# Quality and Metadata at Statistics Denmark

Mogens Grosen Nielsen mgn@dst.dk

Flemming Dannevang fda@dst.dk

STATISTICS DENMARK

# 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

| StatBank | ↔ | Methods/ "Survey" | Methods papers |
|----------|---|-------------------|----------------|

**Methods/ "Survey"**
Quality declaration

**Concept**
Concepts database

**Variable/dataset**
Variable database

**Classifications**
Class database
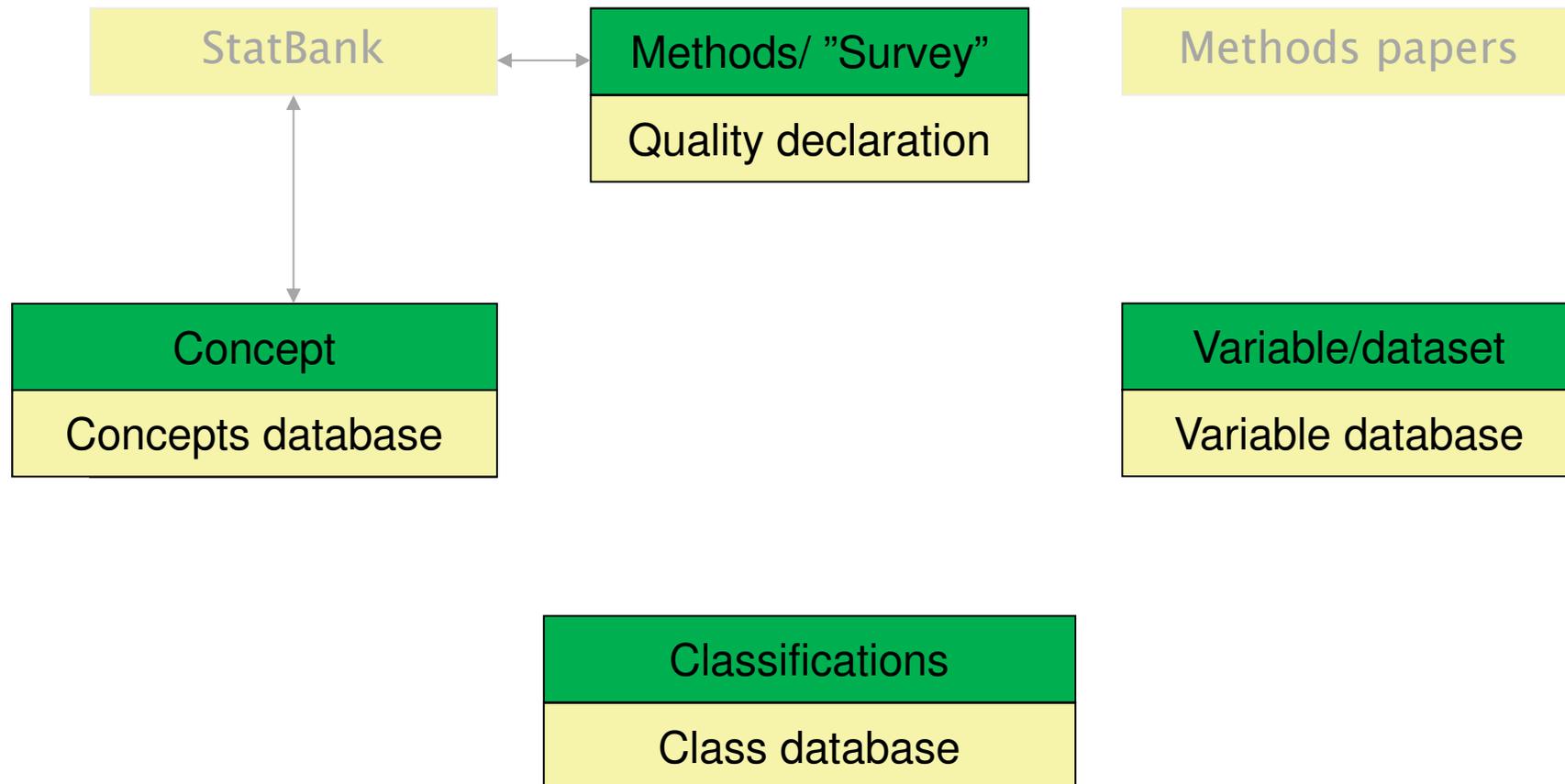
# However…

- 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 ressources
   - Sustainable long term solution

