Agenda

1. Objectives and expected output of mission
2. Short introduction to metadata and metadata questions from ICBS
3. Methodology
4. Current situation
5. Vision (goal, benefits and principles)'
6. Introduction of end-to-end perspective on the use of metadata in the production of statistics;
7. Roadmap / timetable

Use cases – will be presented and discussed during the program

- Use case 1: A simple questionnaire
- Use case 2: A simple data-structure (survey/admin. micro-data)
- Use case 3: How to create an aggregated dataset using N-cube
- Use case 4: A quality-declaration

STATISTICS DENMARK

• 1. Objectives of mission and methodology

STATISTICS DENMARK

STATISTICS DENMARK

Objectives of the mission: Purpose (ToR)

- Development of the part of ICBS' strategy plan regarding metadata and quality.
- ICBS' specific needs regarding choice of data formats and corresponding software solutions will be discussed

Output from the mission (ToR)

1. Mission report with recommendations regarding the metadata and quality strategy, data formats and software solutions.
2. Live demonstration (with Colectica) of the importance, advantages and functionnalities for internal and external use, of a statistical metadata system (SMS) which is centrally integrated. Hands-on experience. Demonstration of user friendliness in different uses.
-

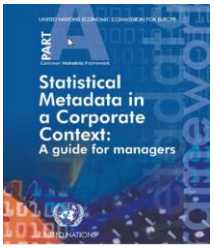
- SHORT INTRODUCTION TO METADATA DEFINITIONS, USERS AND AND PRINCIPLES

"Metadata – an old invention"

- Library-catalogs invented 245 BC by Callimachus in Greece
- Today:
  1. New technology
  2. More content
  3. Same purpose: to help users and producers

STATISTICS DENMARK

Definition of Metadata and Statistical Metadata System (SMS)



- Statistical Metadata: "data about statistical data comprising data and other documentation that describe objects in a formalised way"
- Statistical Metadata System (SMS) "A data processing system that uses, stores and produces statistical metadata". The term system refers to the people, processes and technology involved in managing statistical metadata.

STATISTICS DENMARK

Metadata – definitions #1

- Statistical metadata broadly: information objects in the Generic Statistical Information Model (GSIM).
- A narrower definition: Metadata can splitted into **reference metadata** and **structural metadata**. This definition can be found in the SDMX Metadata Common Vocabulary (MCV):

STATISTICS DENMARK

STATISTICS DENMARK

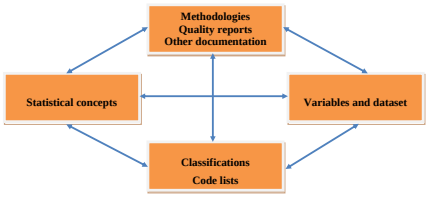
Metadata – definitions #2

- Reference metadata: *"Metadata describing the contents and the quality of the statistical data."*
- Reference metadata includes:
- Conceptual metadata: *"Metadata describing the concepts used and their practical implementation, allowing users to understand what the statistics are measuring and, thus, their fitness for use"*
- Methodological and processing metadata: *"Metadata describing methods used for the generation of the data (e.g. sampling, collection methods, editing processes)"*
- Quality metadata: *"Metadata describing the different quality dimensions of the resulting statistics (e.g. timeliness, accuracy)"*

Metadata – definitions #3

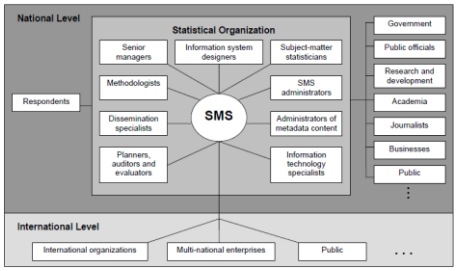
- Structural metadata: *"Structural metadata are metadata that act as identifiers and descriptors of the data. Structural metadata are needed to identify, use, and process data matrixes and data cubes, e.g. names of columns or dimensions of statistical cubes. Structural metadata must be associated with the statistical data, otherwise it becomes impossible to identify, retrieve and navigate the data."*

Integration of "Classical metadata"



STATISTICS DENMARK

Metadata users



Metadata principles – the most important

1. **Reuse:** Reuse metadata where possible for statistical integration as well as efficiency reasons
2. **Statistical business process model**: Manage metadata with a focus on the overall statistical business process model (GSBPM)
3. **Active metadata**: Make metadata active to the greatest extent possible. Active metadata are metadata that drive other processes and actions. Treating metadata this way will ensure they are accurate and up-to-date.

