Metadata and classification system development in Bosnia and Herzegovina

23. april 2012

Mogens Grosen Nielsen

Statistics Denmark

Outline of introduction to metadata project in Bosnia and Hercegovina

1. Introduction
2. Objectives
3. Three main elements in report
4. Issues at three levels
5. Business process model
6. Standards

Objectives for component 3

Component 3: Strengthening the Institutional Capacaty of BIH Statistical System (comprises long term strategy and metatdata)

- *sustainable development*

- *better communication*
  - *between the three statistical institutes*
  - *between providers and users of the statistics.*

Objectives/result component 3.2

Component 3.2: metadata

- *Technical documentation including selection of IT tools for Metadata system prepared*
- *Development plan prepared and preparatory activities carried out for Classification server*

Benchmark

- *Metadata strategy developed by 8th project quarter*
- *Database model with data flow diagram prepared by 8th project quarter*
- *Plan on development of software for classification database by 8th project quarter*

DANMARKS STATISTIK

Objectives – this mission

- *Improved the knowledge of all participants regarding metadata and classification systems;*

- *Defined the basic concept of metadata and classification systems in Bosnia and Herzegovina Statistics*

- *Report detailing the current state of classifications and metadata. Detailed road map for implementation of other activities.*

Suggested activities for this mission

- *basic metadata functionality and classification system from experts' experience*

- *Assessment of current situation*

- *Overview of classifications use, quality systems used, documentation systems, present software.*

- *Determination of possibilities, frame and needs (EU requirements);*

- *Discussion on location and responsibilities for updating metadata/classifications;*

DANMARKS STATISTIK

Three main elements in report

A: The situation today

- **Status / ongoing activities**

- **Issues - recorded**

- **Environmet**
  - *1: government/ regulations etc*
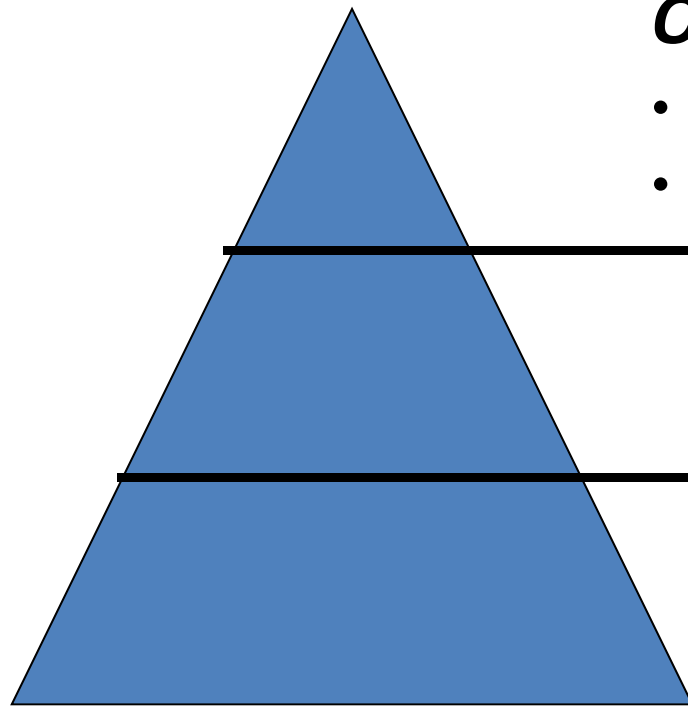 - *2: Technology / People*
  - *3: Users*
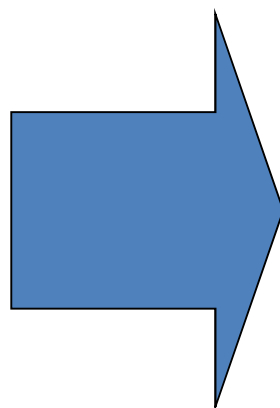
B: Objectives / results

