

Metadata at Statistics Denmark

Lars Thygesen lthygesen@gmail.com

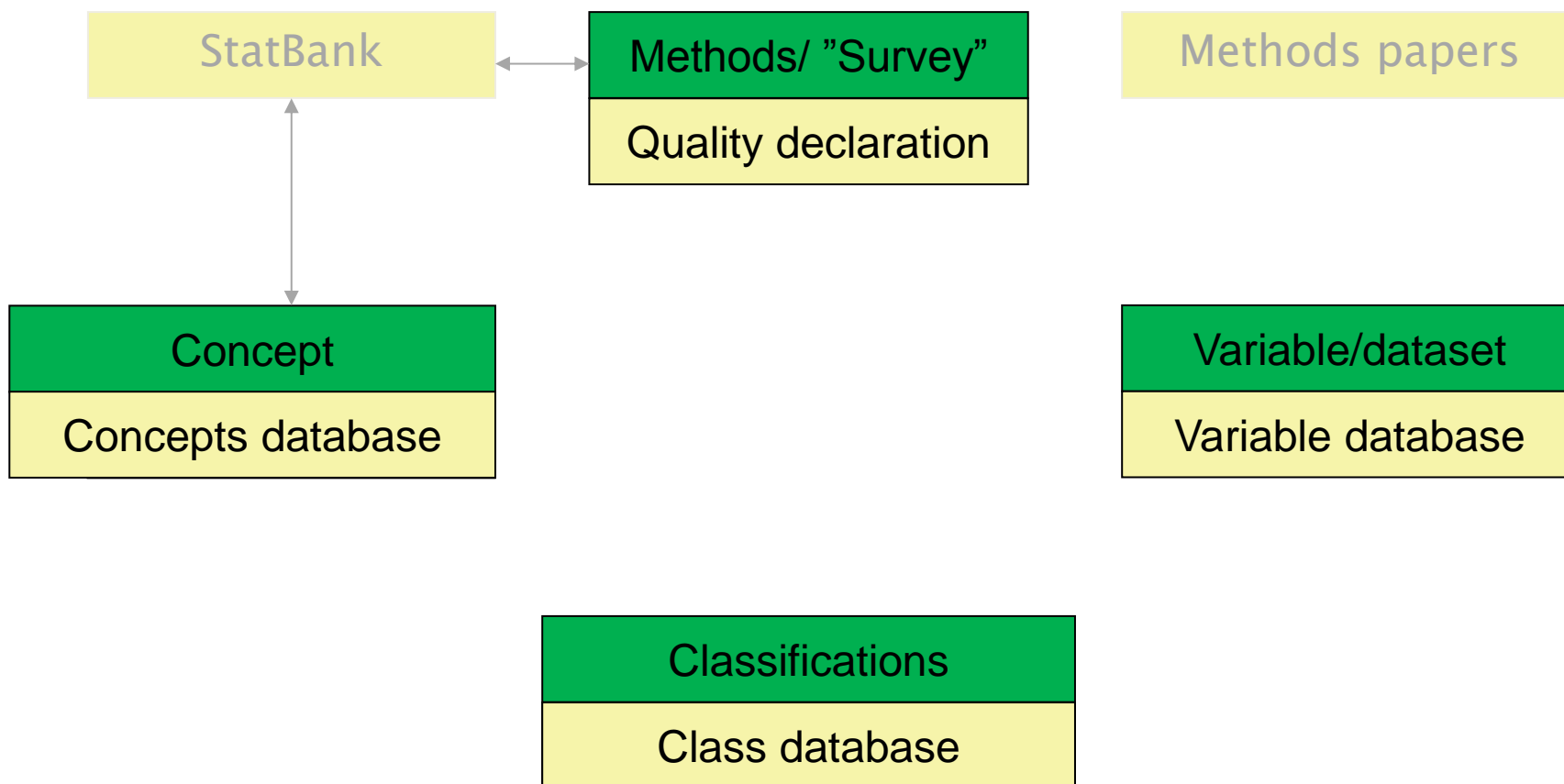
Agenda

1. Metadata Challenges
2. Vision
3. User evaluations and Focus-groups on Metadata
4. Project on implementing quality declarations
5. Implementation using Colectica

Aren't our metadata good enough?