- Questions from ICBS

STATISTICS DENMARK

Introductory discussion – metadata questions

What does DDI particularly relevant and advantegeous to work with as a format/standard?

- DDI (together with SDMX) has been chosen by UN, EU and others
- *Reuse of concepts, classifications, questions, variables due to the integration in DDI*
- *Connect to GSBPM (lifecycle)*
- *Active metadata (metadata-driven production via sharing of software using the same standard)*
- *All metadata in one SMS using one standard*

Colectica

- Who is using Colectica? *Currently New Zealand, Canada and Denmark. Ohters are testing : Eurostat, Ireland, England, Sloveniea and others. rchiving organisations.*
- Alternatives to Colectica? *Primarily a number of open source product. See DDI alliance website. Besides Colectica, there is a commercial tool called Sledgehammer.*
- What is the advantage with Colectica compared to other tools or systems? *a commercial product , standard DDI tool*

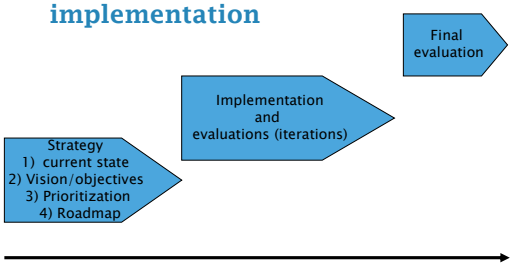
Integration with Statbank and webpage

- *In the longer term all metadata should be in Colectica, og should be reachable through an API-solution, e.g. homepage, StatBank (datawarehouse) etc. In the short term Statistics Denmark focuses on the quality declations.*

STATISTICS DENMARK

•Methodology

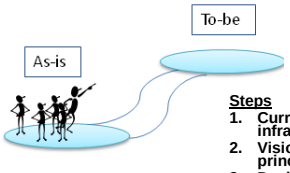
Quality and metadata strategy – development and implementation



STATISTICS DENMARK

STATISTICS DENMARK

Methodology: Business process management – from "as-is" to "to-be"



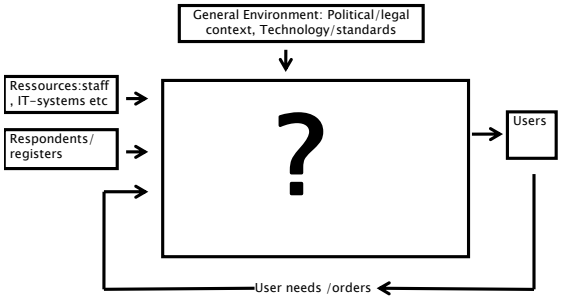
Steps

1. **Current state – processes and infrastructure**
2. **Vision – goals, benefits and principles**
3. **Decide on prioritization of changes – gap analysis**
4. **Create roadmap with iterations**

Implement projects with iterations

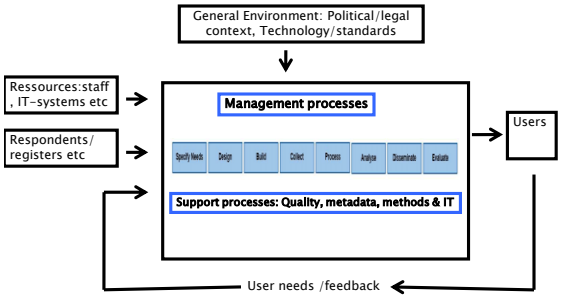
STATISTICS DENMARK

Business proces perspective.



