation formulae based on the models	Software procedures
Confidentiality protection R&T	- Secrecy laws and policies - Models of assumptions made and models for computing disclosure risks, given the assumptions - Rules for actions to be taken, given the models and the estimated disclosure risks	- Software procedures - Metadata about the availability of background data that a potential intruder could use for compromising the confidentiality of data

Figure 24. Examples of R&T metadata objects with implied rules and supporting tools.

Statistical metadata may be stored…

- in direct physical connection with the metadata object they describe, e.g. in a data record in a file or a database
- in some kind of annex referred to from the described metadata object, e.g. an annex of footnotes or comments referred to by some kind of “flags” in a data record
- in an autonomous metadata holding, containing metadata referred to and shared by several metadata objects, e.g. names and descriptions of values in a value set, stored in a classification database shared by all data collections with variables using a particular value set or classification
- in a separate log file, as a more or less temporary holding of metadata concerning events that occur intermittently and more or less frequently

Processes are related to metadata in three different ways

1. processes *use metadata* (about themselves and about other metadata objects)
2. processes *produce metadata* (about themselves and about other metadata objects)
3. processes are *objects of metadata*, carriers of metadata about themselves

Processes and procedures

- Processes are described and explained in terms of procedures
- Procedures are for processes what concepts are for data – abstractions
- A procedure has a function/definition and an operational implementation
- The function may be described by a mathematical formula or a mathematical model
- The operational implementation may be described by means of an algorithm or a heurithm
- A process may be seen as an instantiation of a procedure

Input processes

- frame establishment procedure
- frame procedure
- sampling procedure
- measurement procedure
- microdata preparation procedures
- finalisation of sets of microdata+metadata (matrixes etc)

Thruput processes

- data integration procedures
- aggregation and estimation procedures
- macrodata transformation procedure
- finalisation of sets of macrodata+metadata (cubes etc)

Output processes

- analytical procedures
- compilation procedures
- confidentiality protection procedures
- layout and presentation procedures
- publishing procedures
- dissemination procedures
- user support procedures

Process data

- observations and reflections by human operators
- process data generated by computerised processes: detailed and aggregated/ analysed

Instrumental metadata resources

- ***Primary instrumental metadata resources***, enabling and supporting different kinds of processes associated with a statistical system
  - ***production tools, "rules and tools" (R&T),*** supporting primarily the production processes of the statistical system
  - ***search and retrieval tools***, supporting primarily use processes
  - ***knowledge resources, knowledge and experiences (K&E),*** supporting primarily "intellectual" processes, such as planning and evaluation (P&E), and research and development (R&D)
  - ***administrative data***, supporting primarily the managerial processes
- ***Secondary instrumental resources***, being systematised and reasonably complete *sets* of other metadata resources, which are organised in such way that they can be shared by many statistical systems, or, preferably, by the statistical organisation as a whole: **corporate metadata resources**

Instrumental metadata resources (process enablers)

- *search and retrieval tools* supporting use processes and other processes that need access to statistical data (and metadata)
- *production tools* supporting production processes and procedures
- *knowledge resources* (knowledge and experiences) supporting primarily the “intellectual” processes around statistical systems, such as planning and evaluation, corporate management, and research and development

Many instrumental metadata resources have two characteristics in common

- they are *sharable*, that is, the same metadata resource can be used by multiple processes
- they need to be *systematised and organised collectively*, in order to be easy to find and make use of

In order to ensure that the instrumental metadata resources of a statistical system satisfy the requirements of sharability and efficient availability, one may organise a number of *metadata repositories* as part of – and core of – the corporate data/metadata infrastructure of a statistical organisation. Thus we have identified a fourth category of instrumental data/metadata resources of a statistical system:

- *corporate data/metadata resources* containing sharable, systematised, and reasonably “complete” metadata resources of a certain kind

Three very important metadata resources with multiple roles

- statistical registers
- observation templates (questionnaires)
- classifications (and other value sets)

Example: classification databases

- organise a central unit responsible for all aspects of standard classifications – with a possibility to delegate some maintenance and operations of a certain classification to a unit in the organisation that is the main user of the classification
- establish certain corporate policies and rules as to how non-standard classifications should be managed, maintained, and operated by local users
- ensure that historical and up-to-date versions of all classifications (including non-standard value set used by single surveys), or replicates of them, are made available to the organisation as a whole in a practical way
- use standardised design solutions and software for all classification databases in the organisation, wherever they are physically stored, and whoever is responsible for them