- Dedicated work since 1995
- Followed best practice from UN
- Main elements in place – but...

Metadata elements



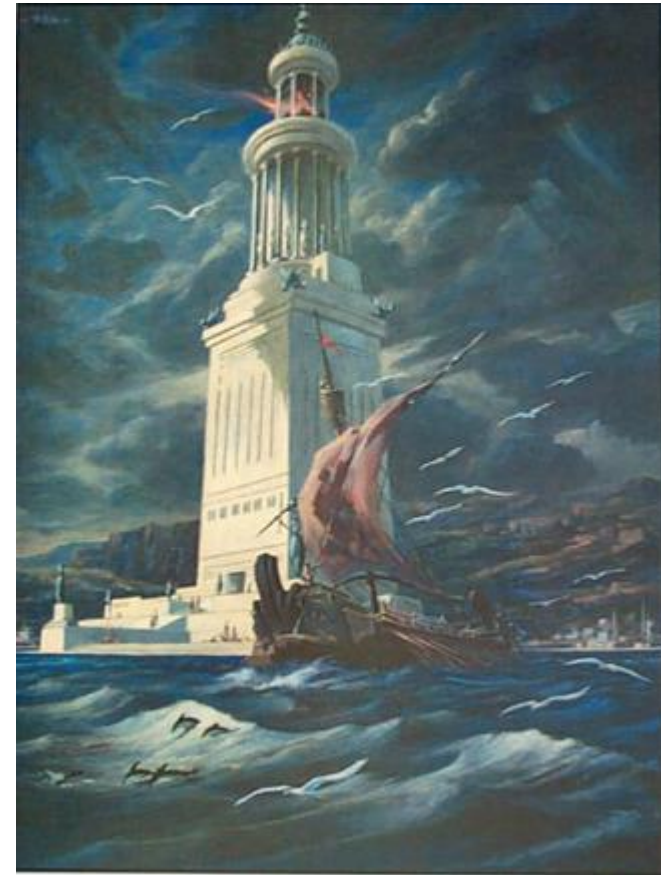
However...

- Consistent claims of users:

Metadata / documentation isn't good enough

Vision and focus of Statistics Denmark

1. Statistical information must guide users in the “turbulent information-sea”
2. Metadata about content and quality must
 - help users in their knowledge processes
 - Help users find the right statistics
 - give users precise information about the products
3. International standards and standard software must enable:
 - Cost efficient solution
 - Gradual implementation with few resources
 - Sustainable long term solution



Challenges on metadata

Standards:

- Introduction of GSBPM in order to reuse common international terms
- SIMS, ESQRS and ESMS from Eurostat

• **External needs (from focus-groups):**

- Difficult to understand content of quality declarations
- Comparability across domains is needed
- Expert knowledge too sectorial
- Better information on data breaks and revisions

Internal

- Fragmented and non-standardised work-processes
- Metadata linked only to final data and no reuse
- Presentation of metadata on Internet is fragmented and incomplete
- Concepts database incomplete
- Classifications and code-lists in many places

User evaluations

- User Committees
- Web satisfaction survey
- Customer satisfaction survey
- Focus-Groups (on metadata)
- Bilateral contacts (on metadata)

FOCUS-GROUPS ON METADATA

Needs for metadata using focus-groups

- 3 groups each with 10-12 handpicked users
- Groups:
 - Intensive users, mostly government
 - Municipal and regional users
 - Education and the media.

Agenda for focus-group meetings

How do you use statistics and what are your needs?

