Progress in the use DDI, SDMX, GSBPM and other standards

Mogens Groesen Nielsen, Statistics Denmark, Anne Sofie Kjeldgaard & Jannik Vestergaard Jensen, Danish Data Archive
Agenda

1. Progress at Statistics Denmark and Danish National Archive (DNA)
2. Metadata portal
Vision and focus at Statistics Denmark

1. Metadata about content and quality must
   - help users in their knowledge processes
   - give users precise information about our products

2. International standards and standard software must enable:
   - Cost efficient solution
   - Gradual implementation with few resources
Metadata – the SDMX definition as the main definition

Reference metadata:
- Conceptual metadata
- Methodological and processing metadata
- Quality metadata

Structural metadata:
- Metadata that act as identifiers and descriptors of the data
Integrated and reuseable metadata using DDI and SDMX ("classical metadata")

Statbank

Methods

Quality Declarations

Concept

Conceptdatabase

Concepts database

Classifications

Classificationsdatabase

Classifications database

Variable/dataset

Variabeldatabase

Papers on methods

“Papers on methods”
Metadata needed to implement Code of Practice (CoP) and Quality Assurance Framework (QAF) from Eurostat
Standard for Quality: Single Integrated Metadata Structure (SIMS)

- Content (population, concepts, reference time etc)
- Statistical processing (sources and methods)
- Information on 5 quality dimensions
  1. Relevance
  2. Accuracy and reliability
  3. Timeliness and punctuality
  4. Comparability
  5. Accessibility and Clarity
Generic Statistical Business Process Model (GSBPM)

- Quality and metadata management processes at the top
- Measuring quality of production processes and quality of products
GSBPM and work processes with focus and quality declarations

Specify Needs
Design
Build
Collect
Process
Analyse
Disseminate
Evaluate

Needs
Prepare user needs etc.

Analyse:
Fill in accuracy etc.
GSBPM and work-processes overall

END USER

1. NEEDS
   IDENTIFY NEEDS
   USE STATISTICS

2. DESIGN
   DECIDE ON PROGRAMME
   APPROVE

3. BUILD
   CONFIRM NEEDS, PLANNING ETC
   DESIGN OUTPUT, COLLECTION ETC
   ADVISE ON COMMON DESIGN ELEMENTS
   BUILD COLLECTION, OUTPUT AND PROCESSING

4. COLLECT
   COMPLETE QUESTIONNAIRE ETC

5. PROCESS
   CONDUCT DATA COLLECTION
   DATA EDIT PROCESS ETC

6. ANALYZE
   ANALYSE
   DISSEMINATE

7. DISSEMINATE
   CONDUCT EVALUATION

7. EVALUATE
   EVALUATE & SUGGEST IMPROVEMENTS
GSBPM and work-processes overall

Reference metadata: needs, purpose etc
GSBPM and work-processes overall

Structural metadata: DDI on questionnaire, variables and cubes and ref. metadata
Reference metadata: concepts, population etc
GSBPM and work-processes overall

Metadata: DDI used to create survey system, databases and output systems etc
Survey system used
GSBPM and work-processes overall

Reference metadata on quality
GSBPM and work-processes overall

Structural and reference metadata used for dissemination.
Output-systems used in the production
GSBPM and work-processes overall

All metadata used for evaluation
THE SOLUTION (QUALITY)

Enter quality-information

Publish at Dst.dk

Publish at the Intranet

Q.D. to EU

Existing metadata

METADATA IN COLECTICA
DDI Structure in Colectica

- **GROUP**
  - **STUDY–UNITS**
  - **CONCEPTS**
  - **QUALITY STATEMENT**

- Retail trade index
  - Retail trade index 2012
  - Retail trade index January 2014
- Data Files
- Metadata Packages
  - Retail trade index January 2014
- Conceptual
  - Concept Sets
    - Retail trade index
    - Universe Sets
    - Organization Sets
    - Quality Statement Sets
- Classifications
- Data Collection
- Data
- Modules
- Quality Statements
Software elements

Customized development:
Integration—components: Web, Eurostat etc
Add-ins to Designer: simplified UI (Task-mode)

Architecture

Colectica Designer
Colectica SDK
Colectica Repository
Web
Eurostat Quality Report Delivery Tool (forthcoming)
EDAMIS
Status at Statistics Denmark

- Quality declarations for all statistics (300+) migrated into a common DDI structure in Colectica
- Work processes in Colectica supporting workflow (including centralized quality control)
- Dissemination at dst.dk developed in collaboration with users
- Automated system to transmit quality reports to Eurostat complying with Metadata Structure Definitions (ESMS and ESQRS)
- Software: Colectica in place with customized interface and integrations-components for EU and for our Content Management System (Sitecore)
Plans for 2015 at Statistics Denmark

- GSBPM, GSIM, DDI and SDMX used for implementation of variables, classifications and concepts in Colectica
- Harmonisation of statistical concepts
- Focus-groups and improved dissemination of integrated metadata
- Use of GSBPM as basis for improvement of a) the Quality Management System and b) reuse of metadata c) metadata driven production
- Metadata portal in collaboration with other organisations
Progress at the Danish National Archive (DNA) / Danish Data Archive

- Statistical package SPSS to DDI-3.1 ingest
- Consolidating metadata services / specificering upgrade decision points
- Integrating DNA methods and standards
- Specified features for longitudinal requirements
- KOOR – Coordinating statistical data / metadata across institutions: Statistical office, National health entity and the national archive
• Metadata portal
Metaportal

• Need for transparency and overview of data available for research
• Combine statistical registers and sample-survey information
• Inter-organisational project (Stat DK, DDA and more)
• Standards: DDI and SDMX
• Flexible and scalable solution
Metadataportal: work-processes, metadata- og data-exchange

Researchers and other users

Metadata og data

METADATA HUB

METADATA & DATA

Statistics Denmark and DNA

Other provider organisations

Metadata og data
Metaportal – technical solution #1

Metadata hub:

- Quick access to search, show og download metadata
- Search on items and thematic (concept-) search
- Metadata items integrated
- Links: search via metadata hub and links back to provider
- The solution must be simple and scalable
Metaportal – technical solution #2

Provider organisations

- Expose their main metadata in the hub via a standard-based interface
- Provide all types of metadata and data as described in the metadata hub
The End!