STATISTICS DENMARK
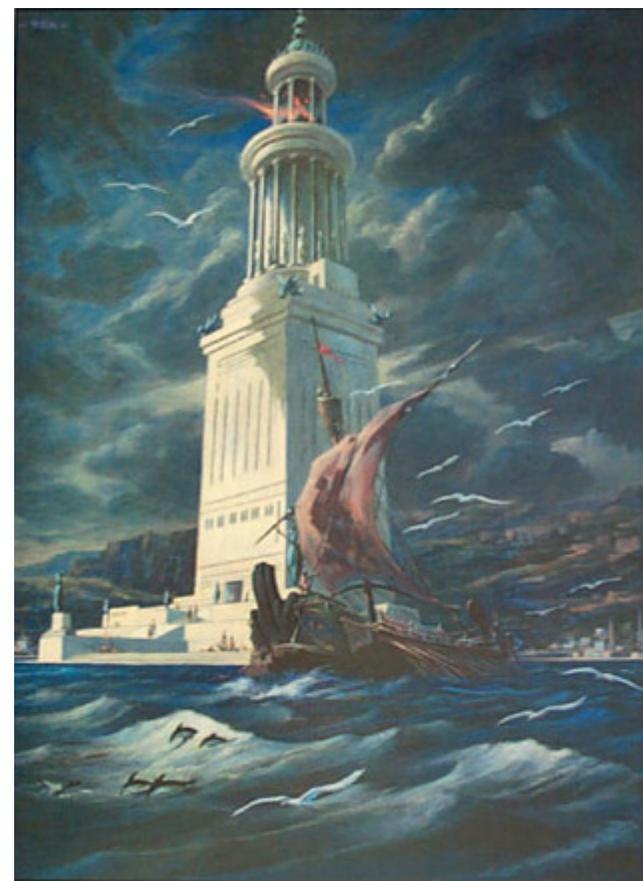
# Implementation of Eurostat Quality Framework (QAF)

Definition of quality in this context: *The degree to which a set of inherent characteristics of processes and products fulfils requirements* (ISO9000)
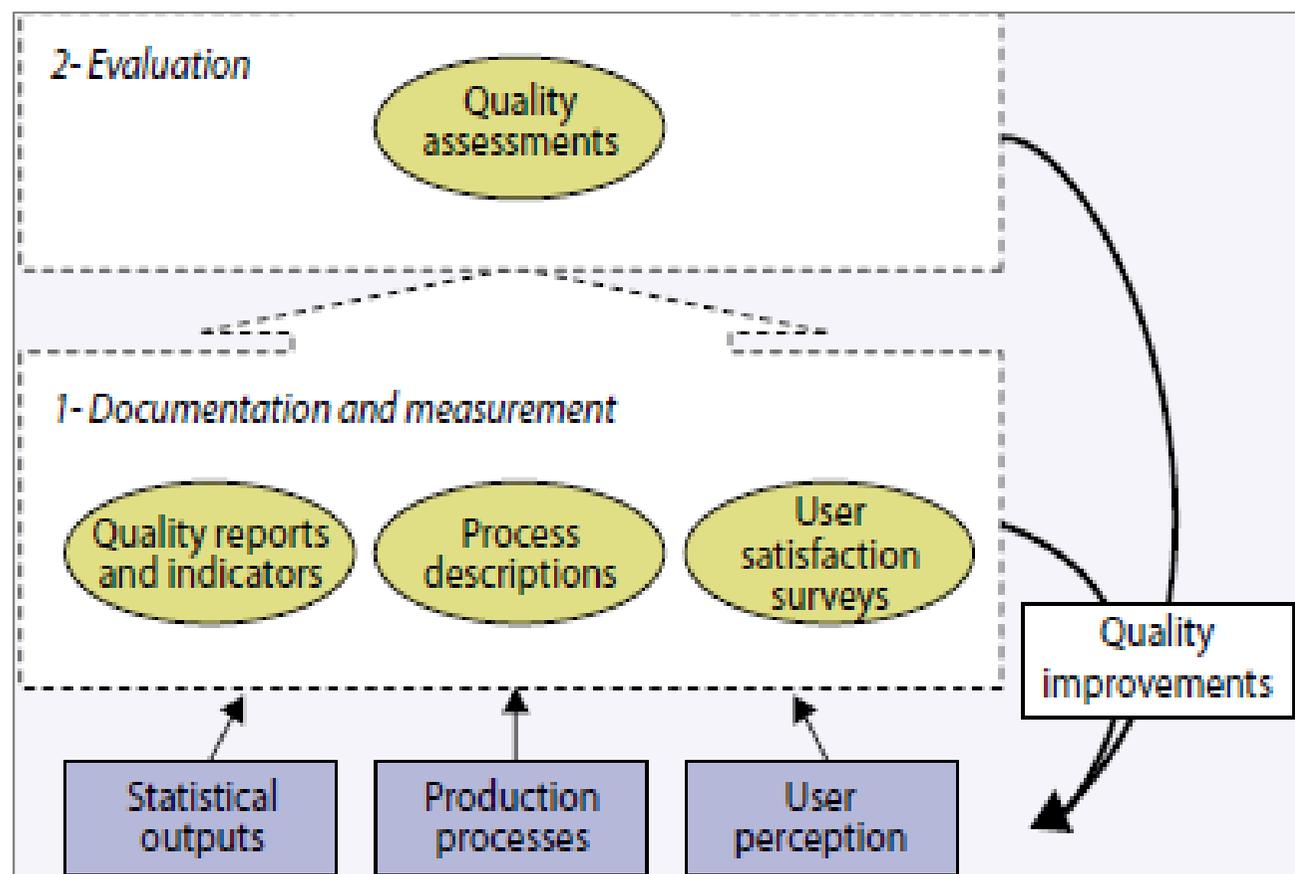
# Requirement for the Quality and metadata management management

