

# Quality and Metadata at Statistics Denmark

Mogens Grosen Nielsen [mgn@dst.dk](mailto:mgn@dst.dk)

Flemming Dannevang [fda@dst.dk](mailto:fda@dst.dk)

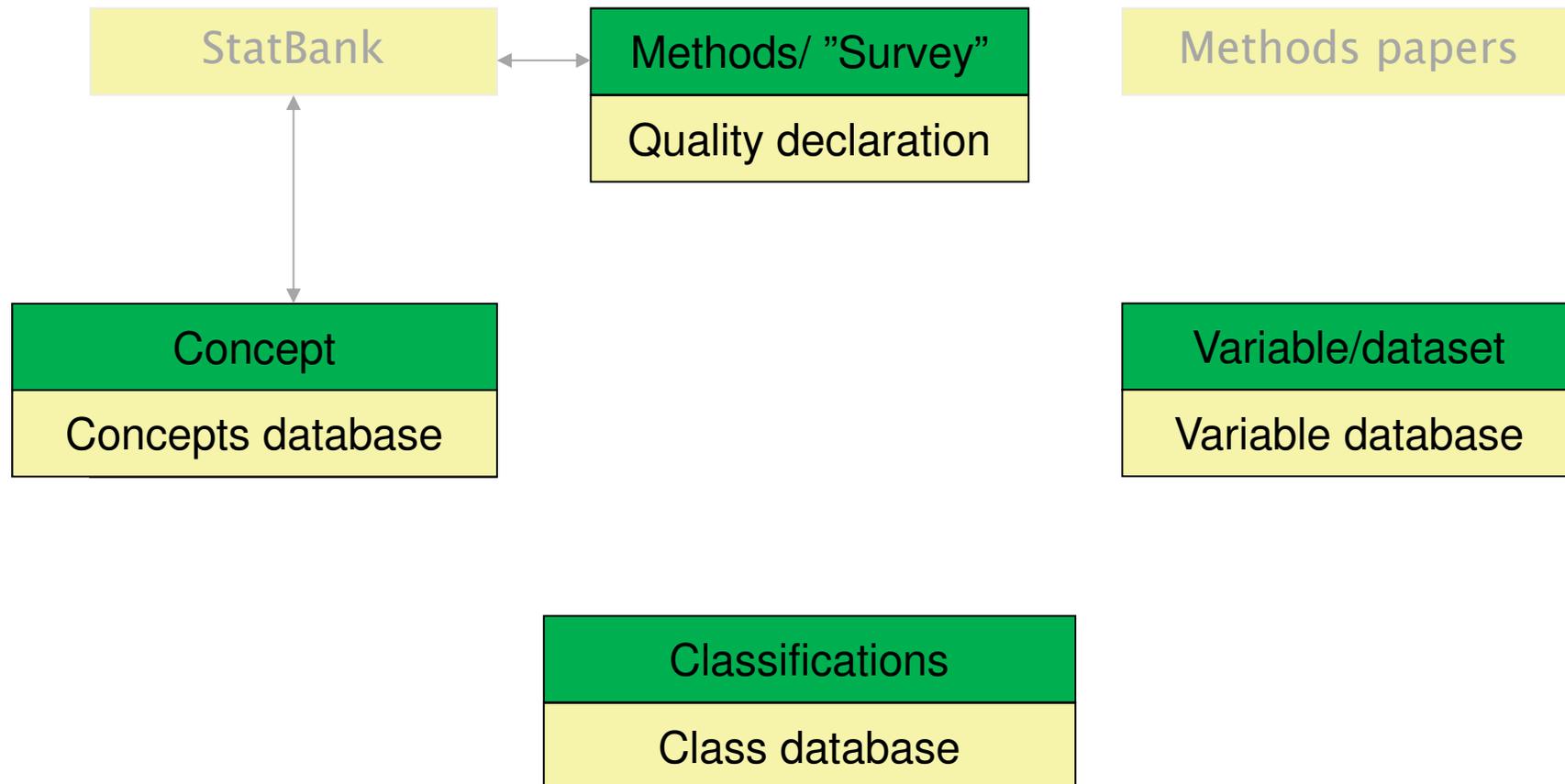
# Agenda

1. Challenges
2. Vision
3. Focus-groups on Metadata
