DANMARKS STATISTIK

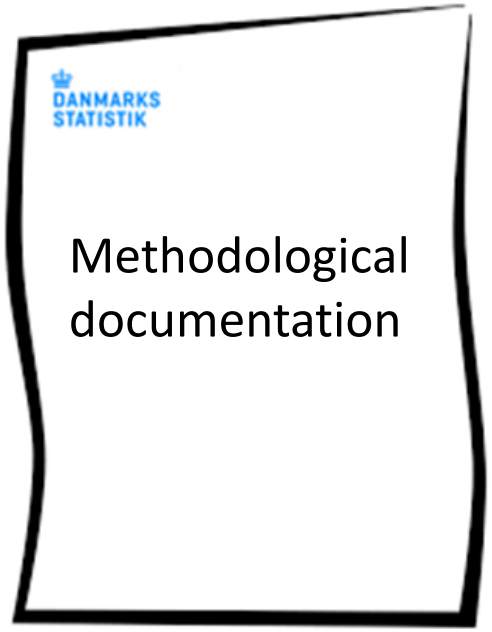
Work on methodological documentation

Activity 2.1.D (RS)

Friday June 4th 2021

Agenda

- Introduction to methodological documentation in Statistics Denmark
  - Why write a methodological documentation?
  - What is described in a methodological documentation
- Introduction to and implementation of Single Integrated Metadata Structure (SIMS)
- Guided tour of methodological documentation at www.dst.dk
- Practical demo of Statistics Denmark's metadata system Colectica



DANMARKS STATISTIK

Why write a methodological documentation

Quality Policy and Code of Practice (CoP)

DANMARKS STATISTIK

Statistics Denmark is a professionally independent institution that develops, produces and disseminates statistics about society. Statistics Denmark's policy is to work according to **sound methodology** and appropriate statistical procedures and with impartiality and objectivity while assuring statistical confidentiality. Our aim is to produce statistics that are accurate and reliable, coherent and comparable.

The statistics are defined according to users' needs and released in a timely and punctual manner. **We strive to present the statistics in a clear and understandable form.** In order to realize this, Statistics Denmark's professional independence is specified in law and it is ensured that staff has adequate training and experience to meet current statistical needs.

Statistics Denmark is responsible for official statistics in Denmark and needs to ensure that statistics are internally coherent and comparable; business and administrative sources are used when possible in order to avoid excessive burden on respondents.

Statistics Denmark puts emphasis on good service, efficiency and cost effectiveness to meet increasing domestic and international demands. Statistics Denmark takes part in international cooperation and is fully compliant with the demands of the European Statistical System.

Statistics Denmark works according to well-defined processes and according to plan. Quality indicators and other important factors regarding processes and **outputs are well-defined** and results are checked accordingly. If quality indicators are not satisfactory, changes will be carried out and improvements made on processes and procedures.

https://www.dst.dk/en/OmDS/strategi-og-kvalitet/kvalitetspolitik

The quality system of Statistics Denmark is based on the **15 principles** of the European Statistics Code of Practice (CoP) published by Eurostat:

Institutional environment

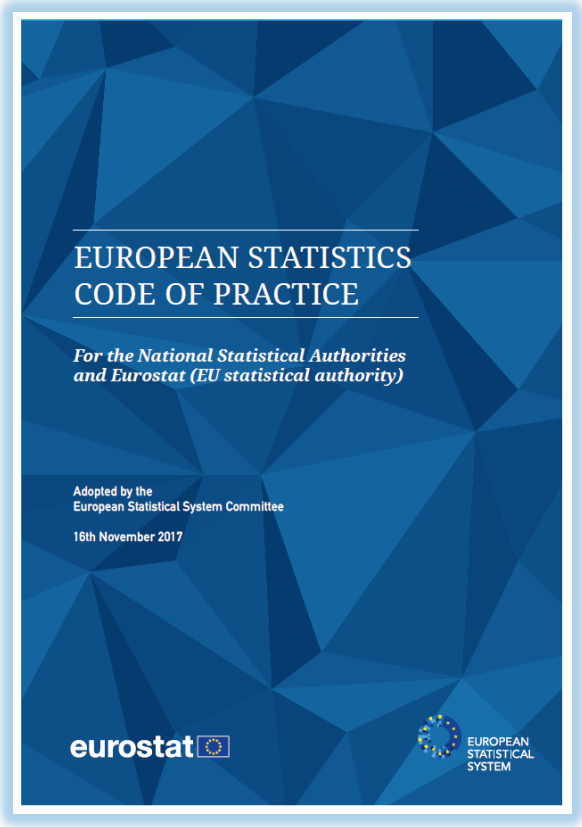
1	Professional Independence
1bis	Coordination and cooperation
2	Mandate for Data Collection
3	Adequacy of Resources
4	Commitment to Quality
5	Statistical Confidentiality

Statistical Processes

6	Impartiality and Objectivity
7	Sound Methodology
8	Appropriate Statistical Procedures
9	Non-excessive Burden on Respondents
10	Cost Effectiveness

Statistical Output

11	Relevance
12	Accuracy and Reliability
13	Timeliness and Punctuality
14	Coherence and Comparability
15	Accessibility and Clarity



Principle 15 Accessibility and Clarity

European Statistics are presented in a clear and understandable form, released in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance.