"Demand"

Presentation of model / solution on integrated metadata and search types

"Supply"

How is the balance between needs and solution

"Match"

Main results

- The integrated model (quality declarations, concepts, variables and classifications) won strong support
 - Good to have a more logical approach to documentation
- Difficult to understand content of quality declarations
- Comparability across domains is needed
- Expert knowledge too sectorial
- Better information on data breaks and revisions

PROJECT ON QUALITY DECLARATIONS

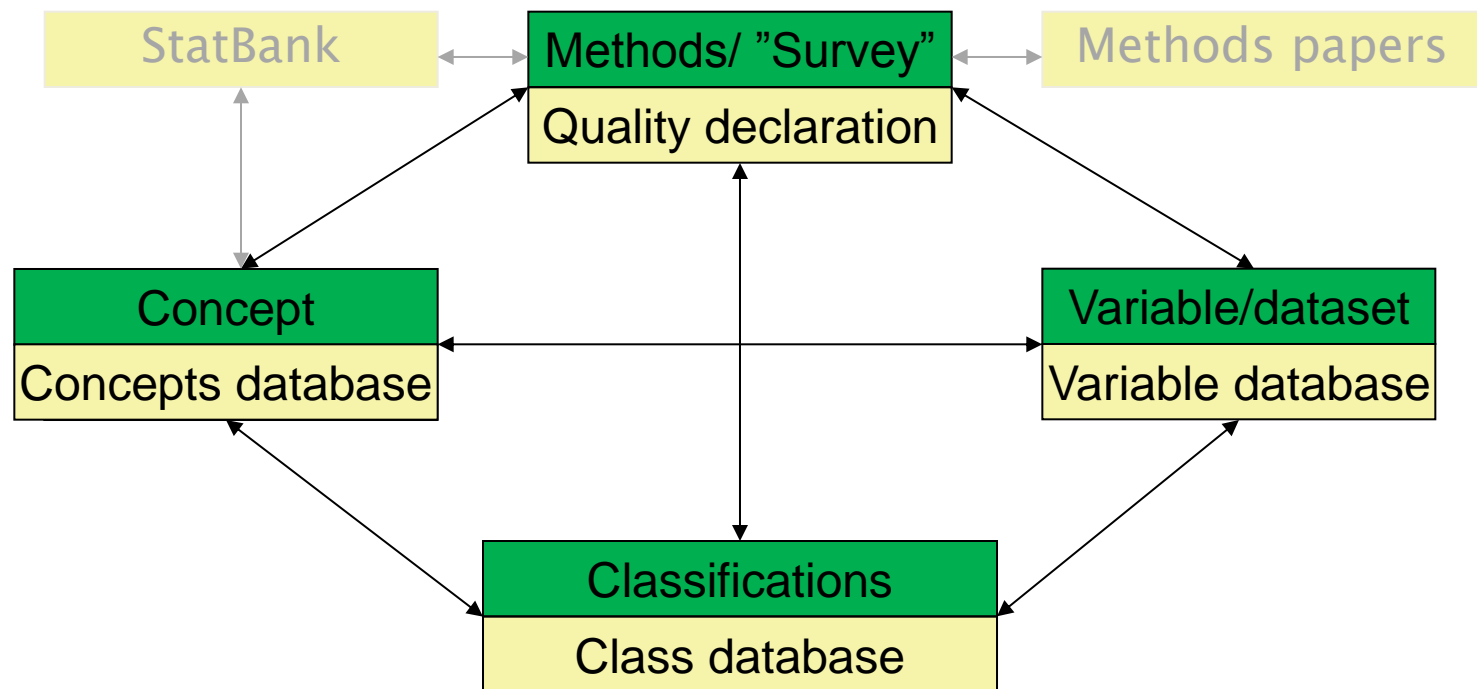
Title of EU Project

"Horizontal and vertical integration: Implementation of technical and statistical standards in the European Statistical System"

Deliverables

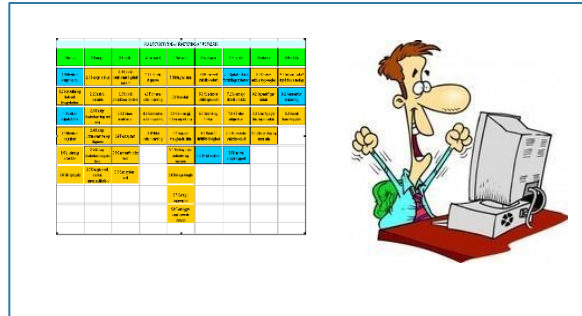
- Quality declarations migrated to new metadata system for all statistics
- Meetings and reports from external users
- Dissemination on dst.dk developed in collaboration with users
- Software (Colectica) installed - internal interface and external interface (dissemination at dst.dk)
- Manuals and courses
- Communication

”Classical” metadata elements implemented using Data Documentation initiative (DDI) and SDMX

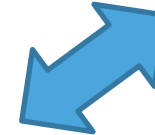
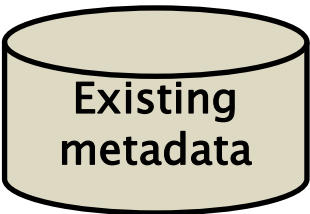




Enter Q.D.

