

Introduction to the Danish model for quality reports

Component A: Quality Management of Official Statistics

Activity A.3: Quality reports, tools and methods

12. – 15. September 2016

Karin Blix, Quality Coordinator, Statistics Denmark, kwb@dst.dk

Peter Stoltze, Head of Division, Statistics Denmark, psl@dst.dk



From ToR

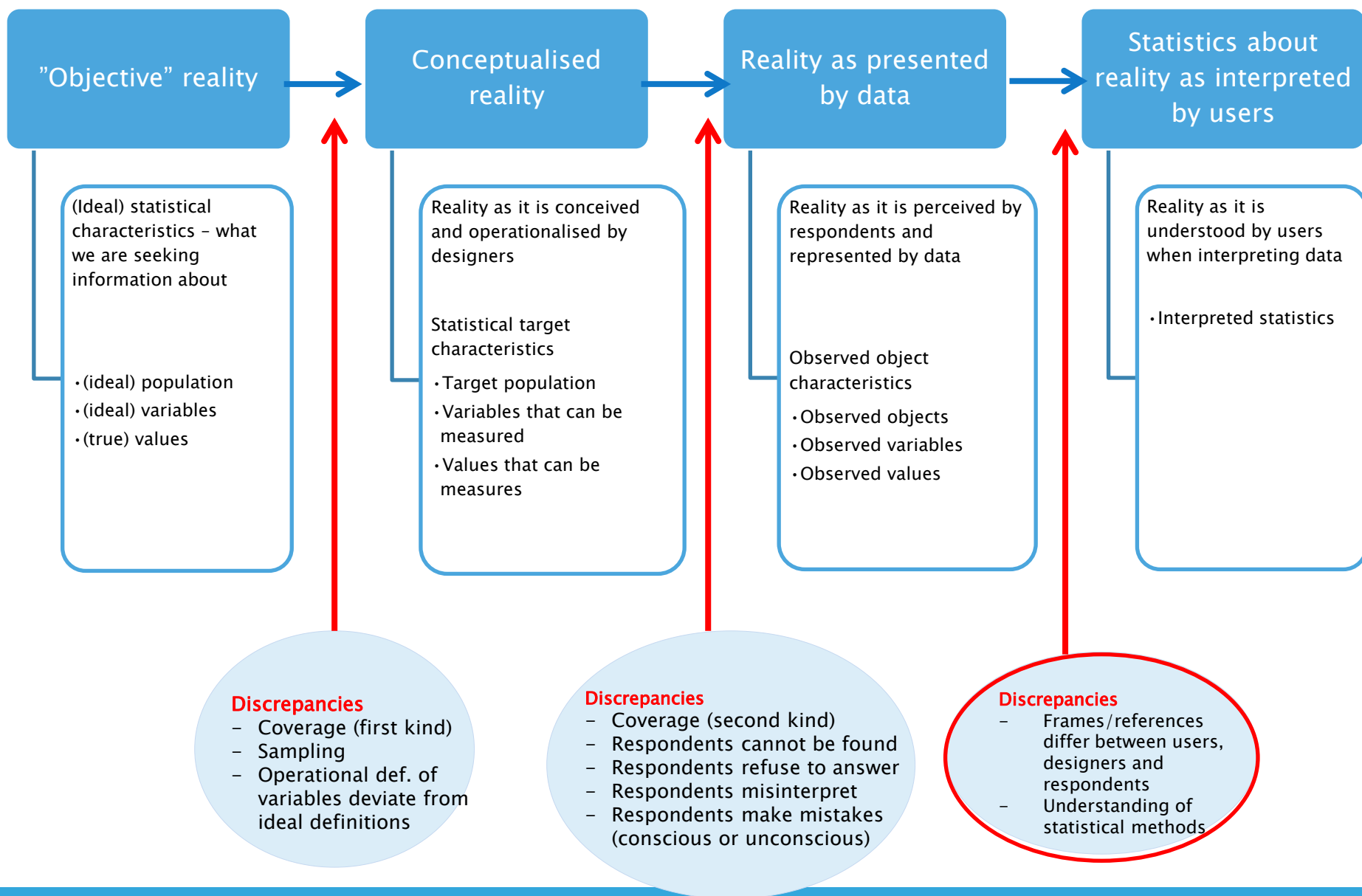
- ***Introduction to the Danish model for quality reports, including training and support to subject units and responsible for webpage metadata***

What is **quality** in official statistics?

- Good or bad quality?
- **Trust** is the key
- UN Fundamental Principles of Official Statistics (1993; 2014)
- European Statistics Code of Practice (2005; 2011)
- EU Quality Assurance Framework (2013)
- OECD Quality Framework (2012)

Producing statistics

- Producing statistics is about describing phenomena in the society
- Not just anything – but some important phenomena in the society – something that some users seek information about
 - Population
 - Education
 - The economy
 - Etc.
 - An illustration is given by the Swede Bo Sundgren in *Statistical systems: Some fundamentals*



Reporting quality to users – Documentation of statistics

- Help for the user to understand the statistics – giving the user information about the frame we have worked within
- Explain the content of the statistics
 - History
 - Purpose of the statistics
 - Content – population, variables etc.
- Quality = Fitness for use
- Quality of contents:
 - Relevance, Accuracy & reliability, Timeliness and punctuality, Coherence & comparability, Accessibility and clearness