Indicator

15.1 Statistics and the corresponding metadata are presented, and archived, in a form that facilitates proper interpretation and meaningful comparisons.

15.2 Dissemination services use modern information and communication technology, methods, platforms and open data standards.

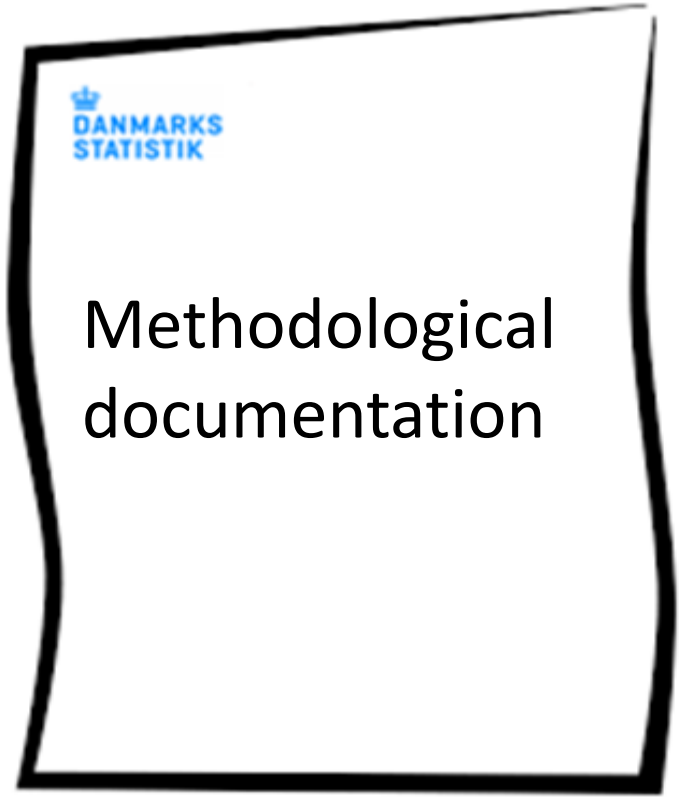
15.3 Custom-designed analyses are provided when feasible and the public is informed.

15.4 Access to microdata is allowed for research purposes and is subject to specific rules and protocols.

15.5 Metadata related to outputs are managed and disseminated by the statistical authority according to the European standards.

15.6 Users are kept informed about the methodology of statistical processes including the use and integration of administrative and other data.

15.7 Users are kept informed about the quality of statistical output with respect to the quality criteria for European Statistics.



ESS handbook for quality and metadata reports

DANMARKS STATISTIK

The general aim of the ESS Handbook for quality and metadata reports is to provide guidelines for the preparation of **producer and user reports for the full range of statistical processes and their outputs** within Member States.

- the term **statistical process** refers to a survey, administrative data process, or macro aggregate compilation conducted by a national statistical authority or by Eurostat
- the term **statistical output** refers to data that are disseminated together with the related services;
- the term metadata refers to descriptions of the statistical process and the concepts underlying its outputs and their quality, more precisely called **reference metadata**;

The ESS handbook for quality and metadata reports provides explicit guidelines for two types of report:

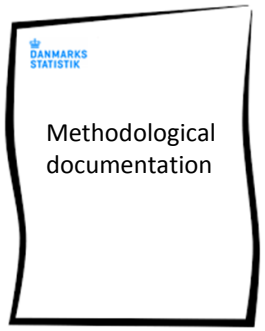
- a producer report, also (more precisely) called a **producer-oriented** report, and (less precisely) a quality report - comprising metadata, especially quality metadata, for use (1) within the NSA to record quality problems and improvements and (2) by Eurostat to review and summarise quality across NSAs;
- a user report, also (more precisely) called a **user-oriented** report, and (less precisely) a metadata report - comprising metadata, including quality metadata, that are intended for users of the statistical outputs, enabling them to ass whether the outputs are appropriate for the purposes they have in mind.