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

# "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? | Præsentation of model / solution on integrated metadata and search types | How is the balance between needs and solution |

*"Demand"*　　　　　　　　*"Supply"*　　　　　　　　*"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 Documenation initiative (DDI) and SDMX
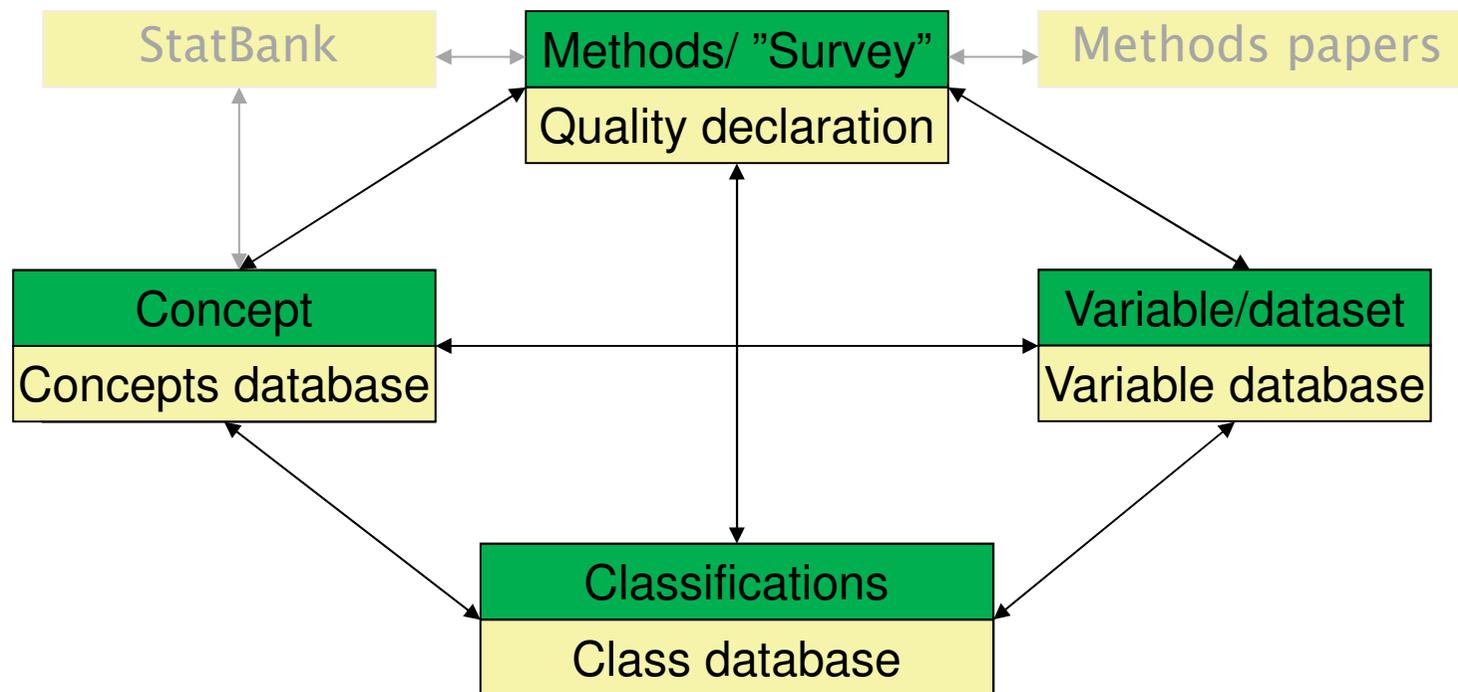
**THE SOLU-TION**