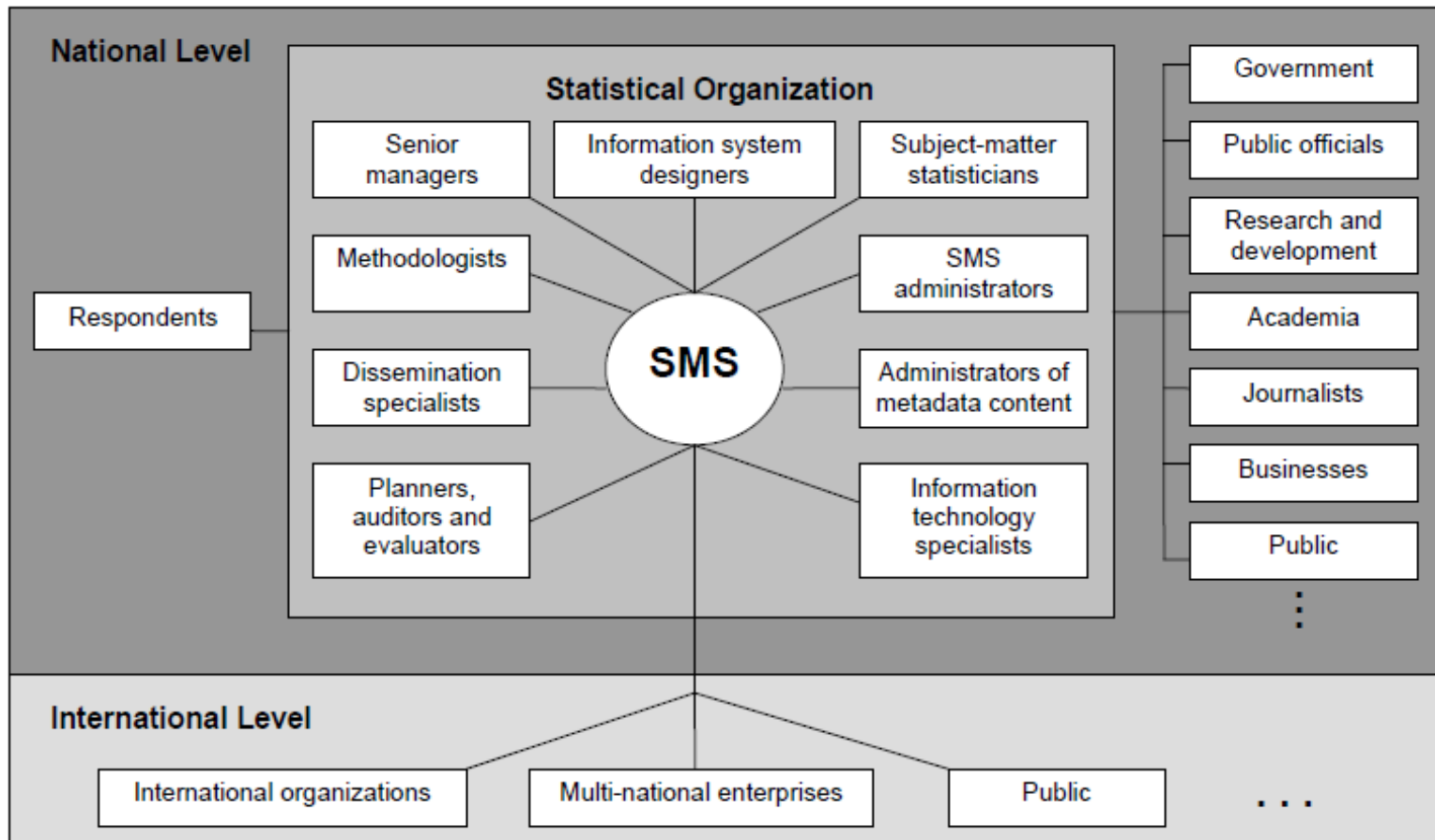
Streamlining and harmonising metadata and quality reporting

- Once for all purposes reporting
 - Each concept is only reported upon once and is re-usable
- Integrated and consistent quality and metadata
 - Reporting framework where the reports are stored in the same database
- A flexible and up to date system
 - Where future extensions are possible by adding new concepts,
- “Single Integrated Metadata Structure” (SIMS)
 - A dynamic and unique inventory of ESS quality and metadata statistical concepts has been created
- In this structure, all statistical concepts of the two existing ESS report structures (ESMS and ESQRS) have been included and streamlined, by assuring that all concepts appear and are therefore reported upon only once

Our users

- Difficulty understanding the content of the quality declarations
- “Front page fields” – easy to understand extraction
- Need information on breaks in data series, information on revisions etc.
- Need of better integration between variables, concepts, classification and quality declarations including links to relevant statistical information

Metadata users



The statisticians

- The quality declarations are made in the "last minute"
- Common "rules" of user friendly content lacking
- Link of content in quality declarations to Statistics Denmark's Process model (based on GSBPM)
- Different reporting of quality for EU, IMF, DST