https://ec.europa.eu/eurostat/web/quality/quality-reporting

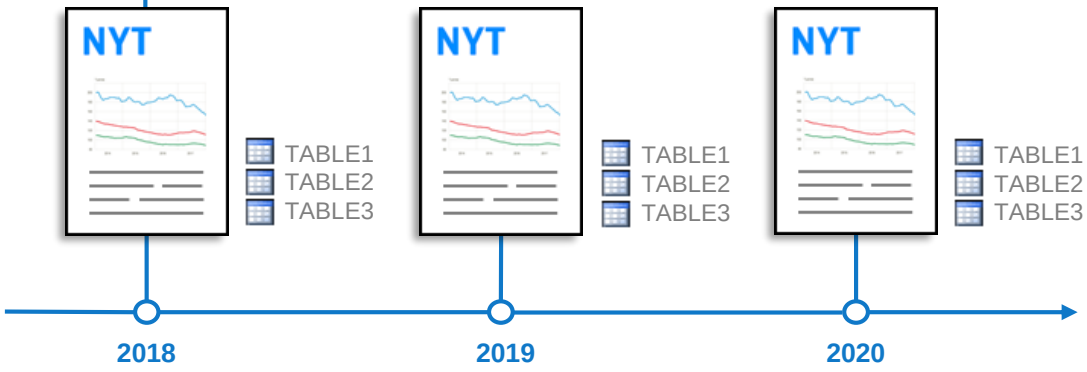
Statistical Products in DST

”The **Statistical Program** is an overview of all statistical products in Statistics Denmark, including a short description of the purpose and content of each product.

”**Statistical Products** are, generally, information dissemination products that are published or otherwise made available for public use that describe, estimate, forecast, or analyze the characteristics of groups, customarily without identifying the persons, organizations, or individual data observations that comprise such groups.”
Source: OECD



”The **Methodolocical documentation** is reference metadata that describes the content of the statistical product, including the tables in the StatBank and how the statistics are produced. In addition, the quality report describe how the statistics are compliant with the quality criteria in the ESS Code of Practice, i.e.; *Relevance*, *Accuracy an Reliability, Timeliness and Punctuality, Coherence and Comparability*, and *Accessibility and Clarity*. The quality report is structured according to the international standard for quality reporting Single Integrated Metadata Structure (SIMS).”



”A **Statistical Release** occurs every time new, preliminary or revised figures for a Statistical Product are made available to the public. A statistical release shows – in numbers – something about a phenomenon in a given reference period, typically a month, quarter or year. A statistical release from Statistics Denmark consist of; new or updated figures in one or more tables in the StatBank, an updated quality report and possibly a press release. A statistical release must have a pre-set publication date in the release calendar.

DANMARKS STATISTIK

What is described in a methodological documentation

Two types of statistical metadata

Structural metadata

- Used to identify statistical data
- Headlines, variable names, unit of measure, reference time mv.
- Must go together with statistical data
- Impossible to interpret statistics without it

Reference metadata

- Describes content, statistical processing, relevance etc.
- Can be detached from the statistical output
- Quality Reports is a type of reference metadata
- …so is methodological metadata

Statistics without metadata

	2 881 620
	2 908 337

	2 868 172
	2 976 785

…with structural metadata

Population

All Denmark	2018Q3
Men	2 881 620
Women	2 908 337

Unit : number

Real estate market value

One-family houses	2016
Brøndby	2 868 172
Vallensbæk	2 976 785

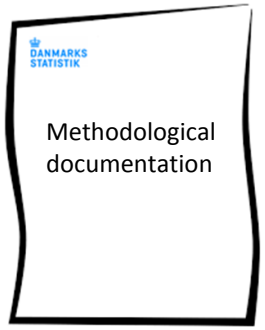
Unit : Average Market value (DKK)

…and reference metadata

Population

All Denmark	
Men	
Women	

Unit : number

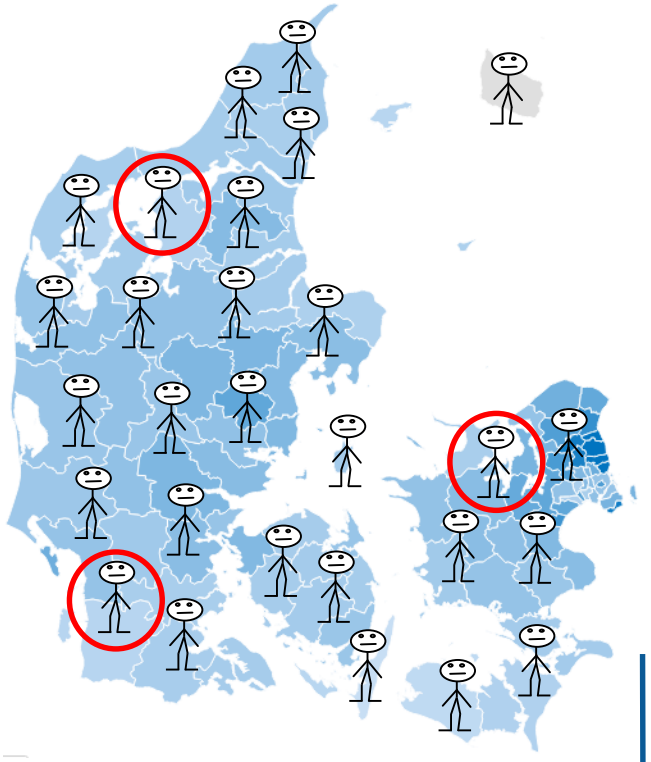


Real estate market

One-family houses	2016
Brøndby	2 868 172
Vallensbæk	2 976 785

Unit : Average Market value (DKK)