**Enter Q.D.**
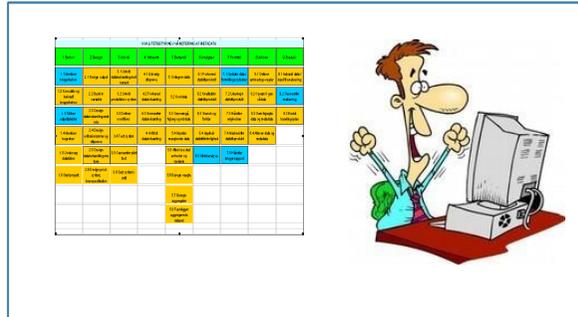
**Publish at Dst.dk**

Existing metadata

**METADATA IN COLECTICA**

**Q.D. to EU**

**Publish at the Intranet**
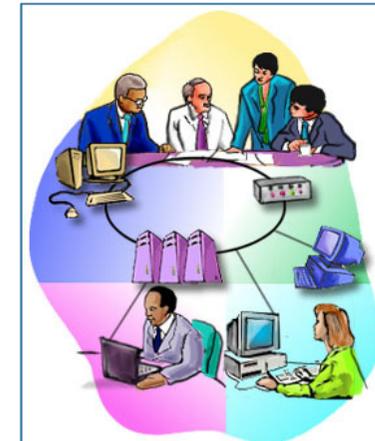
# Single Integrated Metadata Structure (SIMS) and reporting formats: ESMS and ESQRS -

| EURO-SDMX Metadata Structure (Dec 2010) | Single Integrated Metadata Structure | ESS Standard for Quality Reports Structure |
|---|---|---|

| 3 | Statistical presentation | S.4 | Statistical presentation |
|---|---|---|---|
| 3.1 | Data description | S.4.1 | Data description |
| 3.2 | Classification system | S.4.2 | Classification system |
| 3.3 | Sector coverage | S.4.3 | Sector coverage |
| 3.4 | Statistical concepts and definitions | S.4.4 | Statistical concepts and definitions |
| 3.5 | Statistical unit | S.4.5 | Statistical unit |
| 3.6 | Statistical population | S.4.6 | Statistical population |
| 3.7 | Reference area | S.4.7 | Reference area |
| 3.8 | Time coverage | S.4.8 | Time coverage |
| 3.9 | Base period | S.4.9 | Base period |

| 13 | Relevance | S.14 | Relevance | IV | Relevance |
|---|---|---|---|---|---|
| 13.1 | User needs | S.14.1 | User needs | IV.1 | User needs |
| 13.2 | User satisfaction | S.14.2 | User satisfaction | IV.2 | User satisfaction |
| 13.3 | Completeness | S.14.3 | Completeness and *R1. Data completeness - rate for U* | IV.3 | Completeness |
| | | S.14.3.1 | *R1. Data completeness - rate for P* | IV.3.1 | Data completeness - rate |

# Work processes and quality declarations

# Work processes and quality declarations in Colectica – Retail Trade index

# Work processes and quality declarations in Colectica – Retail Trade index

# 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**

# Software elements



Architecture

The End!