STATISTICS DENMARK

Business proces perspective with management- core- and support-processes



STATISTICS DENMARK

STATISTICS DENMARK

- Current state – issues and recommendataions

3 main elements in the analysis from last mission

A: The situation today "as-is"

- ***Environment***
  - ***1: Government/ regulations etc***
  - ***2: Technology***
  - ***3: User needs***
  - ***4: Standards***
- ***Status / ongoing activities***
  - ***Strategic, business processes and technological***

B: Objectives / results "to-be"

General:

1) Cost-efficient production using common metadata 2) Quality of statistical products 3) Fulfillment of user needs on documentation

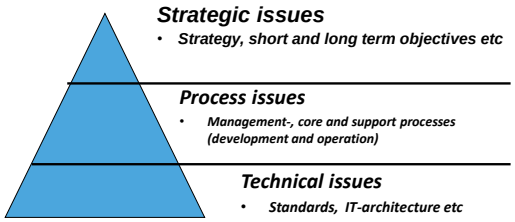
Functions:

1) User needs and feedback 2) Dis-semination of metadata on website 3) Quality Declarations d) common concepts etc 4) GSBPM

C: How to get from A to B

Issues and recommendations

Issues at three levels



STATISTICS DENMARK

Issues and recommendations at the strategic level

Issues

- Awareness and commitment from top level management is needed.
- Metadata initiatives are not connected to General Strategy of ICBS or the Dissemination Strategy of ICBS.

Recommendations:

- Integrate metadata in the ICBS strategy
- Employ a business process model perspective
- Initiate metadata project. Projects on quality, processes and part of dissemination could be integrated

Issues and recommendations related to end users

Issues

- Statistical illiteracy
- Difficult to know what data to use
- Cost-effective transmission of data to international organisations (webservice and sdmx format)
- Metadata on web-site incomplete.

Recommendations:

- Analysis of known (already identified) problems on dissemination (FAQ, typical problem types, e.g. which variables/statistics should be used to shed light on the problems which the user needs to solve)
- Establish so-called focus groups discussing the roles of metadata.

Issues and recommendations at process level

Issues

- Metadata is time consuming for subject matter units
- Lack of systematic planning/methodology.
- Lack of management of processes,
- Lack of integration between processes, e.g. dissemination, researchers. (e.g. GSBPM)
- How should Subject Matter Units be supported using GSBPM and overall framework?
- Extract Transform Load (ETL) has been introduced but difficult to introduce common standards/models to subject-matter statisticians.
-

Recommendations:

- Give general information on benefits etc., awareness of the importance of integration, etc., to all staff
- Implement GSBPM (includes translation and adaptation of descriptions to ICBS).
- Prepare as-is documentation of selected surveys on using simple templates.

STATISTICS DENMARK

Issues on standards and technical implementation

Issues

- Lack of knowledge on standards and harmonization of standards and metadata across departments
- Challenges on how to handle new website technology
- Lack of common transition procedures from paper to web
- How does CBS standards (parent variables, dictionaries, de facto metadata) play together with DDI, SDMX and international standards in general
- Micro vs. Macro. DDI and SDMX uses different approaches – how to reconcile?
- Where to use DDI and where to use SDMX?

Re*commendations*:

- Training, communication of purpose. e.g. terminology on metadata: DDI, SDMX, GSBPM, GSIM etc.
- Harmonization; use SDMX and DDI. Integrate work between Dissemination Unit and subject matter units
  - Concepts, e.g. Statistical Yearbook (Statistical Abstract) and Subject Matter Units domains
  - Code lists
- Continue work on common variables and code-lists with a view to moving towards sdmx and DDI standards
- Implement standard DDI and SDMX-tools in order to be cost-effective and to ensure the use of international standards.