The reality as it really is, objective and true



Conceptualized reality (Frame population)

	Quest 1	Quest 2

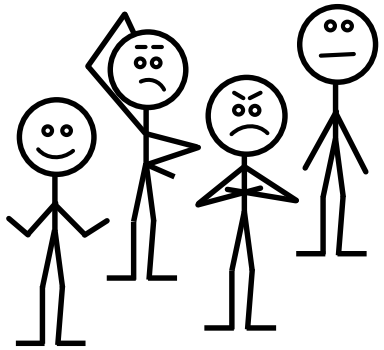
Units that cant be sampled or units outside the sample/population

The reality as it is represented by data

	Quest 1	Quest 2
	Yes	No
	Yes	Yes
	No	Yes
	Yes	Yes
	No idea	No clue
	..	Well..
	NO!	NO!
	..	..

What we ask for, relative to the data we actually get from respondents

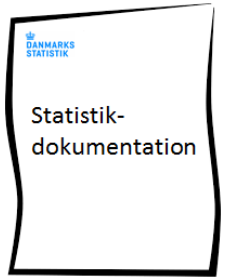
Statistics about reality as interpreted by users



Different frames of reference in interpreting and understanding data

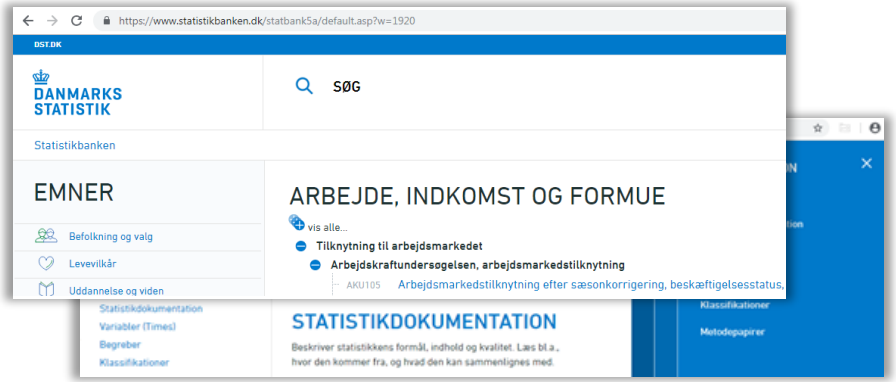
Metadata in Statistics Denmark

DANMARKS STATISTIK



Methodological documentation

StatBank and www.dst.dk



Registers and variables

ID	VAR_1	TYPE	AAR_FRA	AAR_TIL	...	LAND
1	25564	5	2008	2013	...	Danmark
2	72639	..	2004	2016	...	..
3	..	4	2010	2010	...	Ghana
4	34621	5	0001	9999	...	Tyskland
5	52485	4	2010	9999	...	Libyen
6	28815	4	2003	2004	...	Belgien
7	60655	..	2007	9999	...	Danmark
8	74111	5	2009	2009	...	Angola
9	59992	2	2001	2013	...	Polen
10	..	3	1998	2009	...	Schweiz
...	.	.	.	.	.	.
n	44589	3	2009	2009	...	Danmark

colectica

Classifications

- W1: LANDE I ALT
  - E1: EUROPA
  - F1: AFRIKA
    - F4: Nordafrika
      - DZ: Algeriet
      - EG: Egypten
      - LY: Libyen
      - MA: Marokko
      - TN: Tunesien
    - F2: Andre afrikanske lande

Concepts

Personbil: *Bil, der er indrettet til at benyttes til befordring af højst ni personer føreren medregnet. Normalt har en personbil hvide nummerplader, men kan også være for-synet med gule nummerplader. En personbil kan føres af alle med et gyldigt B kørekort.*

Varebil: *Bil, der er indrettet til at benyttes til godsbefordring, og som har en tilladt total-vægt på ikke over 3.500 kg. Er forsynet med gule eller delvist gule (papegøjeplader) nummerplader. En varebil kan køres af alle med et almindeligt kørekort til kategori B.*