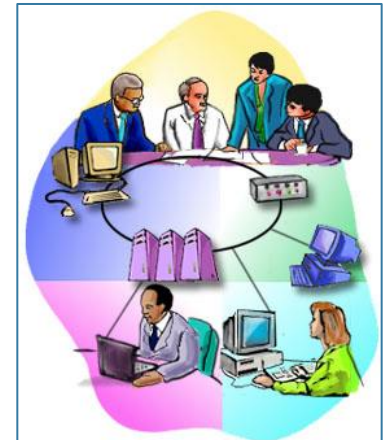


Publish at Dst.dk



Publish at the Intranet

Q.D. to EU



Quality declarations

EU standard *Single Integrated Metadata Structure SIMS*:

Statistical content

Statistical processing

Quality dimensions

Relevance

Accuracy and reliability

Timeliness and punctuality

Comparability

Accessibility and clarity

- Basis for quality assurance of statistical products
- Reporting to Eurostat (ESMS & ESQRS), IMF (SDDS)
- Customized external and internal publishing

Single Integrated Metadata Structure (SIMS) and reporting formats: ESMS and ESQRS -

EURO-SDMX Metadata Structure (Dec 2010)

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

13	Relevance
13.1	User needs
13.2	User satisfaction
13.3	Completeness

Single Integrated Metadata Structure

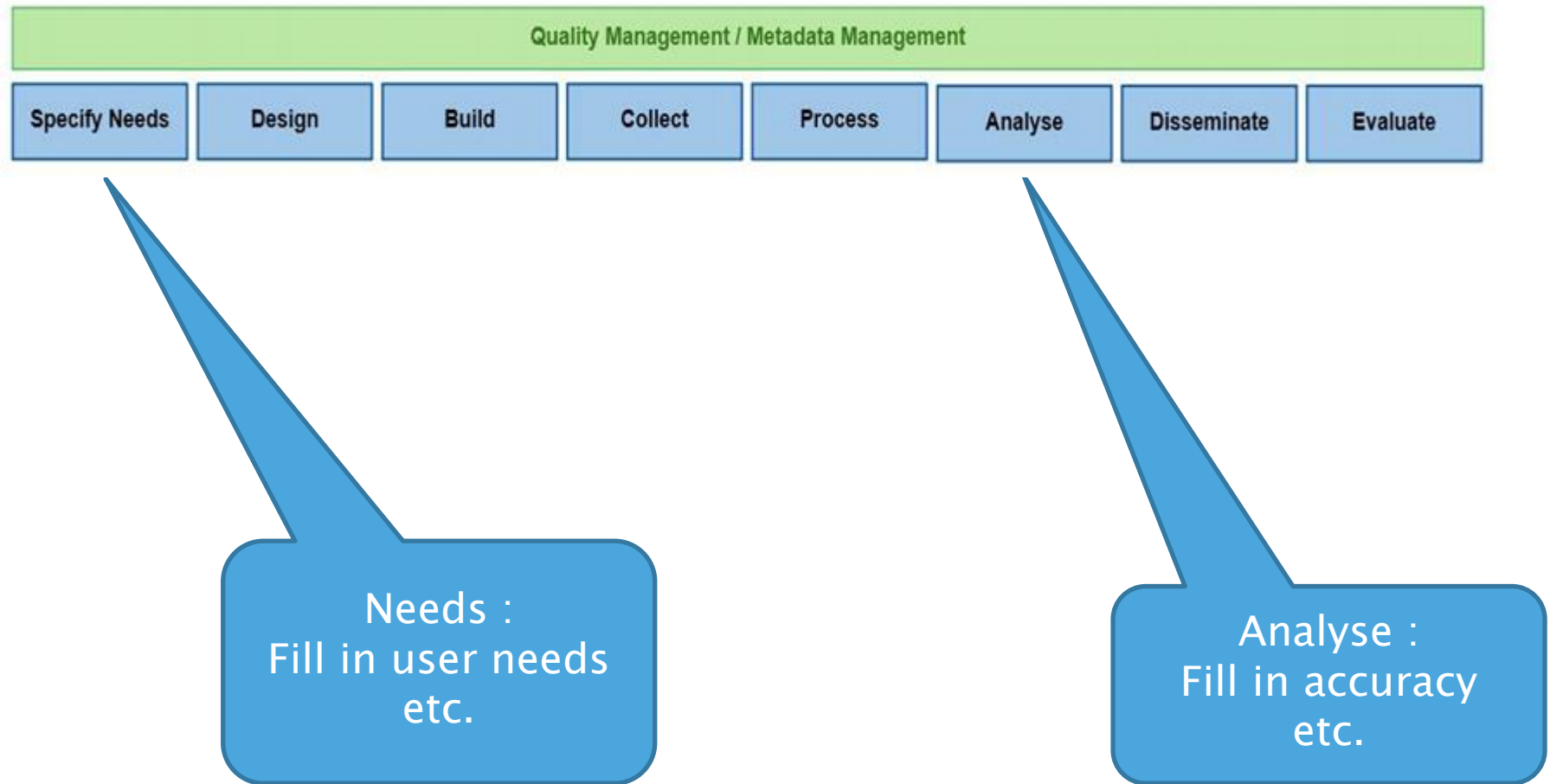
S.4	Statistical presentation
S.4.1	Data description
S.4.2	Classification system
S.4.3	Sector coverage
S.4.4	Statistical concepts and definitions
S.4.5	Statistical unit
S.4.6	Statistical population
S.4.7	Reference area
S.4.8	Time coverage
S.4.9	Base period