- *Sustainable development*

- *Better communication*

- *Strategy for metadata*

- *Database models / flows*

- *Classification*

C: How to get from A to B
Recommendations

Issues at three levels

Organisational issues

- **Strategy**
- **Core-process and support processes**

Process issues

- **Metadata as support processes**
- **Management of metadata etc**

Implementation issues

- **Standards**
- **Technology etc**

DANMARKS STATISTIK

>> The role NSI's

- A lighthouse in the turbulent sea of information

- Focus on metadata to support knowledge processes

- Create metadata to give user exact knowledge on products

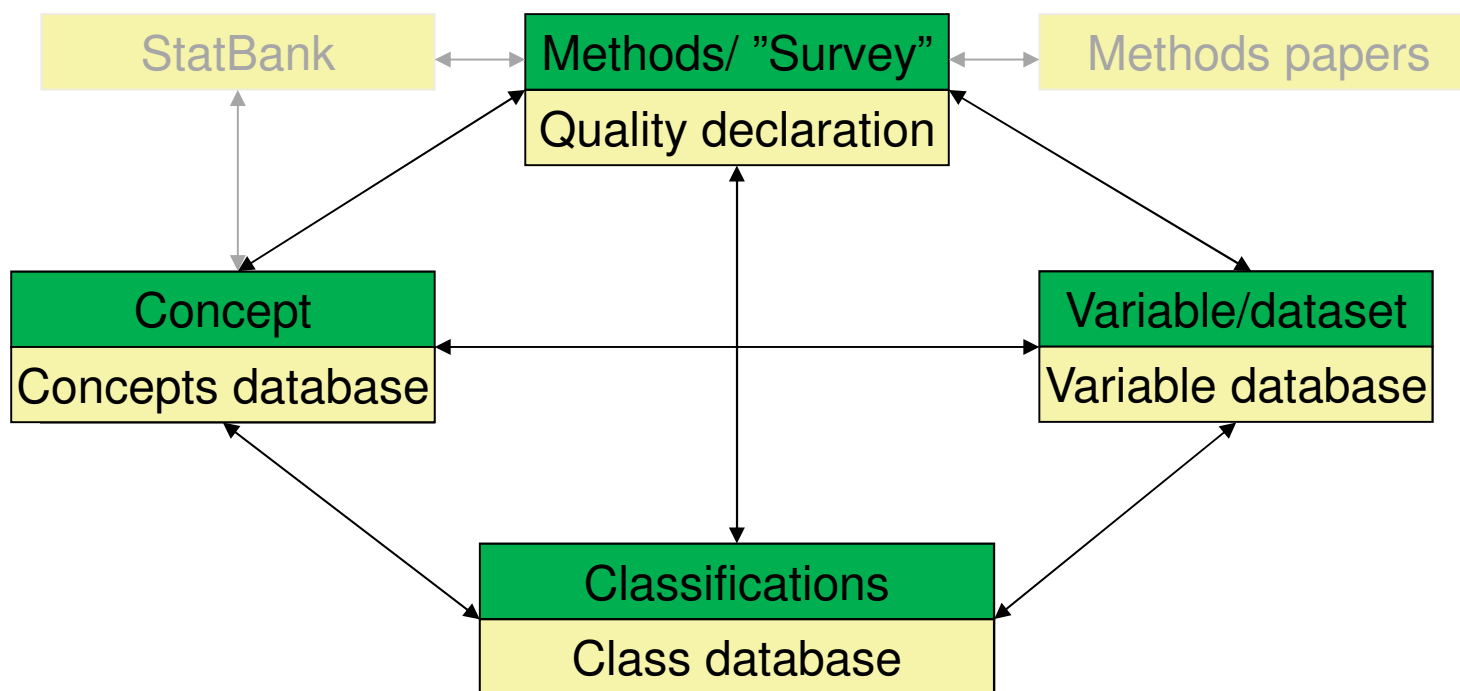# Handling complexity with metadata – what and how

What kind of metadata?