Coherent metadata in Colectica

DANMARKS STATISTIK

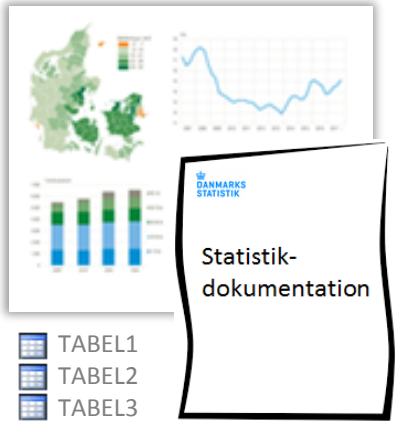
Mikro data

ID	VAR_1	TYPE	AAR_FRA	AAR_TIL	...	LAND
1	25564	5	2008	2013	...	Danmark
2	72639	..	2004	2016	...	..
3	..	4	2010	2010	...	Ghana
4	34621	5	0001	9999	...	Tyskland
5	52485	4	2010	9999	...	Libyen
6	28815	4	2003	2004	...	Belgien
7	60655	..	2007	9999	...	Danmark
8	74111	5	2009	2009	...	Angola
9	59992	2	2001	2013	...	Polen
10	..	3	1998	2009	...	Schweiz
...	.	.	.	.	.	.
n	44589	3	2009	2009	...	Danmark

Classifications

- W1: LANDE I ALT
  - E1: EUROPA
  - F1: AFRIKA
    - F4: Nordafrika
      - DZ: Algeriet
      - EG: Egypten
      - LY: Libyen
      - MA: Marokko
      - TN: Tunesien
    - F2: Andre afrikanske lande

Statistical products



Variables

Kan antage et sæt værdier. Nogle variables udfaldsrum specificeres i tilhørende kodelister, fx TYPE

Variabel	Beskrivelse
VAR_1	...
TYPE	Køretøjtype
AAR_FRA	..
AAR_TIL	...
...	...
LAND	Land

Code lists

TYPE	TYPE_TXT
1	Personbil
2	Bus
3	Varebil
4	Lastbil
5	Påhængsvogn
6	Mototcykel
7	Knallert 45
8	Traktor
9	Campingvogn
10	Øvrige

Concepts

Personbil: *Bil, der er indrettet til at benyttes til befordring af højst ni personer føreren medregnet. Normalt har en personbil hvide nummerplader, men kan også være forsynet med gule nummerplader. En personbil kan føres af alle med et gyldigt B kørekort.*

Varebil: *Bil, der er indrettet til at benyttes til godsbefordring, og som har en tilladt totalvægt på ikke over 3.500 kg. Er forsynet med gule eller delvist gule (papegøjeplader) nummerplader. En varebil kan køres af alle med et almindeligt kørekort til kategori B.*

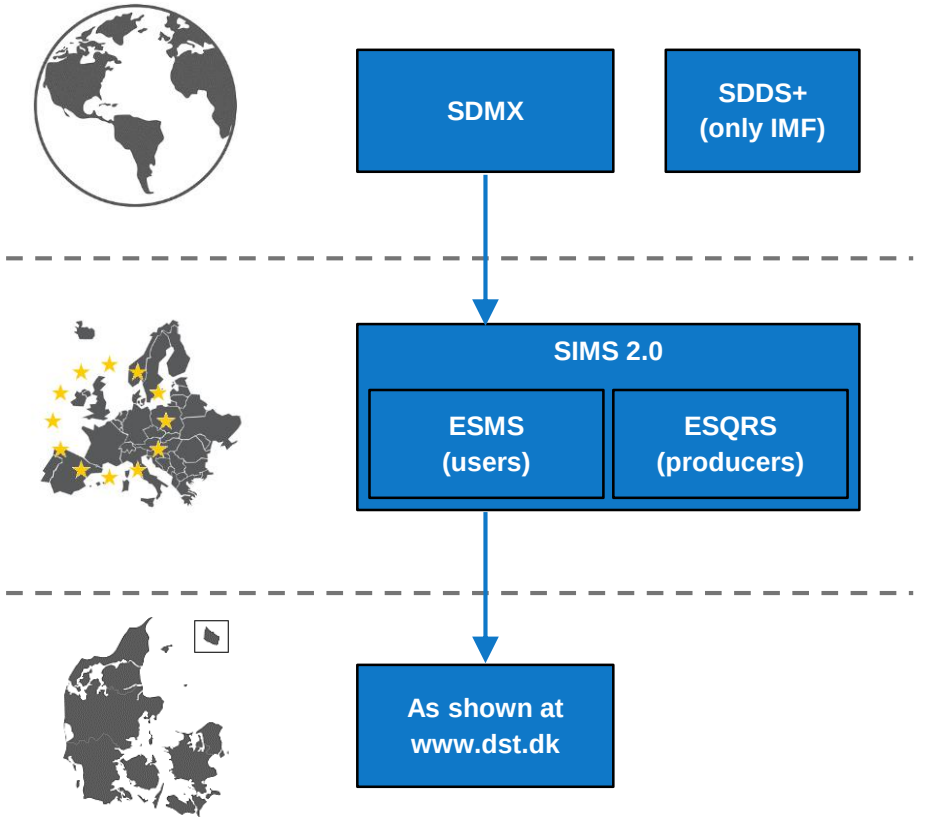
DANMARKS STATISTIK

Implementation of Single Integrated Metadata Structure (SIMS)