S.14	Relevance
S.14.1	User needs
S.14.2	User satisfaction
S.14.3	Completeness and <i>R1. Data completeness - rate for U</i>
S.14.3.1	<i>R1. Data completeness - rate for P</i>

ESS Standard for Quality Reports Structure

IV	Relevance
IV.1	User needs
IV.2	User satisfaction
IV.3	Completeness
IV.3.1	Data completeness - rate

Work processes and quality declarations



Work processes and quality declarations in Colectica – Retail Trade index

The screenshot displays the Colectica Designer application window. The title bar reads "Colectica Designer". The menu bar includes "Filer", "Colectica", "Import", "Explore", "Help", and "GSBPM". A breadcrumb trail shows "Quality Statement". The main content area is titled "Quality Statement" and contains the following information:

- Quality Statements
- Quality Standard
- External Materials
- Extended
- Identification

Below these tabs, a row of buttons represents the work process steps: "General", "1. Specify Needs", "2. Design", "5. Process", "6. Analyse", "7. Disseminate", "8. Archive", "9. Evaluate", and "All". The "1. Specify Needs" button is currently selected.

The "User Needs" section is expanded, showing a list of topics on the left and a detailed description on the right.

User Needs

Description of users and their respective needs with respect to the statistical data.

Many users who monitor the current business trends share a great interest for the published statistics of retail trade. The statistics is in demand from trade associations, bank and finance sector, politicians, public and private institutions, researchers, enterprises, news media and Eurostat.

It is also of great interest to many enterprises within the retail trade sector, as publications allow them to compare their own sales to those of their trade industry.

Furthermore, the statistics provide important input to the national accounts statistics, published quarterly by Statistics Denmark.

Some users have a need for more commodity-based figures. The detailed industry figures are based solely on industry activity, thus enterprises with a broad commodity selection, in reality covering more industries, causes problems

en...