Metadata functions and tasks decided on last mission

Functions	Tasks
User needs and feedback	• Analysis of known problems (FAQ, typical problem types, e.g. what variables to use) • Focus groups discussing the roles of metadata
Dissemination of statistical information (website)	• Define how metadata can support the information on the website • Integration with StatBank and other dissemination products
Improving the quality of statistical data and transparency of methodologies	• Implementation of Quality Declarations
Methodological activities	• Documentation, standardization and harmonization of common concepts, classifications, code-lists • Enter the common elements (concepts etc.) into IT solution
GSBPM: Planning, implementation, evaluation and documentation of statistical processes + guidelines	• Adoption of GSBPM • Development of guidelines • Documentation of existing processes (as-is) • Development of guidelines on the use and production of metadata in the production processes • Training and implementation

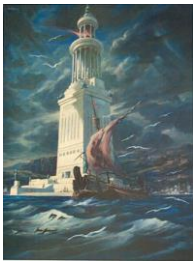
Recommendation for long term strategy on metadata from *Activity C.5 Awareness about the National Statistical System*

- Adoption of international standards (structured according to 'the diamond')
- Implementation of a central statistical metadata system (SMS)
- Standardization of maintenance of metadata
- Measurement of users' satisfaction (with metadata)

STATISTICS DENMARK

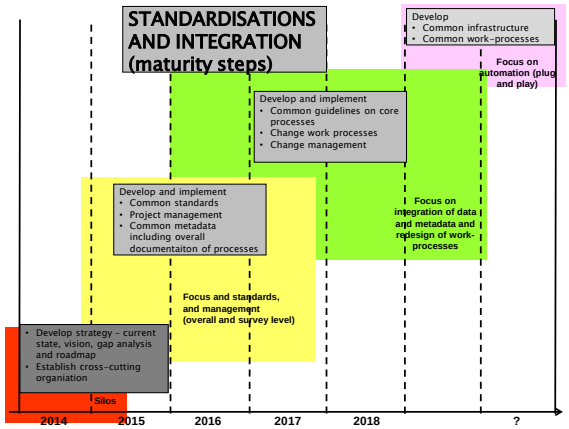
- Vision, objectives, principles and benefits

Vision and focus at Statistics Denmark



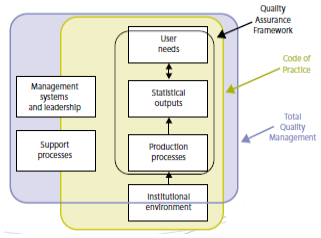
1. Statistical information must help users in the "turbulent information-sea"
2. Metadata about content and quality must
   - help users in their knowledge processes
   - give users precise information about our products
3. International standards and standard software must enable:
   - Cost efficient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

STATISTICS DENMARK



STATISTICS DENMARK

Principles and quality frameworks



CODE OF PRACTICE

Institutional environment

- P1: professional independence
- P2: mandate for data collection
- P3: adequacy of resources
- P4: quality commitment
- P5: statistical confidentiality
- P6: impartiality and objectivity.

Statistical procedures

- P7: sound methodology
- P8: appropriate statistical procedures
- P9: non-excessive burden on respondents
- P10: cost effectiveness.

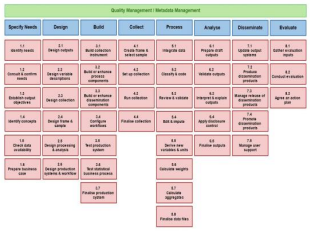
Statistical results

- P11: relevance
- P12: accuracy and reliability
- P13: timeliness and punctuality
- P14: coherence and comparability*
- P15: accessibility and clarity

STATISTICS DENMARK

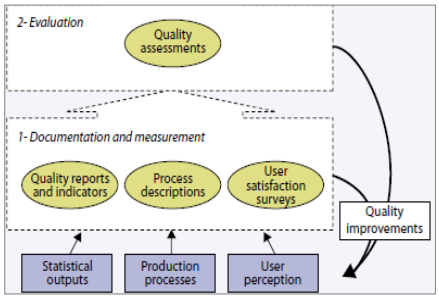
GSBPM – quality and metadata management processes at the top

- Quality and metadata management processes at the top
- Measuring quality of production processes and quality of products



STATISTICS DENMARK 35

Implementation of Eurostat Quality Framework



STATISTICS DENMARK 36

STATISTICS DENMARK

Standardized metadata systems in QAF

- **Indicator 15.5. Metadata are documented according to standardized metadata systems**