International standards for quality reporting

- ✓ In Statistics Denmark we use **SIMS**
- ✓ We have translated SIMS to Danish
- ✓ SIMS is denoted in parenthesis, e.g. **4. Relevance (s12)**
- ✓ We ‘speak the same language’ as Eurostat and other countries
- ✓ English version is uploaded directly in ESS Metadata Handler

SDMX:	Statistical Data and Metadata eXchange
SDDS+:	Special Data Dissemination Standard
SIMS:	Single Integrated Metadata Structure
ESMS:	EURO-SDMX Metadata Structure
ESQRS:	ESS Quality Reporting Structure



ESMS - Euro-SDMX Metadata Structure

DANMARKS STATISTIK

- User-oriented documentation because it contains a “basic level” of information
- However, the main purpose of the ESMS is not to report on data quality but to document the production and output of European statistics for data users.
- Therefore, the ESMS also includes other statistical concepts that are not directly related to quality

Item No	Concept Name
1	**Contact**
1.1	Contact organisation
1.2	Contact organisation unit
1.3	Contact name
1.4	Contact person function
1.5	Contact mail address
1.6	Contact email address
1.7	Contact phone number
1.8	Contact fax number
2	**Metadata update**
2.1	Metadata last certified
2.2	Metadata last posted
2.3	Metadata last update
3	**Statistical presentation**
3.1	Data description
3.2	Classification system
3.3	Sector coverage
3.4	Statistical concepts and definitions
3.5	Statistical unit
3.6	Statistical population
3.7	Reference area
3.8	Time coverage
3.9	Base period
4	**Unit of measure**
5	**Reference period**
6	**Institutional mandate**
6.1	Legal acts and other agreements
6.2	Data sharing

Item No	Concept Name
7	**Confidentiality**
7.1	Confidentiality - policy
7.2	Confidentiality - data treatment
8	**Release policy**
8.1	Release calendar
8.2	Release calendar access
8.3	User access
9	**Frequency of dissemination**
10	**Accessibility and clarity**
10.1	News release
10.2	Publications
10.3	On-line database
10.4	Micro-data access
10.5	Other
10.6	Documentation on methodology
10.7	Quality documentation
11	**Quality management**
11.1	Quality assurance
11.2	Quality assessment
12	**Relevance**
12.1	User needs
12.2	User satisfaction
12.3	Completeness
13	**Accuracy and reliability**
13.1	Overall accuracy
13.2	Sampling error
13.3	Non-sampling error

Item No	Concept Name
14	**Timeliness and punctuality**
14.1	Timeliness
14.2	Punctuality
15	**Coherence and comparability**
15.1	Comparability - geographical
15.2	Comparability - over time
15.3	Coherence - cross domain
15.4	Coherence - internal
16	**Cost and burden**
17	**Data revision**
17.1	Data revision - policy
17.2	Data revision - practice
18	**Statistical processing**
18.1	Source data
18.2	Frequency of data collection
18.3	Data collection
18.4	Data validation
18.5	Data compilation
18.6	Adjustment
19	**Comment**

ESQRS - ESS Standard for Quality Reports

DANMARKS STATISTIK

- Producer-oriented documentation because it contains a more detailed level of information
- Focuses more on quality performance indicators (QPI) in statistical process
- SDMX-compliant and technically implemented as metadata structure definition

Item no	Concept Name
1	**Contact**
1.1	Contact organisation
1.2	Contact organisation unit
1.3	Contact name
1.4	Contact person function
1.5	Contact mail address
1.6	Contact email address
1.7	Contact phone number
1.8	Contact fax number
2	**Statistical presentation**
2.1	Data description
2.2	Classification system
2.3	Sector coverage
2.4	Statistical concepts and definitions
2.5	Statistical unit
2.6	Statistical population
2.7	Reference area
2.8	Time coverage
2.9	Base period
3	**Statistical processing**
3.1	Source data
3.2	Frequency of data collection
3.3	Data collection
3.4	Data validation
3.5	Data compilation
3.6	Adjustment
4	**Quality management**
4.1	Quality assurance
4.2	Quality assessment
5	**Relevance**
5.1	User Needs
5.2	User Satisfaction
5.3	Completeness
5.3.1	Data completeness - rate