4. Ongoing 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

# Metadata elements



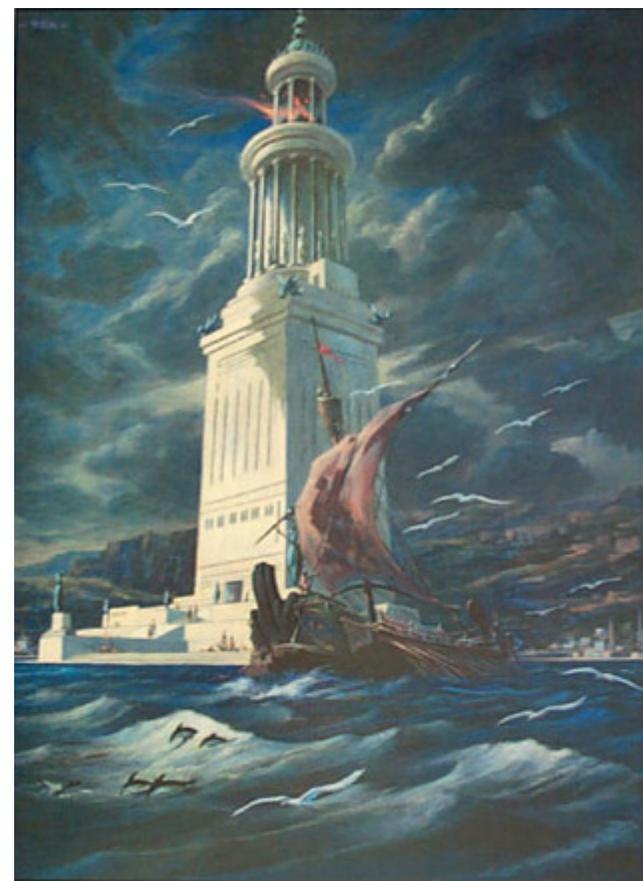
## However...

- Consistent claims of users:

Metadata / documentation isn't good enough

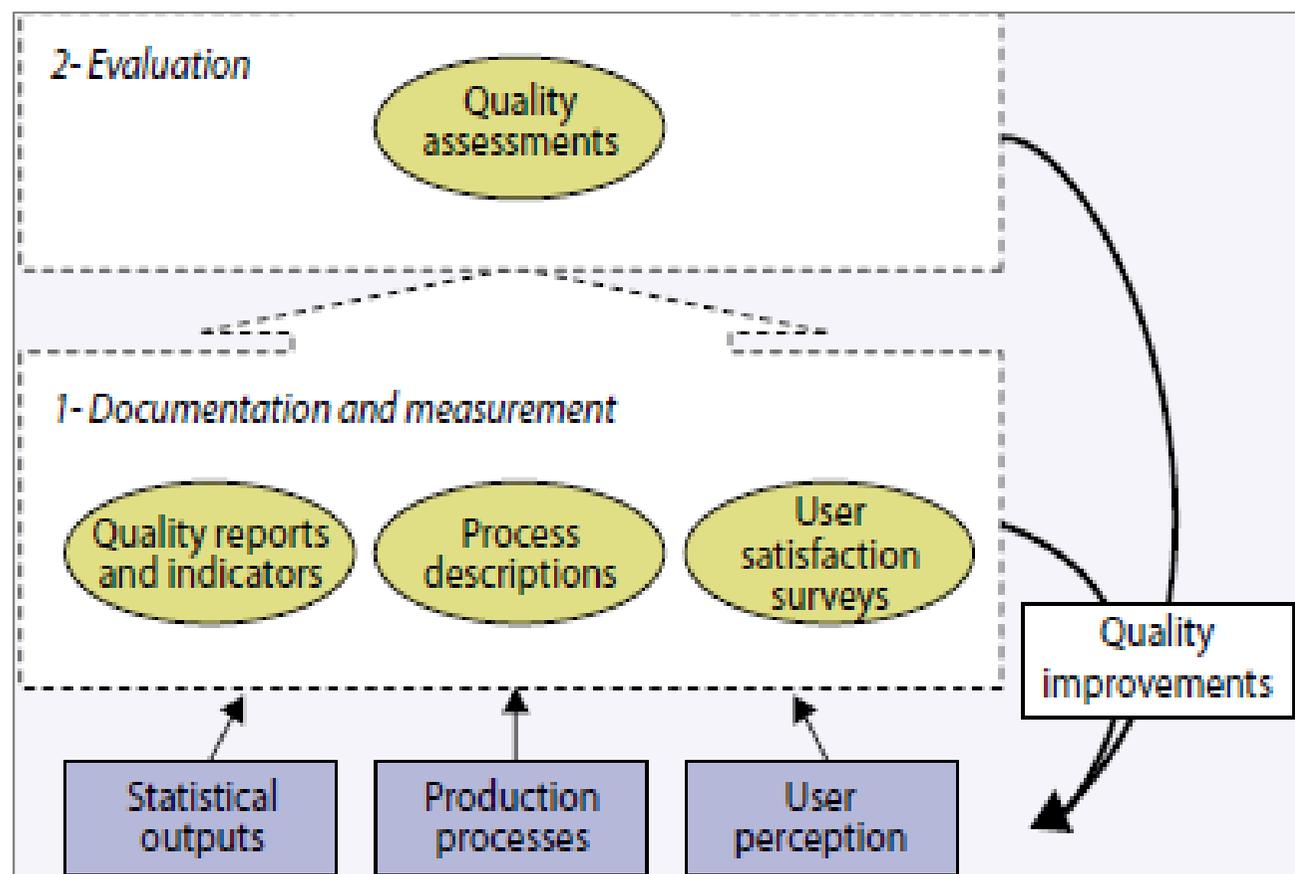
# Vision and focus at Statistics Denmark