Metadata principles – the most important

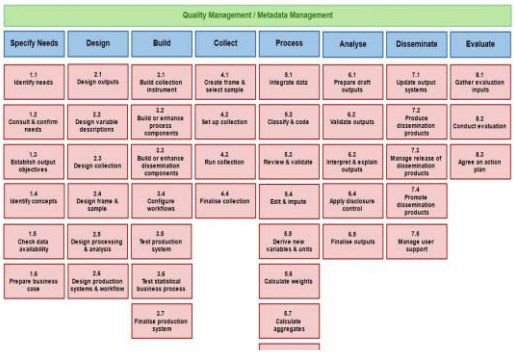
1. **Reuse:** Reuse metadata where possible for statistical integration as well as efficiency reasons
2. **Statistical business process model**: Manage metadata with a focus on the overall statistical business process model (GSBPM)
3. **Active metadata**: Make metadata active to the greatest extent possible. Active metadata are metadata that drive other processes and actions. Treating metadata this way will ensure they are accurate and up-to-date.

Benefits – short and long term

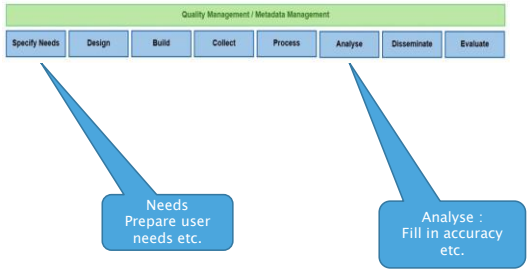
- Cost efficiency
  - …
- Fulfilment of user needs
  - …
- Improved quality
  - …