Item no	Concept Name
6	**Accuracy and reliability**
6.1	Accuracy - overall
6.2	Sampling error
6.2.1	Sampling error - indicators
6.3	Non-sampling error
6.3.1	Coverage error
6.3.1.1	Over-coverage - rate
6.3.1.2	Common units - proportion
6.3.2	Measurement error
6.3.3	Non response error
6.3.3.1	Unit non-response - rate
6.3.3.2	Item non-response - rate
6.3.4	Processing error
6.3.4.1	Imputation - rate
6.3.5	Model assumption error
6.4	Seasonal adjustment
6.5	Data revision - policy
6.6	Data revision - practice
6.6.1	Data revision - average size
7	**Timeliness and punctuality**
7.1	Timeliness
7.1.1	Time lag - first result
7.1.2	Time lag - final result
7.2	Punctuality
7.2.1	Punctuality - delivery and publication
8	**Coherence and comparability**
8.1	Comparability - geographical
8.1.1	Asymmetry for mirror flow statistics - coefficient
8.2	Comparability - over time
8.2.1	Length of comparable time series
8.3	Coherence - cross domain
8.4	Coherence - sub annual and annual statistics
8.5	Coherence - National Accounts
8.6	Coherence - internal

Item no	Concept Name
9	**Accessibility and clarity**
9.1	News release
9.2	Publications
9.3	Online database
9.3.1	Data tables - consultations
9.4	Microdata access
9.5	Other
9.6	Documentation on methodology
9.7	Quality documentation
9.7.1	Metadata completeness - rate
9.7.2	Metadata - consultations
10	**Cost and Burden**
11	**Confidentiality**
11.1	Confidentiality - policy
11.2	Confidentiality - data treatment
12	**Comment**

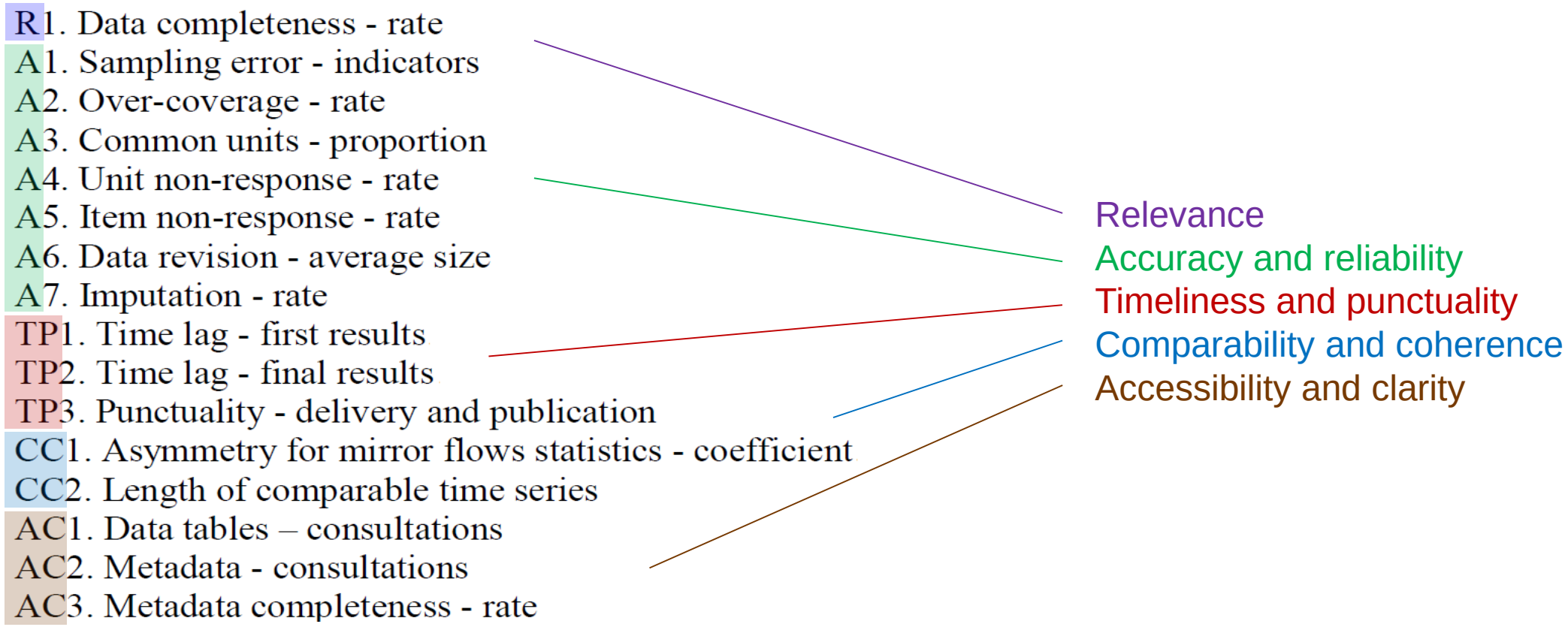
SIMS – Single Integrated Metadata Structure