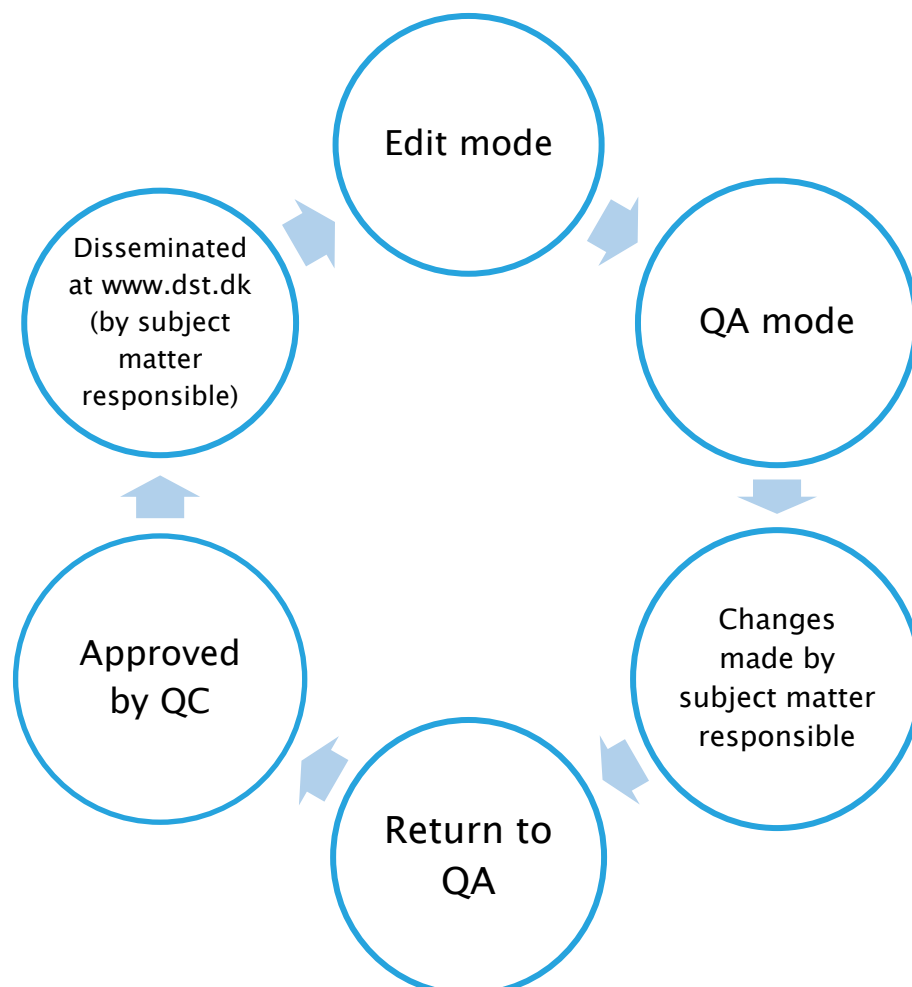
Documentation of Statistics

- Re-organisation 2014 following the ESS handbook
- Three levels
 1. “Front page” to appear at the webpage of Statistics Denmark, with a short description of the 9 headlines in the Structure. From the front page one can open around 100 specified topics (SIMS)
 2. SIMS topics cover the more detailed quality report (see guidelines in Annex 2). From level 2 one can open annexes for further description
 3. Annexes
- The idea is in one product to cover all customers (national, international, EU).
- Prepared in Danish and English

Documentation of statistics

- Cycle:
 - Every new dissemination of statistics followed by an updated quality declaration
 - Send to QA
 - QA reads and comments
 - Dissemination centre reads and comments on "front-pages"
 - Responsible updates – new QA
 - Ready for dissemination

Cycle for Documentation of statistics



Documentation of statistics on www.dst.dk

The screenshot shows the Statistics Denmark website with the URL <http://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/retail-trade-index> in the browser address bar. The page features the Statistics Denmark logo and a navigation menu with links to CONTACT, PRESS, INFORMATION SERVICES, and DANSK. A search bar is also present. The main content area is titled "Retail Trade Index" and includes a description of the index, contact information for Lina Pedersen, and links to statistical presentation and processing. A sidebar on the left lists various documentation topics, and a "Denmark in figures" box is on the right.

Documentation of statistics

- Retail Trade Index**
- Statistical presentation
- Statistical processing
- Relevance
- Accuracy and reliability
- Timeliness and punctuality
- Comparability
- Accessibility and clarity

Classifications

- Methodology
- Major revision of national accounts

Retail Trade Index

The Retail Trade Index shows the development in turnover within the retail trade sector. The statistics is published monthly and is primarily used as short term indicator for private consumption as well as the general business cycle movement.

Contact info

Short Term Statistics, Business Statistics
Lina Pedersen
+45 39 17 36 75
lip@dst.dk

Get as PDF

[Retail Trade Index](#)

[Show archive](#)

Denmark in figures

Get a quick overview of the developments in the Danish society in this colourful and easily read booklet.

[Denmark in Figures](#)

Statistical presentation

Retail trade indices are published for 43 industries and for three main commodity groups: food and other everyday commodities, clothing etc., and other commodities. Furthermore special industry aggregates are produced for Eurostat. Value as well as volume indices are produced. The volume indices only for the main commodity groups and the indices for Eurostat. The statistics is based on survey data from all large retail trade enterprises and a sample of the remaining retail trade enterprises. seasonal adjustment is performed on the main commodity groups and the total.

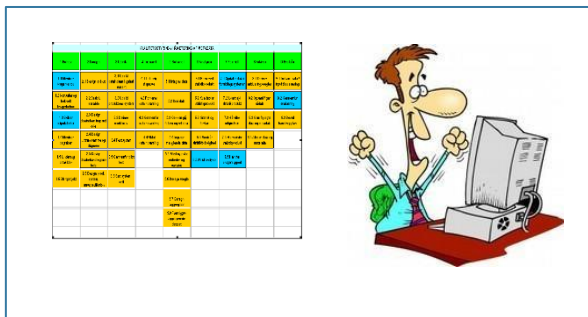
[Read more about statistical presentation](#)

Statistical processing

Turnover figures are collected each month from a sample of 3.500 retail trade enterprises, reporting the figures either by web or by dial-in. The submitted data is error checked by a number of measures, e.g. by comparing the turnover growth in similar enterprises.

Survey data is grossed up in part by including information from administrative sources on the VAT turnover in the previous quarter for the entire population.

Enter SIMS fields



Publish at www.dst.dk

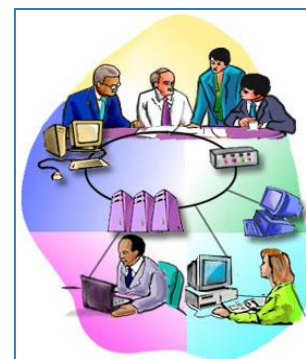


**METADATA
IN
COLECTICA**

Send Quality report to EU



Publish at the Intranet



EURO-SDMX Metadata Structure (Dec 2010)

1	Contact
2	Metadata update
3	Statistical presentation
4	Unit of measure
5	Reference period
6	Institutional mandate
7	Confidentiality
8	Release policy
9	Frequency of dissemination
10	Dissemination format
11	Accessibility of documentation
12	Quality management
12.1	Quality assurance
12.2	Quality assessment
13	Relevance
14	Accuracy and reliability
15	Timeliness and punctuality
16	Comparability
17	Coherence
18	Cost and burden
19	Data revision
20	Statistical processing
21	Comment