21

Work processes and quality declarations in Colectica – Retail Trade index

The screenshot displays the Colectica Designer application window. The title bar reads "Colectica Designer". The menu bar includes "Filer", "Colectica", "Import", "Explore", "Help", and "GSBPM". The main workspace is titled "Quality Statement" and shows a document icon. Below this, a header bar indicates the project: "Retail trade index January 2014", "Agency dk.dst", "Revision 64", and "Last Updated 2014-02-03 11:04". The interface features a tabbed view with "Quality Statements" selected. A row of buttons represents the workflow: "General", "1. Specify Needs" (highlighted), "2. Design", "5. Process", "6. Analyse", "7. Disseminate", "8. Archive", "9. Evaluate", and "All". On the left, a list of sections is shown, with "Introduction" selected. The right pane displays the content of the "Introduction" section, which includes a general description of the statistical process and its outputs, and a paragraph explaining the purpose of the Retail Trade Index.

Colectica Designer

Filer Colectica Import Explore Help GSBPM

Quality Statement

Quality Statement

Retail trade index January 2014 Agency dk.dst Revision 64 Last Updated 2014-02-03 11:04

Quality Statements Quality Standard External Materials Extended Identification

General 1. Specify Needs 2. Design 5. Process 6. Analyse 7. Disseminate 8. Archive 9. Evaluate All

Statistical presentation
User Needs
Introduction
Data description
Institutional mandate
Legal acts and other agreements
Confidentiality
Confidentiality - policy

Introduction
A general description of the statistical process and its outputs, and their evolution over time.
The purpose of the Retail Trade Index is to analyze and explain the development in turnover within the different sectors of retail trade, which constitutes an important part of private consumption in Denmark. The statistics is primarily a short term indicator and is being used for assessing the fluctuations of the market.

en...

Editor DDI

Quality declaration: Retail ... x Skriv en omtale af Candela... x Lars's Profil - Airbnb x +

dst.dk/en/Statistik/dokumentation/declarations/retail-trade-index.aspx

Mest besøgte Ferieboliger, private v... Log på Netbank International Consulta...

STATISTICS DENMARK

CONTACT PRESS INFORMATION SERVICES DANSK

Search our website Search

FIND STATISTICS PRODUCTS & SERVICES CONSULTING ABROAD ABOUT US

SUBJECT PAGES STATBANK PUBLICATIONS SCHEDULED RELEASES DOCUMENTATION IMF: SDDS

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
Søren Kuhl Andersen
+45 3917 3561
ska@dst.dk
Hent som PDF
[Retail Trade Index](#)
[» Show archive](#)

PC-Axis
You can download the PC-AXIS software free of charge
[Read about PC-AXIS](#)

Quality Declarations
Retail Trade Index
Accessibility and clarity
Statistical presentation
Statistical processing
Relevance
Accuracy and reliability
Timeliness and punctuality
Comparability

Classifications
Methodology
Major revision of national accounts

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.

Transmission to Eurostat and customized presentation in reports and at dst.dk

- Many views into quality declarations to support various users
- Example: Extract from report with summary information for Retail Trade Index

Introduction

The purpose of the Retail Trade Index is to analyze and explain the development in turnover within the different sectors of retail trade, which constitutes an important part of private consumption in Denmark. The statistics is primarily a short term indicator and is being used for assessing the fluctuations of the market.

Data description

The retail trade index shows the development of turnover from sales of commodities to private persons in Danish retail enterprises. The retail trade index is published as value and volume indices. The value indices show the development of the turnover in current prices. Value indices are published on 43 industries within the retail trade sector and for three main commodity groups: ...