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



# Implementation of Eurostat Quality Framework (QAF)

Definition of quality in this context: *The degree to which a set of inherent characteristics of processes and products fulfils requirements* (ISO9000)



# Requirement for the Quality and metadata management

1. Define content: SIMS
2. Guidelines for production processes
3. User needs: survey specific and quality / metadata specific
4. Quality and metadata organisation

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			

# “Challenges on metadata ...”

## **Standards:**

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

## **Eksternal (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 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

# **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"*

Præsentation 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

**ON GORING  
PROJECT TOLD  
SHORTLY**

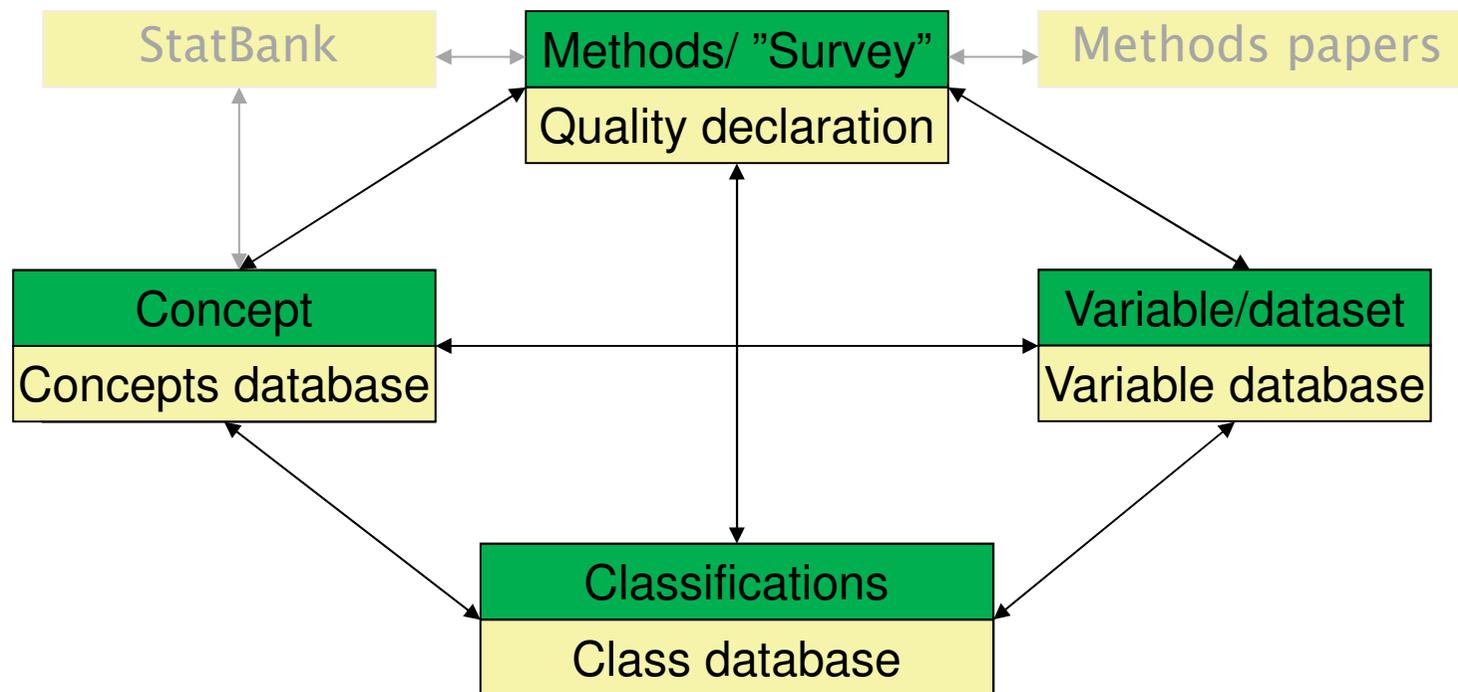
# TITLE

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

# DELIVERABLES

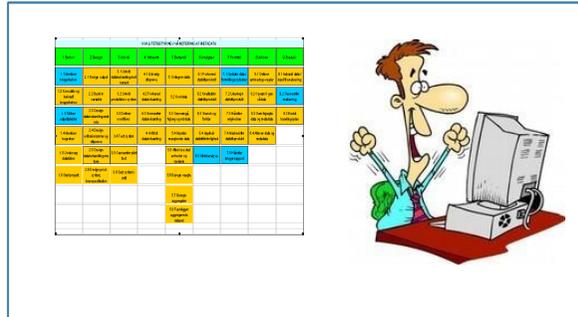
- Quality declarations migrated to new metadata system for all statistics