- Quality report (methodology, relevance, accuracy, comparability etc)

- Other metadata (concepts, variables and classifications)

How: processes and databases

- Processes to handle feedback and knowledge in relation to users (GSBPM)

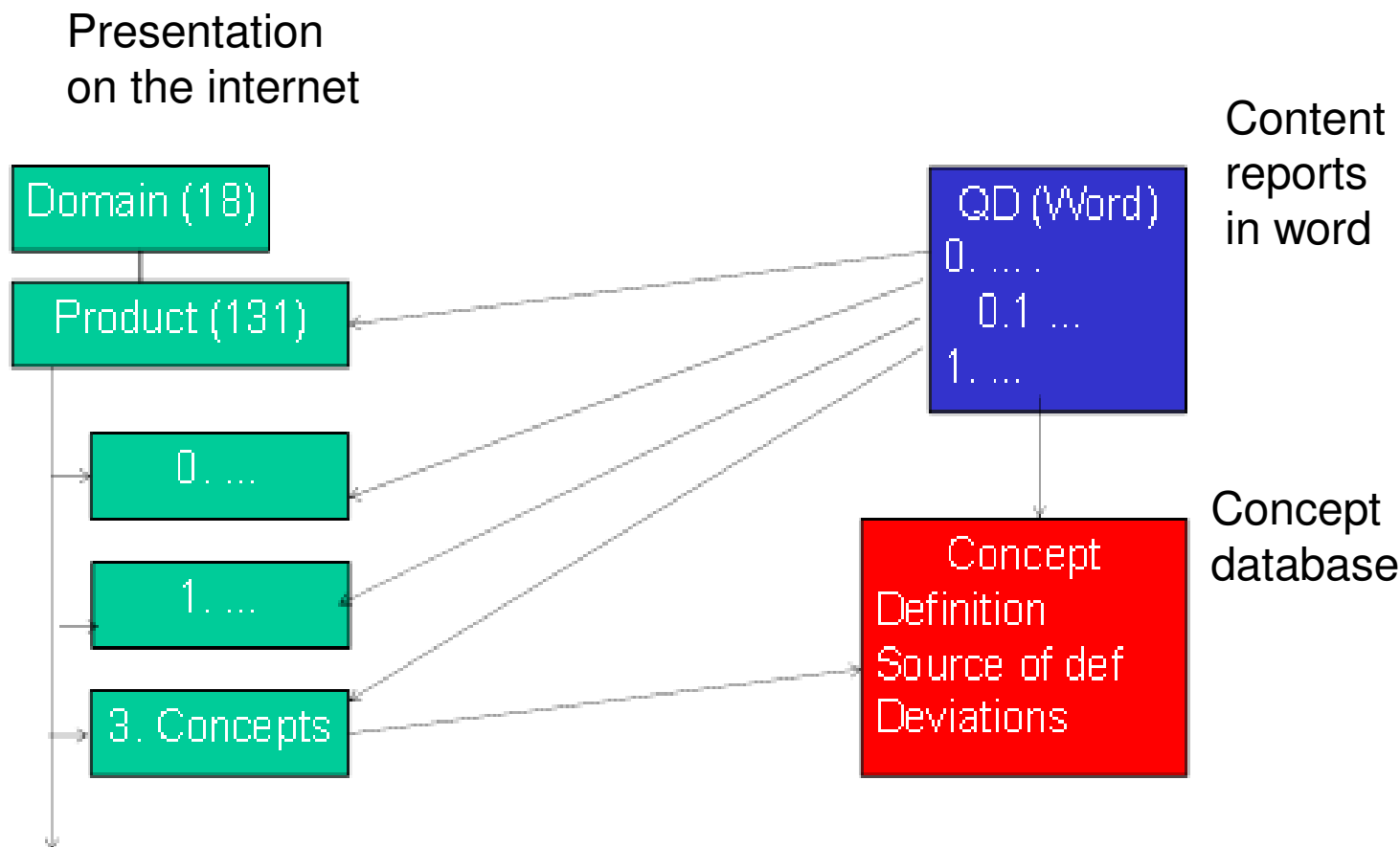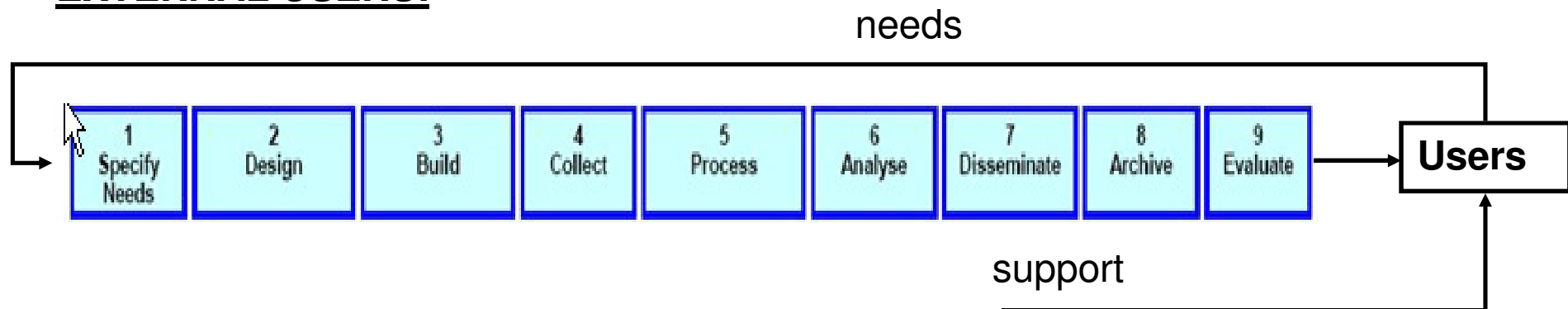- Databases with searchable integrated metadata directed towards user-groups