- End-to-end perspective on work-processes, quality and metadata

The Generic Statistical Process model

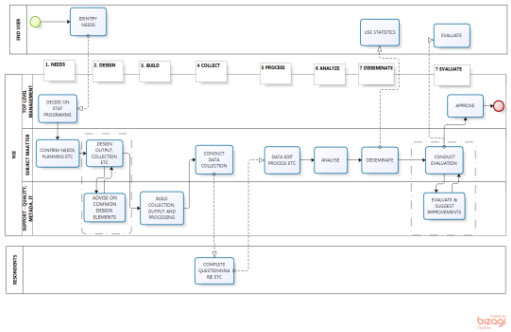


GSBPM and work processes with focus and quality declarations

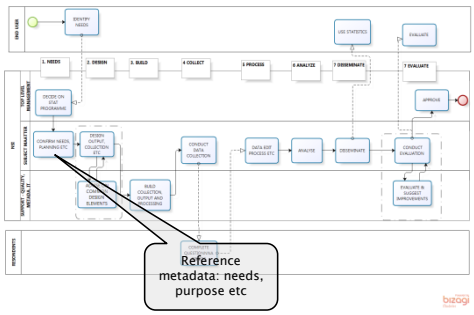


STATISTICS DENMARK

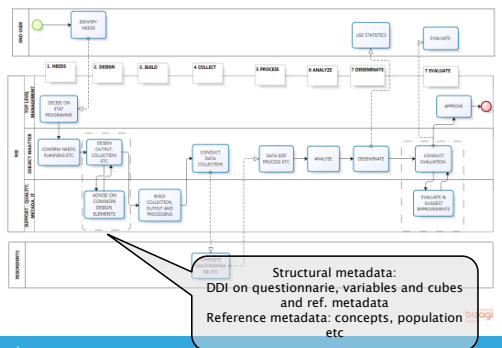
GSBPM and work-processes overall



GSBPM and work-processes overall

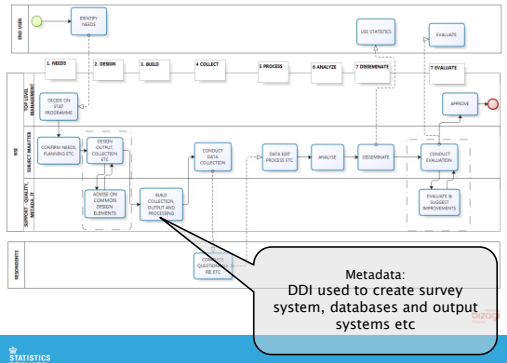


GSBPM and work-processes overall

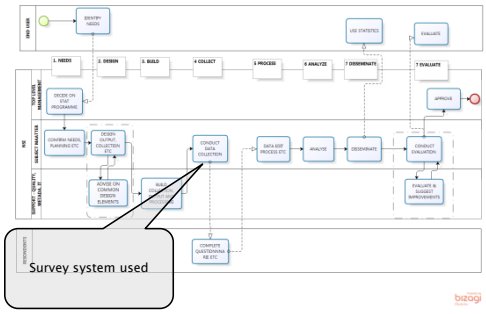


STATISTICS DENMARK

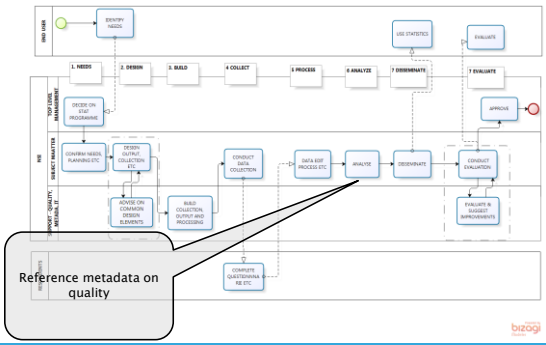
GSBPM and work-processes overall



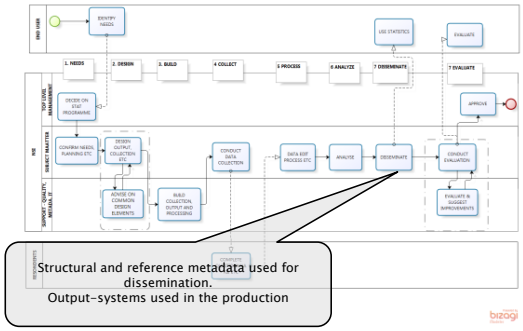
GSBPM and work-processes overall



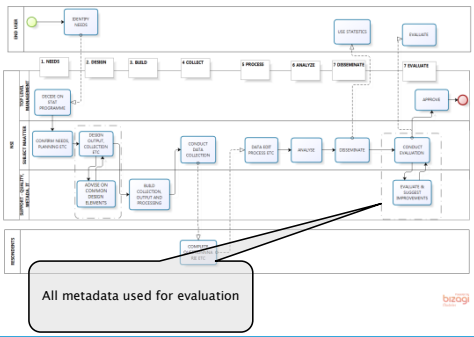
GSBPM and work-processes overall



GSBPM and work-processes overall



GSBPM and work-processes overall

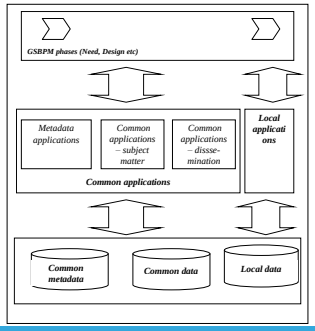


Guidelines with focus on metadata phase 1: Needs

Elements	Description of element
Description of the work processes	In this phase subject matter statistician prepares document with user needs and objectives and a work plan. Steps:
Input	Requirements from Ministries etc
Output	Decisions / project plan / Objectives of the survey (opportunities)
People involved	Management, subject matter statistician, external users
Regulations and guidelines	
Metadata used produced	Quality declarations: user needs, purpose, content
It solution	

STATISTICS DENMARK

GSBPM processes, applications and data



Use cases

- Use case 1: A simple questionnaire
- Use case 2: A simple data-structure (survey/admin. micro-data)
- Use case 3: How to create an aggregated dataset using N-cube
- Use case 4: A quality-declaration

- Strategy including roadmap

STATISTICS DENMARK

Strategy outline

1. Introduction, including current state
2. International standards, definition of metadata and roles of a Statistical Metadata System
3. Vision and objectives (short and long term), principles and benefits
4. Tasks and time-table (roadmap)
5. Organisation

Annex 1. Guidelines for production and use of metadata

Annex 2. IT-solution (short and long term)

1. High level architecture
2. Applications and databases

Roadmap / timetable