- Meetings and reports from external users
- Dissemination of dst.dk developed in collaboration with users
- Software (Colectica) installed - internal interface and external interface (dissemination at dst.dk)
- Manuals and courses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

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



**THE SOLUTION**

Enter Q.D.



Publish at Dst.dk

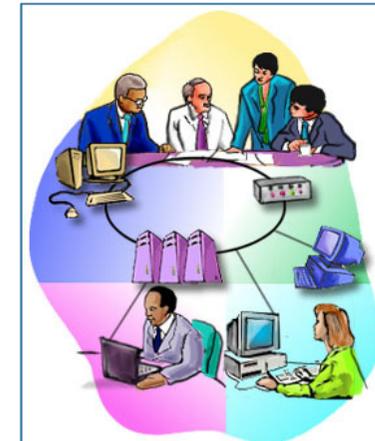


Existing metadata



METADATA IN COLECTICA

Publish at the Intranet



Q.D. to EU



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

**EURO-SDMX Metadata Structure (Dec 2010)**

<b>3</b>	<b>Statistical presentation</b>
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

<b>13</b>	<b>Relevance</b>
13.1	User needs
13.2	User satisfaction
13.3	Completeness

**Single Integrated Metadata Structure**

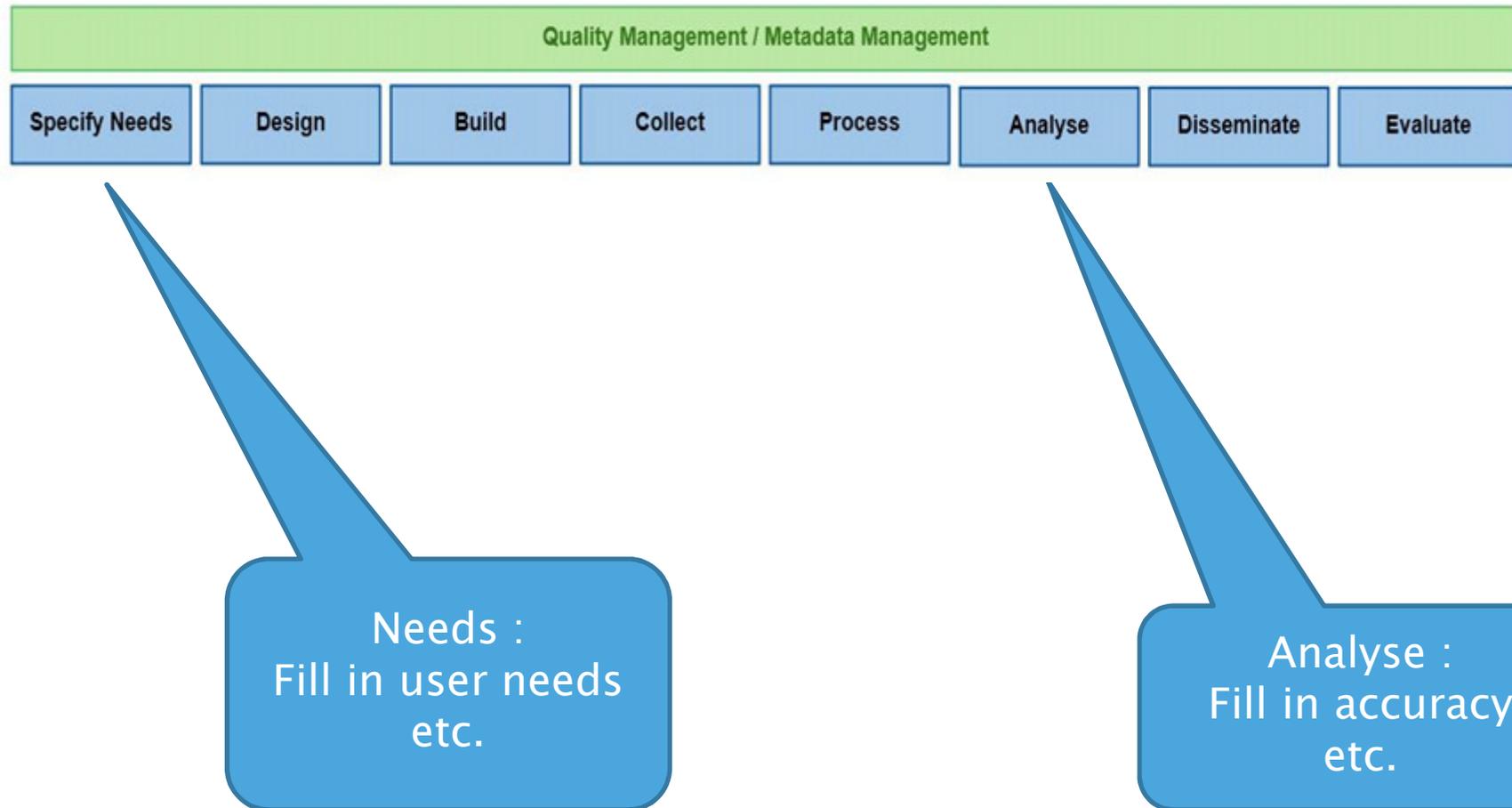
<b>S.4</b>	<b>Statistical presentation</b>
S.4.1	Data description
S.4.2	Classification system
S.4.3	Sector coverage
S.4.4	Statistical concepts and definitions
S.4.5	Statistical unit
S.4.6	Statistical population
S.4.7	Reference area
S.4.8	Time coverage
S.4.9	Base period