Single Integrated Metadata Structure

S.1	Contact
S.2	Introduction
S.3	Metadata update
S.4	Statistical presentation
S.5	Unit of measure
S.6	Reference period
S.7	Institutional mandate
S.8	Confidentiality
S.9	Release policy
S.10	Frequency of dissemination
S.11	Dissemination format, Accessibility and clarity
S.12	Accessibility of documentation
S.13	Quality management
S.13.1	Quality assurance
S.13.2	Quality assessment
S.14	Relevance
S.15	Accuracy and reliability
S.16	Timeliness and punctuality
S.17	Comparability
S.18	Coherence
S.19	Cost and burden
S.20	Data revision
S.21	Statistical processing
S.22	Comment

ESS Standard for Quality Reports Structure

I	Contact
II	Introduction
XI	Confidentiality
VII	Accessibility and clarity
III	Quality assessment
IV	Relevance
V	Accuracy and reliability
VI	Timeliness and punctuality
VIII	Comparability
IX	Coherence
X	Cost and Burden
V.3.6	Data revision
XII	Statistical Processing
XIII	Comment

ESS handbook - Purpose

Handbook is addressed to

- NSO for their own internal assessment of process and output quality
- NSO as the starting point for preparing user-oriented quality reports
- NSO for producer-oriented quality reports to Eurostat

Single metadata structure should promote

- Both user-oriented and producer-oriented should be derived from the same source
- Maximum re-use of information in the metadata system
- Reduction and simplification of documents
- The user-oriented quality reports should be improved

ESS Handbook - Structure

1. Introduction
2. Relevance, assessment of user needs
3. Accuracy and reliability
4. Timeliness and punctuality
5. Accessibility and clarity
6. Coherence and comparability
7. Cost and burden
8. Confidentiality
9. Statistical processing

2-6: output components 7-9: process components

Statistical Processes

The ESS Handbook applies to the following statistical processes:

1. Sample survey
2. Census
3. Statistical process using administrative source(s)
4. Statistical process involving multiple data sources
5. Price or other economic index process
6. Statistical compilation assembling a variety of primary sources (e.g. National Accounts)

Guidelines for preparing detailed quality reports

- For all 9 headlines in the structure
(Relevance, Accuracy ...)
- For all statistical process on the whole
and where relevant
- For the 6 types of statistical processes
(Sample Survey, Census ...)

Content of the quality declarations

1 Introduction (s2):

2 Statistical presentation (s4):

2.1 Data description (s4.1):

2.2 Classification system (s4.2):

2.3 Sector coverage (s4.3):

2.4 Statistical concepts and definitions (s4.4):

2.5 Statistical unit (s4.5):

2.6 Statistical population (s4.6):

2.7 Reference area (s4.7):

2.8 Time coverage (s4.8):

2.9 Base period (s4.9):

2.10 Unit of measure (s5):

2.11 Reference period (s6):

2.12 Frequency of dissemination (s10):

2.13 Legal acts and other agreements (s7.1):

2.14 Cost and burden (s19):

2.15 Comment (s22):

Content of the quality declarations

3 Statistical processing (s21):

3.1 Source data (s21.1):

3.2 Frequency of data collection (s21.2):

3.3 Data collection (s21.3):

3.4 Data validation (s21.4):

3.5 Data compilation (s21.5):

3.6 Adjustment (s21.6):

4 Relevance (s14):

4.1 User Needs (s14.1):

4.2 User Satisfaction (s14.2):

4.3 Data completeness rate (s14.3):

Contents (cont.)

5 Accuracy and reliability (s15):

5.1 Overall accuracy (s15.1):

5.2 Sampling error (s15.2):

5.3 Non-sampling error

5.4 Quality management (s13):

5.5 Quality assurance (s13.1):

5.6 Quality assessment (s13.2):

5.7 Data revision - policy (s20.1):

5.8 Data revision practice (s20.2):

6 Timeliness and punctuality (s16):

6.1 Timeliness and time lag - final results (s16.1):

6.2 Punctuality (s16.2):

Contents (cont.)

7 Comparability (s17):

- 7.1 Comparability - geographical (s17.1):
- 7.2 Comparability over time (s17.2):
- 7.3 Coherence - cross domain (s18.1):
- 7.4 Coherence - internal (s18.2):

8 Accessibility and clarity (s11):

- 8.1 Release calendar (s9.1):
- 8.2 Release calendar access (s9.2):
- 8.3 User access (s9.3):
- 8.4 News release (s11.1):
- 8.5 Publications (s11.2):
- 8.6 On-line database (s11.3):
- 8.7 Micro-data access (s11.4):
- 8.8 Other (s11.5):
- 8.9 Confidentiality - policy (s8.1):
- 8.10 Confidentiality - data treatment (s8.2):
- 8.11 Documentation on methodology (s12.1):
- 8.12 Quality documentation (s12.2):

Contents (cont.)