DANMARKS STATISTIK

Metadata with links

StatBank	Methods/ "Survey"	Methods papers
	Quality declaration	

Concept	Variable/dataset
Concepts database	Variable database

Classifications
Class database

How to integrate and publish content-report and other metadata

Presentation
on the internet

Content
reports
in word

Domain (18)

Product (131)

0. ...

1. ...

3. Concepts

QD (Word)
0. ... .
  0.1 ...
1. ...

Concept
database

Concept
Definition
Source of def
Deviations

DANMARKS
STATISTIK

How to : Business processes / users / metadata

EXTERNAL USERS:

needs

1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Archive	9 Evaluate

Users

support

INTERNAL USERS (support of proces 1-9)
a) metadata used in browse- and search- applications
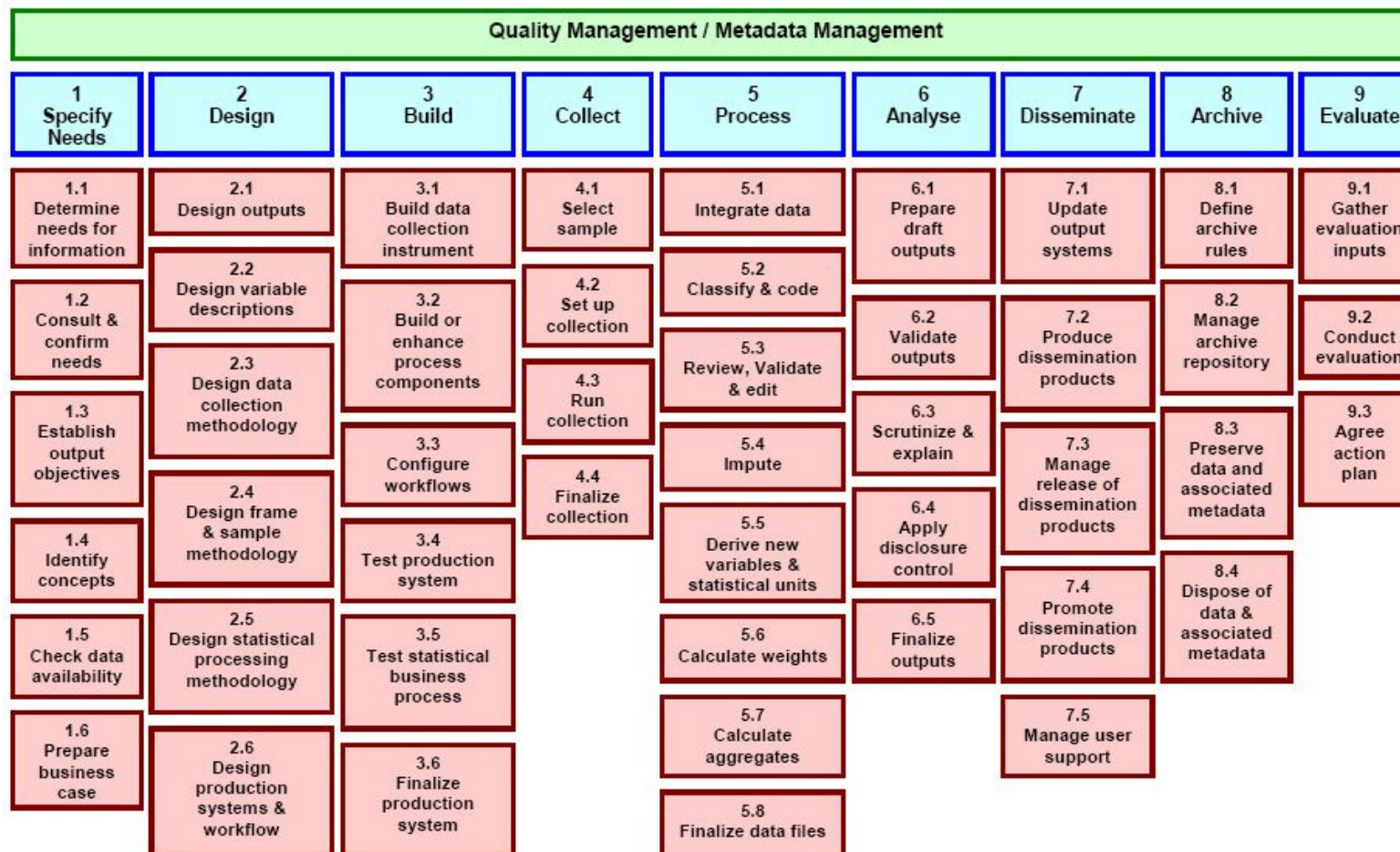b) metadata used to support IT (system-to-system interface)