DANMARKS STATISTIK

EURO-SDMX Metadata Structure		Single Integrated Metadata Structure		ESS Standard for Quality Reports Structure	
5	Reference period	S.5	Reference period		
6	Institutional mandate	S.6	Institutional mandate		
6.1	Legal acts and other agreements	S.6.1	Legal acts and other agreements		
6.2	Data sharing	S.6.2	Data sharing		
7	Confidentiality	S.7	Confidentiality	11	Confidentiality
7.1	Confidentiality - policy	S.7.1	Confidentiality - policy	11.1	Confidentiality – policy
7.2	Confidentiality - data treatment	S.7.2	Confidentiality - data treatment	11.2	Confidentiality – data treatment
8	Release policy	S.8	Release policy		
8.1	Release calendar	S.8.1	Release calendar		
8.2	Release calendar access	S.8.2	Release calendar access		
8.3	User access	S.8.3	User access		
9	Frequency of dissemination	S.9	Frequency of dissemination		
10	Accessibility and clarity	S.10	Accessibility and clarity	9	Accessibility and clarity
10.1	News release	S.10.1	News release	9.1	News release
10.2	Publications	S.10.2	Publications	9.2	Publication
10.3	On-line database	S.10.3	On-line database	9.3	On-line database
		S.10.3.1	*AC1. Data tables - consultations*	9.3.1	Data tables - consultations
10.4	Micro-data access	S.10.4	Micro-data access	9.4	Micro-data access
10.5	Other	S.10.5	Other	9.5	Other
		S.10.5.1	*AC 2. Metadata - consultations*	9.7.2	Metadata - consultations
10.6	Documentation on methodology	S.10.6	Documentation on methodology	9.6	Documentation on methodology
		S.10.6.1	*AC 3. Metadata completeness - rate*	9.7.1	Metadata completeness – rate
10.7	Quality documentation	S.10.7	Quality documentation	9.7	Quality documentation
11	Quality management	S.11	Quality management	4	Quality management
11.1	Quality assurance	S.11.1	Quality assurance	4.1	Quality assurance
11.2	Quality assessment	S.11.2	Quality assessment	4.2	Quality assessment
12	Relevance	S.12	Relevance	5	Relevance
12.1	User needs	S.12.1	User needs	5.1	User needs
12.2	User satisfaction	S.12.2	User satisfaction	5.2	User satisfaction
12.3	Completeness	S.12.3	Completeness and *R1. Data completeness - rate for U*	5.3	Completeness
		S.12.3.1	*R1. Data completeness - rate for P*	5.3.1	Data completeness - rate

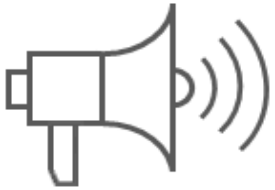
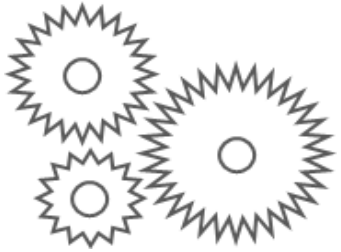
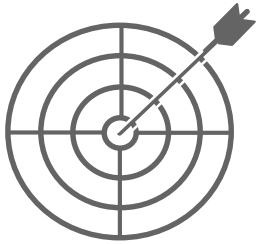
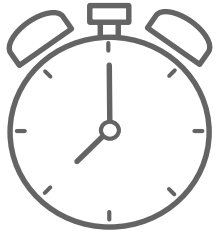
Quality and performance indicators (QPI)

- As a general rule, it is recommended that both, producer- and user-oriented quality reports contain as many Quality and Performance Indicators (QPIs) of the standard ESS list as possible.
- The list and description of the 16 standard ESS Quality Performance Indicators is published on the website of Eurostat



Methodological documentation content

DANMARKS STATISTIK

Introduction 	Statistical presentation 	Statistical processing 	Relevance 
Accuracy and reliability 	Timeliness and punctuality 	Comparability 	Accessibility and clarity 

Methodological documentation and CoP

DANMARKS STATISTIK

Introduction	Statistical presentation	Statistical processing	Relevance PRINCIPLE 11 Relevance
Accuracy and reliability PRINCIPLE 12 Accuracy and Reliability	Timeliness and punctuality PRINCIPLE 13 Timeliness and Punctuality	Comparability PRINCIPLE 14 Coherence and Comparability	Accessibility and clarity PRINCIPLE 15 Accessibility and Clarity

DANMARKS STATISTIK

Guided tour of methodological documentation in Statistics Denmark

DANMARKS STATISTIK

Practical demo of Statistics Denmark's metadata system Colectica

Thank you for your attention

DANMARKS STATISTIK