9 Contact (s1):

9.1 Contact organisation (s1.1):

9.2 Contact organisation unit (s1.2):

9.3 Contact name (s1.3):

9.4 Contact person function (s1.4):

9.5 Contact mail address (s1.5):

9.6 Contact email-address (s1.6):

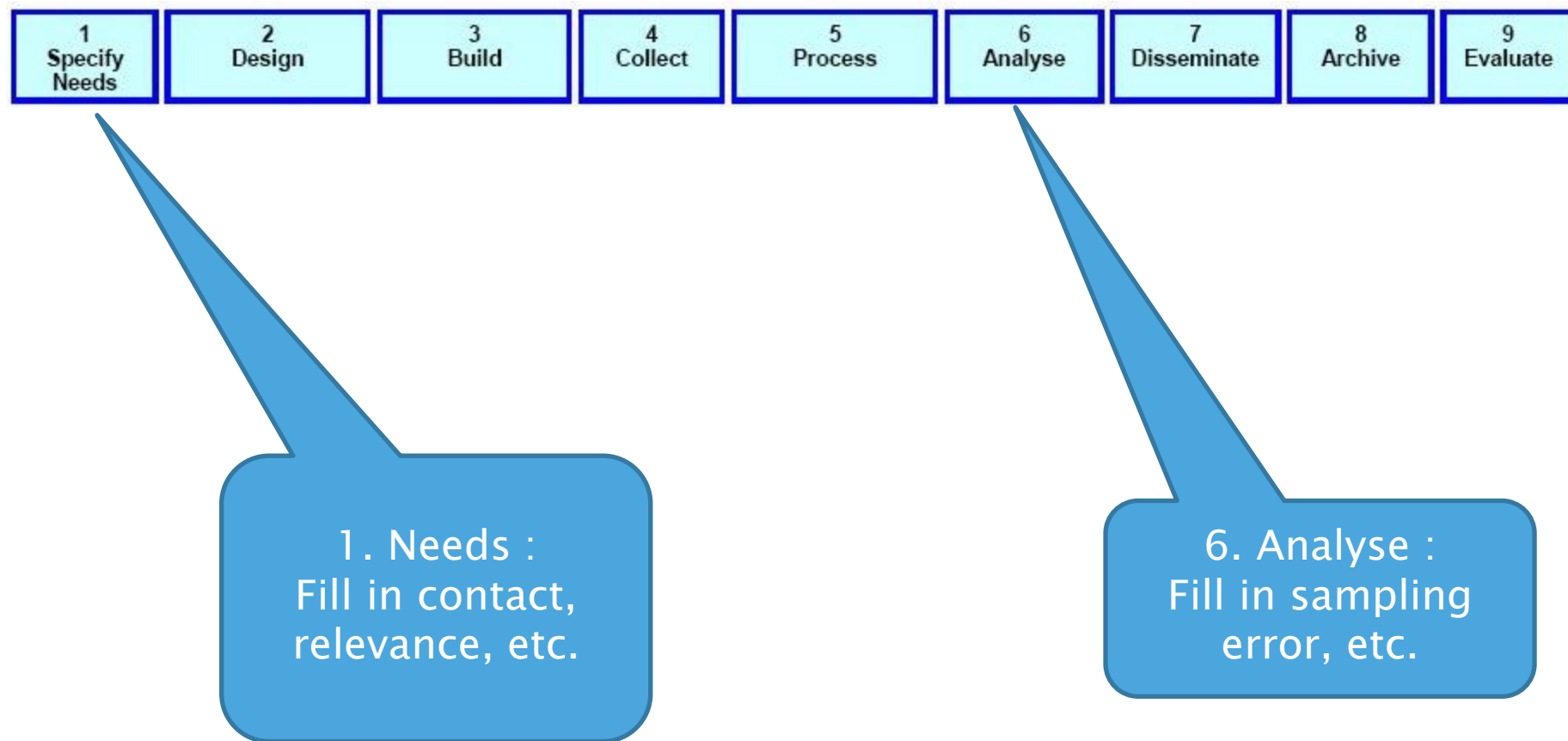
9.7 Contact phone number (s1.7):

9.8 Contact fax number (s1.8):

Process model

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Work processes and quality declarations



Quality and Performance Indicators

16 quality indicators to be reported in the ESS Quality Reports to Eurostat

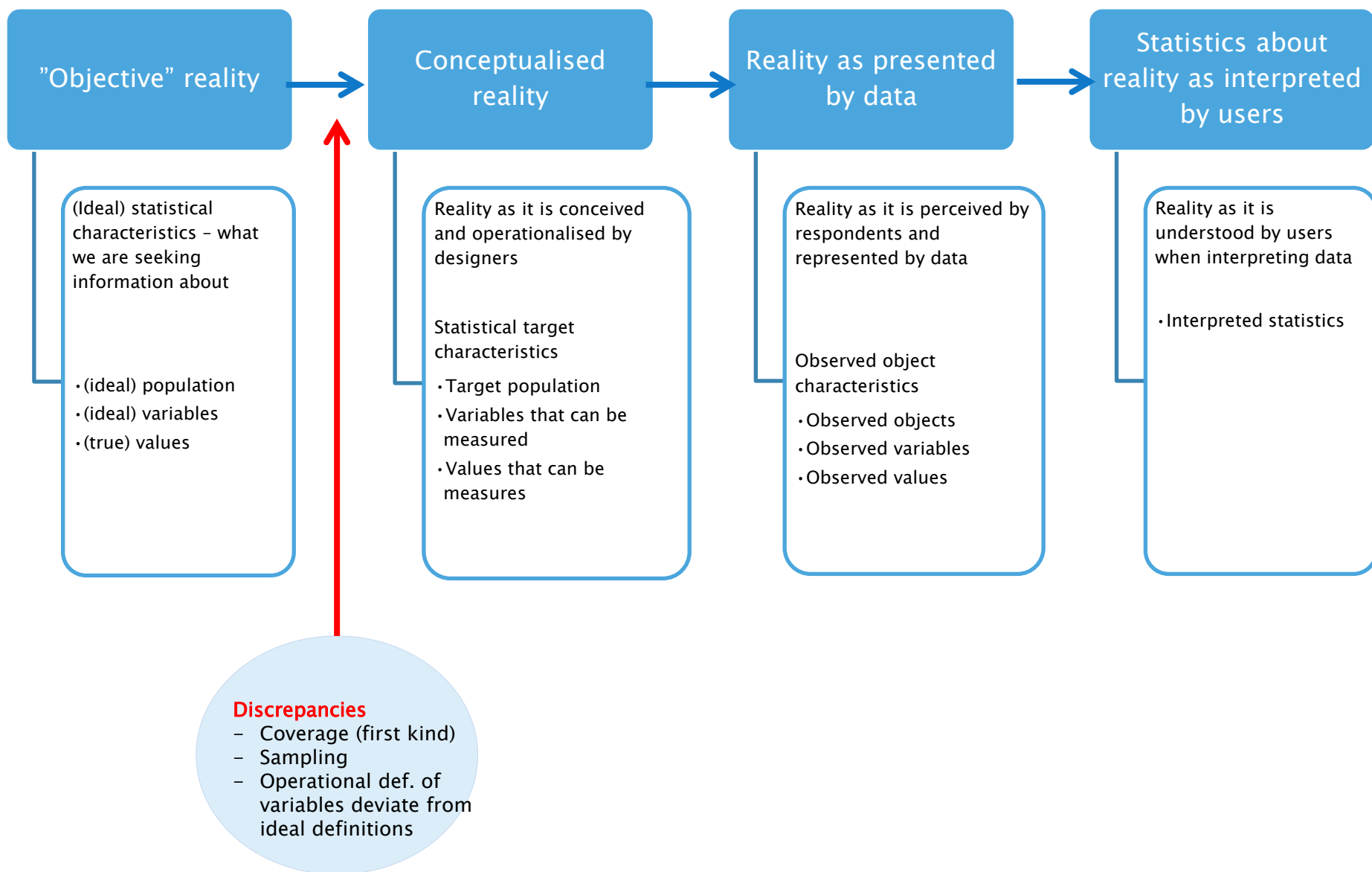
- | | |
|-------------------------------|-----------|
| • Relevance | R1 |
| • Accuracy and Reliability | A1 – A7 |
| • Timeliness and Punctuality | TP1 – TP3 |
| • Coherence and Comparability | CC1 – CC2 |
| • Accessibility and Clarity | AC1 – AC3 |