DANMARKS STATISTIK

Why have a common business proces model

- **Common language: Generic description of the workflow in the production of official statistics**

- **Sharing of knowledge: GSBPM is an international standard used by many NSI'S**

- **Process orientation is the starting point for**

- Quality models (CoP, ISO EFQM ...)

- Project management models

- Metadata models (DDI, SDMX ETC)

- Models that handle feedback and knowledge in relation to user

Generic Statistical Business proces model

Quality Management / Metadata Management								
1 Specify Needs	**2 Design**	**3 Build**	**4 Collect**	**5 Process**	**6 Analyse**	**7 Disseminate**	**8 Archive**	**9 Evaluate**
1.1 Determine needs for information	2.1 Design outputs	3.1 Build data collection instrument	4.1 Select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Define archive rules	9.1 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 Manage archive repository	9.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design data collection methodology	3.3 Configure workflows	4.3 Run collection	5.3 Review, Validate & edit	6.3 Scrutinize & explain	7.3 Manage release of dissemination products	8.3 Preserve data and associated metadata	9.3 Agree action plan
1.4 Identify concepts	2.4 Design frame & sample methodology	3.4 Test production system	4.4 Finalize collection	5.4 Impute	6.4 Apply disclosure control	7.4 Promote dissemination products	8.4 Dispose of data & associated metadata	
1.5 Check data availability	2.5 Design statistical processing methodology	3.5 Test statistical business process		5.5 Derive new variables & statistical units	6.5 Finalize outputs	7.5 Manage user support		
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Finalize production system		5.6 Calculate weights				
				5.7 Calculate aggregates				
				5.8 Finalize data files				

DANMARKS STATISTIK

Standards - Neuchatel

- Common language for classifications

- Concepts (semantic level)

- No managament modul

Classification family

Classification

Classification version

Classification level

Classification item

>> Standards: SDMX

It consists of

- Information model for data and metadata
- Content oriented guidelines
- IT-architeture for data and metadata exchange
- Supporting tools

Advantages – as exchange standard

- Common language and understanding "2+2=5"
- It is global - BIS, ECB,Eurostat, IMF, OECD, UN joined forces

DANMARKS
STATISTIK

The SDMX-elements

>> Standards: DDI-standard

What is it?

Documentation standard, expressed in open XML standard

Many years of experience including use in NSI's

Advantages

Common language and understanding "2+2=5"

Integration of concepts, variables, classifications quality

Both for schema and register based statistics

Model currently used in Australia, New Zealand, Canada etc.(together with sdmx)

Many tools

DANMARKS STATISTIK

≫ DDI – types of metadata (WHAT)

Metadatatypes:

· Concepts ("terms")

· Studies ("surveys", "collections", "data sets", "samples", "censuses", "trials", "experiments", etc.)