Two kinds of secondary metadata resources

- listings of metadata objects
  - catalogues, registers, directories, indexes
- repositories of metadata object descriptions
  - databases, libraries, archives

Metadata object type	*Examples of subtypes*	*Metadata object descriptions*	*Links to*
Production tool (Rules&Tools object) (R&T object)	Register Questionnaire Classification Coding R&T Editing R&T Estimation R&T	Rules&Tools object documentation	
Search and retrieval tool	Search engine Dictionary, Thesaurus Conceptual map - object graph - process graph - system flow	Search and retrieval tool documentation	Catalogues
Knowledge& Experiences object (K&E object)	Theory Method Practice Standard Law Administrative rule or procedure	Documented knowledge and/or experiences about the K&E object	Related metadata objects
Administrative data object	Person Organisational unit IT resource Equipment resource Office resource	Administrative data	Non-statistical enterprise systems for administrative data, e.g. personnel and accounting systems

Figure 23. Overview of metadata about some primary instrumental metadata resources.

Different ways of organising the corporate data/metadata resources

- in a completely centralised way, with centrally placed organisational units being fully responsible for them, or as decentralised networks, mainly controlled by standardised interfaces and procedures, or as federated databases
- example: classification databases

Organising and maintaining the data/metadata infrastructure

- How to organise and store microdata, macrodata, and associated metadata
- How to obtain and capture statistical metadata and how to keep them updated

Criteria for choosing place

- redundance considerations
- access speed considerations
- formally structured metadata vs free-text
- compactness
- procedural vs non-procedural metadata
- natural birth events
    - design decisions
    - operation process events

Levels for storing metadata in direct physical connection with the data

- for observation data
  - an elementary observation in a collection of observation data
  - a complex observation corresponding to an input form instance
  - a row or a column in an observation matrix
  - an observation matrix
  - an observation register or database
- for statistics
  - an estimated value of an elementary statistical characteristic (a cell in a cube)
  - a classification dimension of a cube
  - the population dimension of a cube
  - a parameter dimension of a cube
  - a cube
  - a collection of cubes or statistical database

Developing a data/metadata infrastructure for a statistical organisation

- A development and implementation stategy
  - Benchmarking – “business intelligence”
  - “Who? Why? What? How?” analysis
  - A vision of the “ideal” statistical metainformation system
  - The road towards the vision: priority setting
  - Detailed planning
- Golden rules for statistical metadata projects
  - If you are a designer…
  - If you are the project co-ordinator…
  - If you are the top manager…

Who? Why? What? How?

- Who are the users of the corporate data/metadata infrastructure?
- Why, for which purposes, do they need the data/metadata infrastructure?
- What kind of data/metadata do they require from the infrastructure?
- How can the data/metadata infrastructure be created and maintained?

If you are a designer…

- Make metadata-related work an integrated part of the business processes of the organisation.
- Capture metadata at their natural sources, preferably as by-products of other processes.
- Never capture the same metadata twice.
- Avoid un-coordinated capturing of similar metadata – build value chains instead.
- Whenever a new metadata need occurs, try to satisfy it by using and transforming existing metadata, possibly enriched by some additional, non-redundant metadata input.
- Transform data and accompanying metadata in synchronised, parallel processes, fully automated whenever possible.
- Do not forget that metadata have to be updated and maintained, and that old versions may often have to be preserved.

If you are the project co-ordinator…

- Make sure that there are clearly identified “customers” for all metadata processes, and that all metadata capturing will create value for stakeholders.
- Form coalitions around metadata projects.
- Make sure that top management is committed. Most metadata projects are dependent on constructive co-operation from all parts of the organisation.
- Organise the metadata project in such a way that it brings about concrete and useful results at regular and frequent intervals.

If you are the top manager…

- Make sure that your organisation has a metadata strategy, including a global architecture and an implementation plan, and check how proposed metadata projects fit into the strategy.
- Either commit yourself to a metadata project – or don't let it happen. Lukewarm enthusiasm is the last thing a metadata project needs.
- If a metadata project should go wrong – cancel it; don't throw good money after bad money.
- When a metadata project fails, make a diagnosis, learn from the mistakes, and do it better next time.
- Make sure that your organisation also learns from failures and successes in other statistical organisations.
- Make systematic use of metadata systems for capturing and organising tacit knowledge of individual persons in order to make it available to the organisation as a whole and to external users of statistics.