(cover the output components in slide 7)

U (user) and P (producer) in SIMS

- P-fields are *producer-oriented* and are normally just the figure(s) asked for
- U-fields are *user-oriented* and are an annotated version of the figure(s) asked for

Indicator	U SIMS ID	P SIMS ID	Indicator	U SIMS ID	P SIMS ID	Indicator	U SIMS ID	P SIMS ID
A1	15.2	15.2.1	A7		21.5.1	CC1		17.1.1
A2		15.3.1.1	TP1		16.1.1	AC1		11.3.1
A3		15.3.1.2	TP2	16.1	16.1.2	AC2		11.5.1
A4	15.3.3	15.3.3.1	CC2	17.2	17.2.1	AC3		12.1.1
A5	15.3.3	15.3.3.2	R1	14.3	14.3.1			
A6	20.2	20.2.1	TP3	16.2	16.2.1			



A1 Sampling error

The sampling error is calculated for key variables

The sampling error is calculated either

- Coefficient of variation $CV(\hat{\theta}) = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}}$ or
- Confidence interval, symmetric around $\hat{\theta}$

A2 Over-coverage

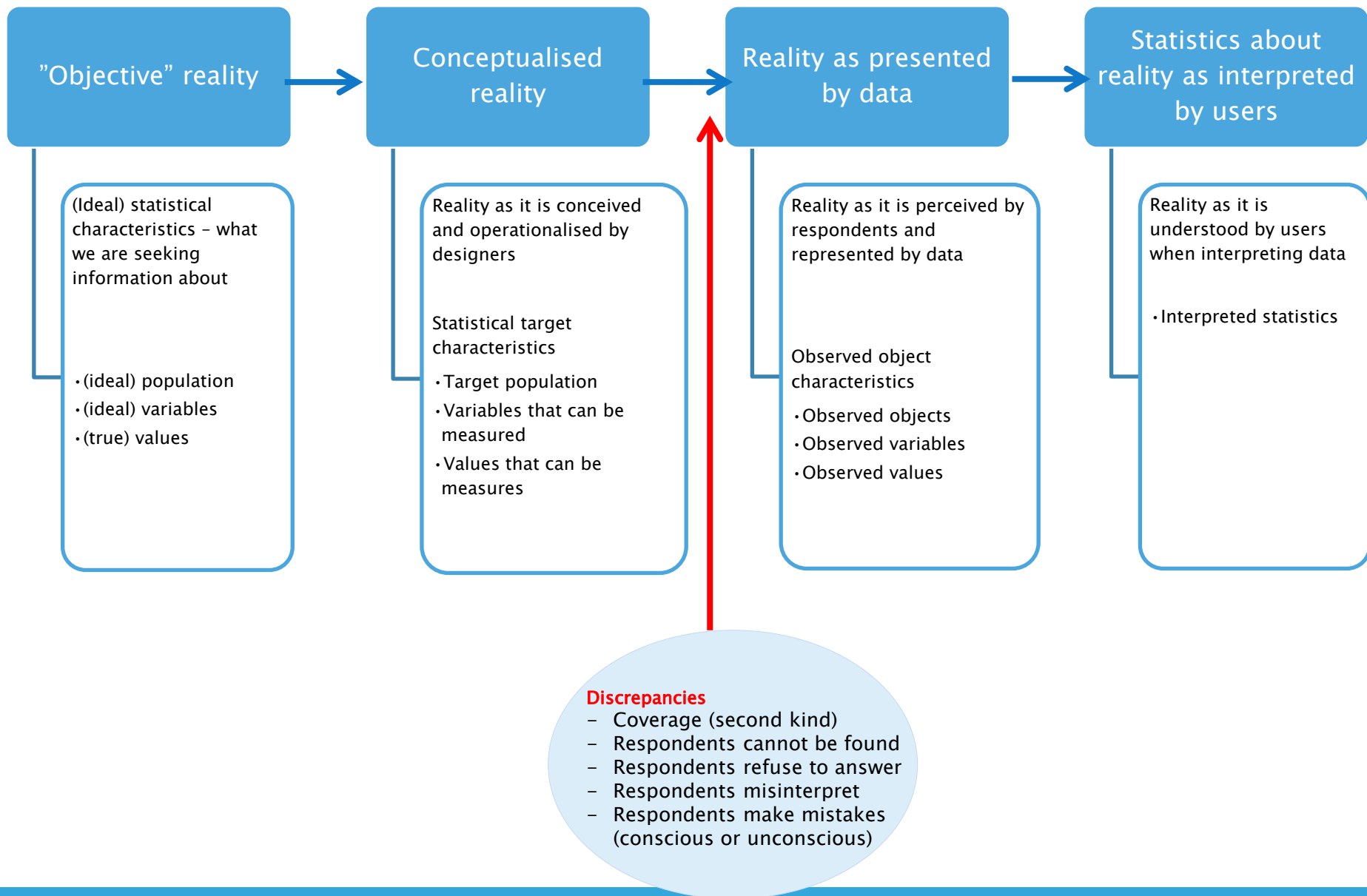
- The rate of over-coverage is the proportion of units accessible via the *frame population* that do not belong to the *target population*.