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

**ESS Standard for Quality Reports Structure**

<b>IV</b>	<b>Relevance</b>
IV.1	User needs
IV.2	User satisfaction
IV.3	Completeness
<b>IV.3.1</b>	<b>Data completeness - rate</b>

# 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'. The main content area is titled 'Quality Statement' and shows a breadcrumb trail: 'Retail trade index January 2014' (Agency dk.dst, Revision 64, Last Updated 2014-02-03 11:04). Below this, there are tabs for 'Quality Statements', 'Quality Standard', 'External Materials', 'Extended', and 'Identification'. A process flow bar contains buttons for 'General', '1. Specify Needs', '2. Design', '5. Process', '6. Analyse', '7. Disseminate', '8. Archive', '9. Evaluate', and 'All'. The '1. Specify Needs' button is selected. On the left, a list of sections includes 'Statistical presentation', 'User Needs', 'Introduction', 'Data description', 'Institutional mandate', 'Legal acts and other agreements', 'Confidentiality', and 'Confidentiality - policy'. The 'User Needs' section is selected and expanded, showing a description: 'Description of users and their respective needs with respect to the statistical data.' The text in the expanded section reads: '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, is realizing more industries, cause problems en...'. A vertical scrollbar is visible on the right side of the text area. On the far right, a vertical sidebar contains 'History', 'Markdown Preview', and 'Links'.

# 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 content area is titled "Quality Statement" and shows a breadcrumb trail: "Retail trade index January 2014" (Agency dk.dst, Revision 64, Last Updated 2014-02-03 11:04). Below this, there are tabs for "Quality Statements", "Quality Standard", "External Materials", "Extended", and "Identification". A row of process steps is visible: "General", "1. Specify Needs", "2. Design", "5. Process", "6. Analyse", "7. Disseminate", "8. Archive", "9. Evaluate", and "All". The "1. Specify Needs" step is selected. On the left, a list of sections includes "Statistical presentation", "User Needs", "Introduction", "Data description", "Institutional mandate", "Legal acts and other agreements", "Confidentiality", and "Confidentiality - policy". The "Introduction" section is selected and its content is displayed in the main area. The content of the "Introduction" section is: "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." The text "en..." is visible at the bottom right of the content area. At the bottom left, there is an "Editor" tab and a "DDI" label.

# Transmission to Eurostat and customized presentation in reports and at [dst.dk](http://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

# Software elements

Architecture