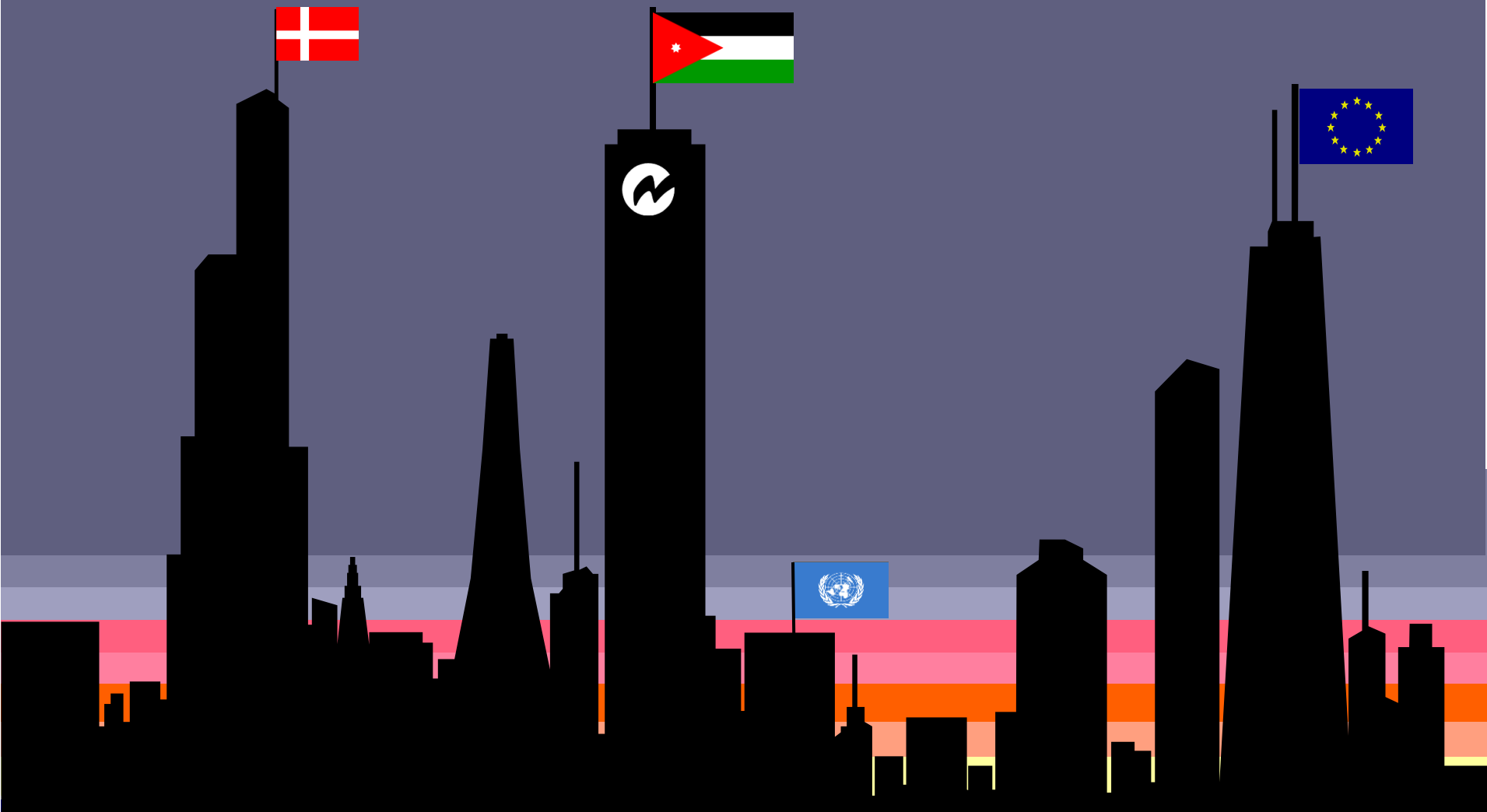
Relevance

ESS Guidelines for the Implementation of the ESS Quality and Performance Indicators (QPI)

- R1. Data completeness - rate
- A1. Sampling error - indicators
- A2. Over-coverage - rate
- A3. Common units - proportion
- A4. Unit non-response - rate
- A5. Item non-response – rate
- A6. Data revision - average size
- A7. Imputation - rate
- TP1. Time lag - first results
- TP2. Time lag - final results
- TP3. Punctuality - delivery and publication
- CC1. Asymmetry for mirror flows statistics – coefficient
- CC2. Length of comparable time series
- AC1. Data tables – consultations
- AC2. Metadata - consultations
- AC3. Metadata completeness - rate

Examples of guidelines for statistical methods

- There are written guidelines for:
 - Sampling, editing/ imputation and seasonal adjustment
- The organization of the individual statistics is based on a common basis, and this common ground is available for all at Statistics Denmark
- Guidelines are at a principal level and no detailed technical instructions in "how to do"
- Prepared because it makes sense, and to fulfill the recommendation in the indicator 7.1 of QAF



The End!