- Definition:

$$\frac{\text{Units in the } \textit{frame population} \text{ not in the } \textit{target population}}{\text{Units in the } \textit{frame population}}$$

- Three main cases

- Un-weighted rate $(w_j = 1)$
- Design-weighted rate $(w_j = d_j)$
- Size-weighted rate $(w_j = d_j * x_j)$



A3 Common units

Mixed statistical processes where some variables or data for some units come from *survey* data and others from *administrative sources*

- Measure for agreement between different sources
- Definition:

Units in both the *survey* and the *administrative source*

Units in the *survey*

A4 Unit non-response

Unit non-response cover units with no information or not usable information. Unit non-response is a source of error

- It reduces the number of responses
- It might introduce bias

- Definition:

Units with no or not usable information
Units in the *survey*

- Three main cases

- Un-weighted rate $(w_j = 1)$
- Design-weighted rate $(w_j = d_j)$
- Size-weighted rate $(w_j = d_j * x_j)$

A5 Item non-response

- Non response for certain variables
- Key variables and variables with major non-response
- Definition:

Units who should have but do not have answered variable X

Units who should have answered variable X

- Three main cases
 - Un-weighted rate $(w_j = 1)$
 - Design-weighted rate $(w_j = d_j)$
 - Size-weighted rate $(w_j = d_j * x_j)$

A6 Data revision

Can be made very complicated – read it yourself

Example: Quarterly national accounts (DK)

- Revision measured on one variable – GDP
- Two measures calculated
 - First release (P) and release after 1 year (L), as the major changes happens in this period
 - First release (P) and release after 3 year (L), after the yearly National Accounts statistics has been finalized
- Two formulas used (with $n = 12$ quarters)
 - $MAR = \frac{1}{n} \sum_{t=1}^n |X_{Lt} - X_{Pt}|$ as a measure for the variation
 - $MR = \frac{1}{n} \sum_{t=1}^n (X_{Lt} - X_{Pt})$ as a measure for the bias

A7 Imputation

Imputation is the process used to assign replacement values for missing, invalid or inconsistent data

- Rate calculated for key variables
- Definition:

$$\frac{\text{Units where variable X has been imputed}}{\text{Units who should have answered variable X}}$$

- Three main cases
 - Un-weighted rate $(w_j = 1)$
 - Design-weighted rate $(w_j = d_j)$
 - Size-weighted rate $(w_j = d_j * x_j)$

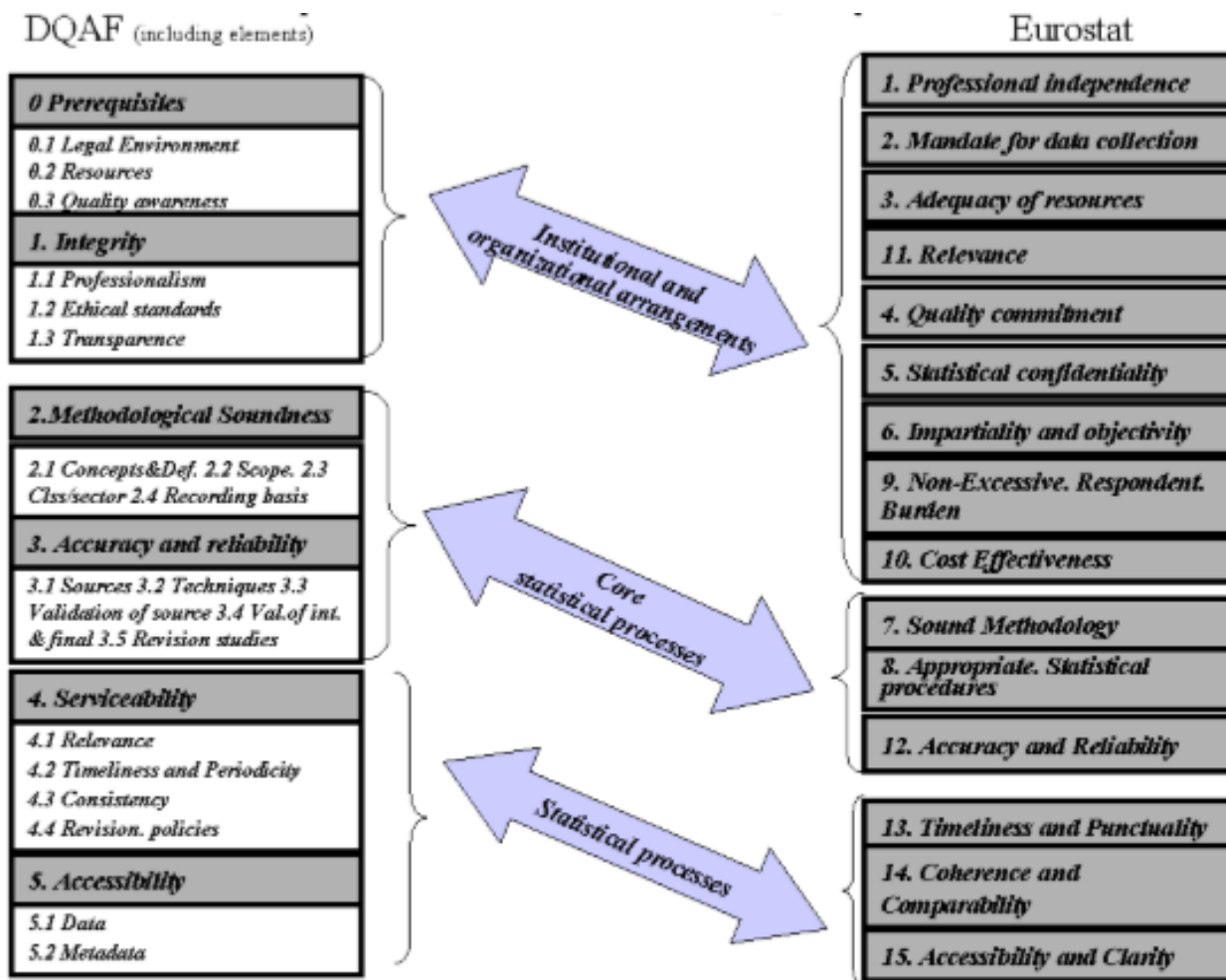
SDDS

- The Special Data Dissemination Standard (SDDS)
- Established in 1996 to guide members that have, or might seek access to international capital markets
- Enhance the availability of timely and comprehensive statistics
- Release calendar
- Detailed information on statistical practices or metadata
- DQAF – data quality assessment framework

DQAF – data quality assessment framework

- Five dimensions
 - assurances of integrity
 - methodological soundness
 - accuracy and reliability
 - serviceability
 - Accessibility
- Special for each of the domains:
 - National Accounts Statistics
 - Consumer Price Index
 - Producer Price Index
 - Government Finance Statistics and Public Sector Debt Statistics
 - Monetary Statistics
 - Balance of Payments and International Investment Position Statistics
 - External Debt Statistics
 - External trade in goods

Comparison of IMF and EU quality dimensions (2006)



o. Prerequisites of quality

0.1 Legal and institutional environment

The responsibility for collecting, processing, and disseminating the statistics is clearly specified, Data sharing and coordination among data-producing agencies are adequate, Individual reporters' data are kept confidential and used for statistical purposes only, Statistical reporting is ensured through legal mandate and/or measures to encourage response.

0.2 Resources

Staff, facilities, computing resources, and financing are commensurate with statistical programs, Measures to ensure efficient use of resources are implemented,

0.3 Relevance

The relevance and practical utility of existing statistics in meeting users' needs are monitored.

0.4 Other quality management

Processes are in place to focus on quality and to monitor quality during the planning and implementation of the statistical program.

1. Assurances of integrity

1.1 Institutional Integrity

Statistics are produced on an impartial basis, data sources and statistical techniques, as well as decisions about dissemination, are informed solely by statistical considerations, the appropriate statistical entity is entitled to comment on erroneous interpretation and misuse of statistics.

1.2 Transparency

The terms and conditions under which statistics are collected, processed, and disseminated are available to the public. Internal governmental access to statistics prior to their release is publicly identified. Products of statistical agencies/units are clearly identified as such. Advance notice is given of major changes in methodology, source data, and statistical techniques.

1.3 Ethical standards

Guidelines for staff behavior are in place and are well known to the staff.

2. Methodological soundness

2.1 Concepts and definitions

The overall structure in terms of concepts and definitions follows internationally accepted standards, guidelines, or good practices.

2.2 Scope

The scope is broadly consistent with internationally accepted standards, guidelines, or good practices.

2.3 Classification/sectorization

Classification/sectorization systems used are broadly consistent with internationally accepted standards, guidelines, or good practices.

2.4 Basis for recording

Market prices are used to value flows and stocks, recording is done on an accrual basis. Grossing/netting procedures are broadly consistent with internationally accepted standards, guidelines, or good practices.

3. Accuracy and reliability

3.1 Source data

Source data are obtained from comprehensive data collection programs that take into account country-specific conditions, reasonably approximate the definitions, scope, sectorization, classifications, valuation, and time of recording required and source data are timely.

3.2 Assessment of source data

Source data—including censuses, sample surveys, and administrative records—are routinely assessed, e.g., for coverage, sample error, response error, and nonsampling error; the results of the assessments are monitored and made available to guide statistical processes.

3.3 Statistical techniques

Data compilation employs sound statistical techniques to deal with data sources, other statistical procedures (e.g., data adjustments and transformations, and statistical analysis) employ sound statistical techniques.

3.4 Assessment and validation of intermediate data and statistical outputs

Intermediate results are validated against other information, where applicable, statistical discrepancies in intermediate data are assessed and investigated. Statistical discrepancies and other potential indicators of problems in statistical outputs are investigated.

3.5 Revision studies

Studies and analyses of revisions and/or updates are carried out and used internally to inform statistical processes (see also 4.3.3.)

4. Serviceability

4.1 Periodicity and timeliness

Periodicity and timeliness follows dissemination standards.

4.2 Consistency

Statistics are consistent within the dataset or reconcilable over a reasonable period of time and with those obtained through other data sources and/or statistical frameworks.

4.3 Revision policy and practice

Revisions and/or updates follow a regular and transparent schedule. Preliminary and/or revised/updated data are clearly identified. Studies and analyses of revisions are made public.

5. Accessibility

5.1 Data accessibility

Statistics are presented in a way that facilitates proper interpretation and meaningful comparisons (layout and clarity of text, tables, and charts).

Dissemination media and format are adequate.

Statistics are released on a preannounced schedule.

Statistics are made available to all users at the same time.

Statistics not routinely disseminated are made available upon request.

5.2 Metadata accessibility

Documentation on concepts, scope, classifications, basis of recording, data sources, and statistical techniques is available, and differences from internationally accepted standards, guidelines, or good practices are annotated.

Levels of detail are adapted to the needs of the intended audience.

5.3 Assistance to users

Contact points are publicized.

Publications, documents, and other services, including information on any charges, are widely available.

ESS Quality Reporting

<http://ec.europa.eu/eurostat/web/quality/quality-reporting>

you will find:

[ESS handbook for quality reports 2014](#)

[ESS Quality and Performance Indicators 2014](#)

[Single Integrated Metadata Structure and its Technical Manual](#)

[ESS Quality Glossary](#)

[Handbook on Data Quality - Assessment Methods and Tools](#)

[Handbook on improving quality by analysis of process variables](#)