· Variables ("data elements", "columns")

· Codes & categories ("classifications", "codelists")

· Universes ("populations", "samples")

· Data files ("data sets", "databases")

For surveys:

· Survey instruments ("questionnaire", "form")

· Questions ("observations")

· Responses

Proces-model and DDI-modules (HOW)
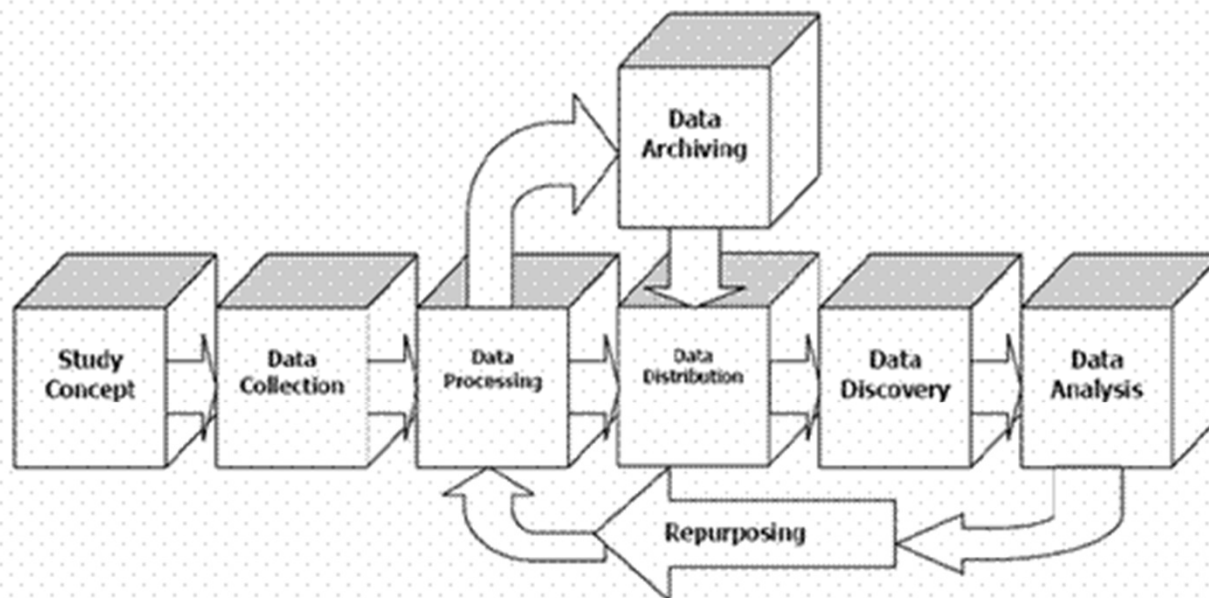
Figure: Combined Life Cycle Model
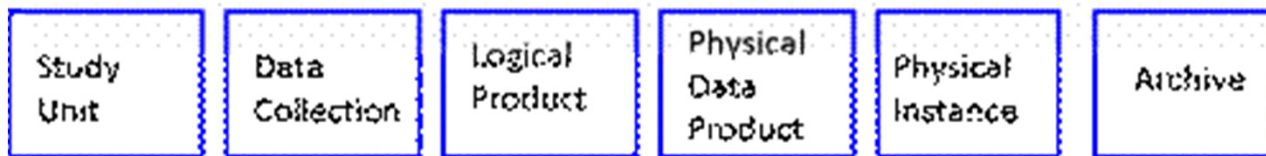
GSBPM – DDI AND SDMX
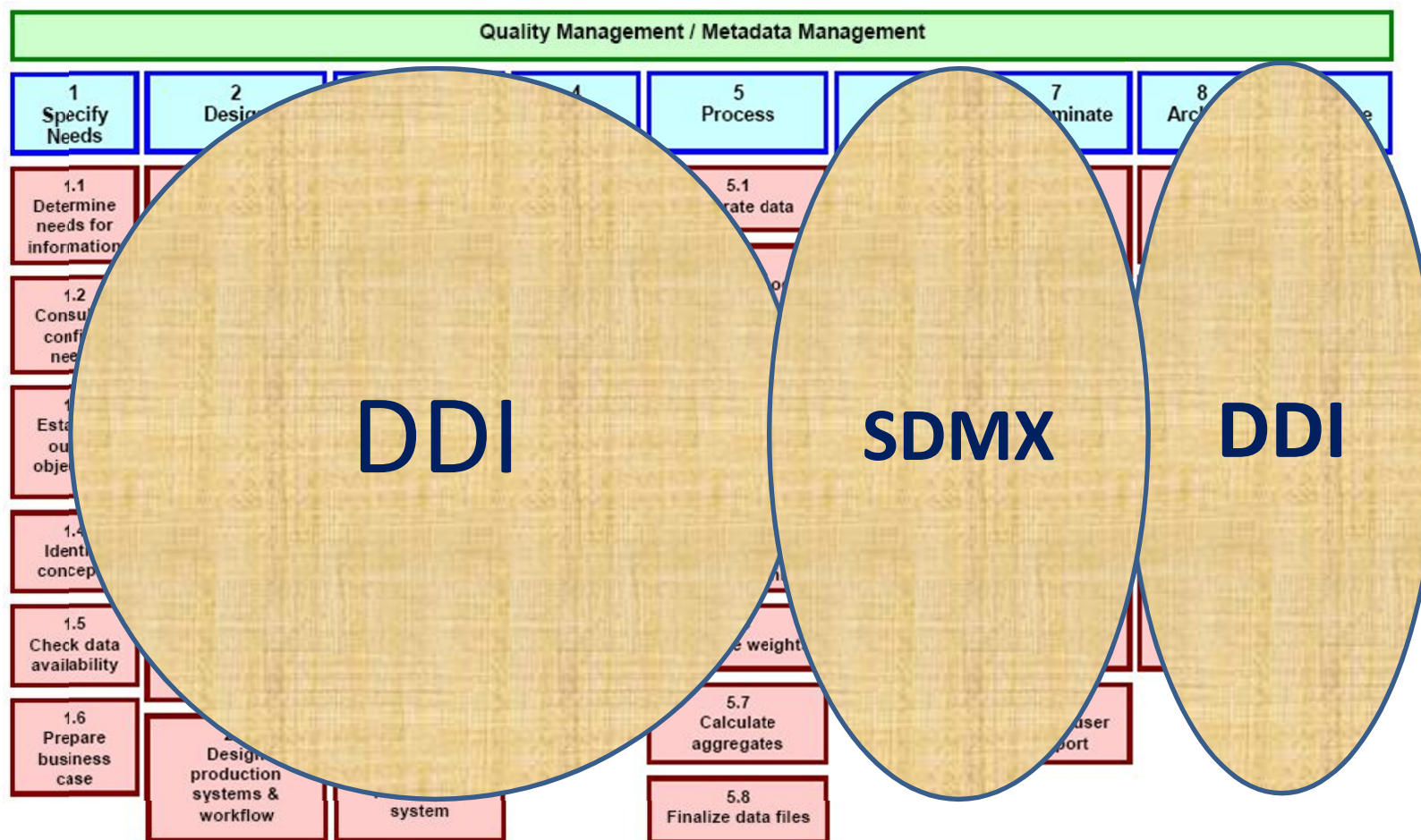
METADATA STRATEGY
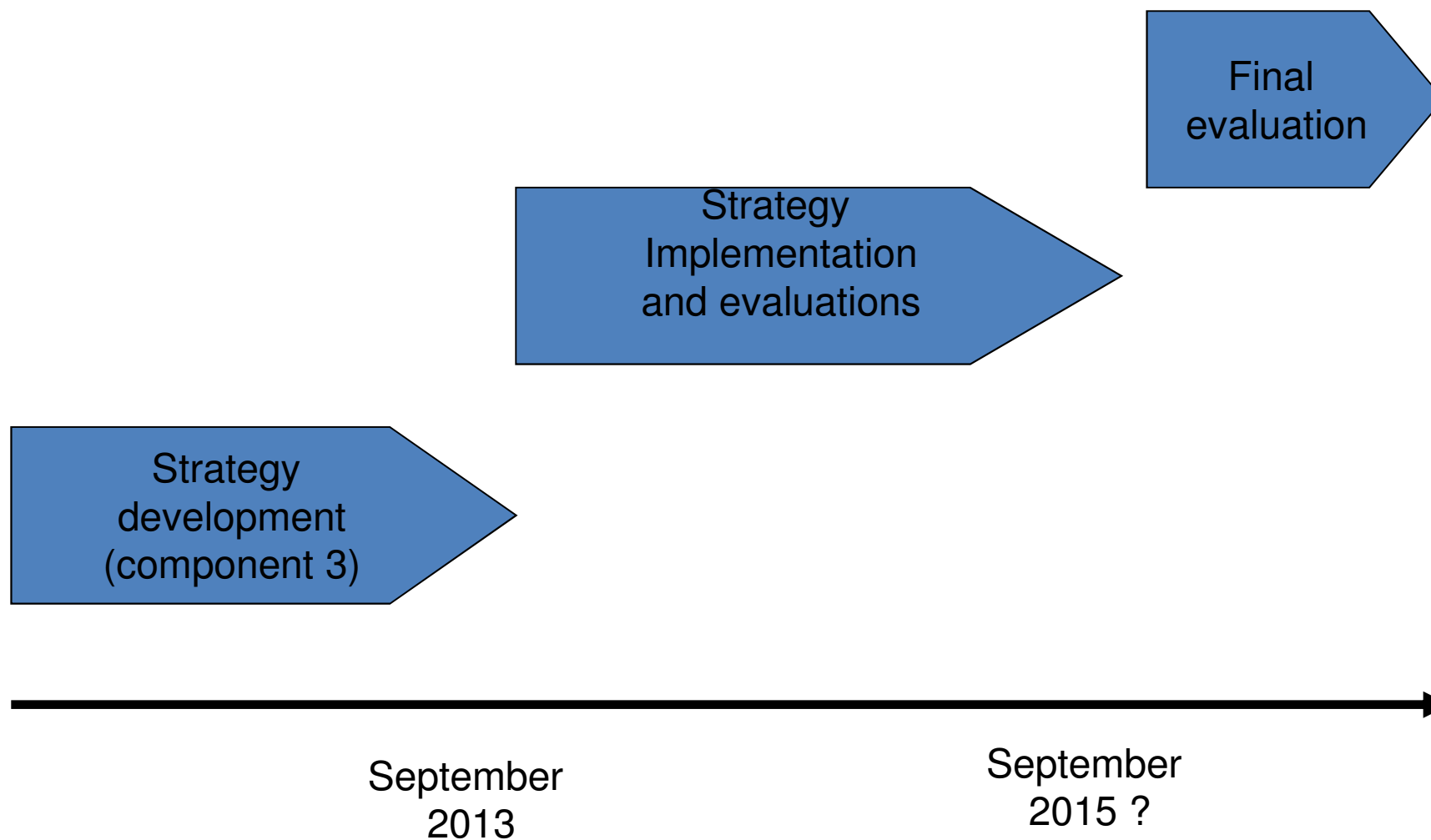
≫ **Strategy outline**

1. **Introduction, including situation today**
2. **Objectives and results**
3. **Tasks and time-table**
4. **Organisation**

Annex 1. Guidelines for production and use of metadata
- **Development processes**
- **Operation processes**
- **Metadata support proces**

Annex 2. IT-solution
1. **High level architecture**
2. **Standards**
3. **Applications and databases**
4. **Data flow**

DANMARKS STATISTIK

» Objectives and results

Objectives

- **Sustainable development**
- **Better communication between the three statistical institutes**
- **Better communication between providers and users of the statistics.**

Results

- **Business processes that ensure fulfilment of internal and external user need**
- **Databases with searchable integrated metadata directed towards user-groups**
- **Details: 1) Metadata and classifications have to be stored and updated in one place: the Agency. 2) RSIS and FIS will have web access to metadata for compilation and dissemination of statistics. 3) RSIS and FIS will report to the Agency and ensure that the same classifications etc are used in all three statistical organisations**

DANMARKS STATISTIK

Guidelines for processes (needs from both external and internal users)

NEEDS

| 1 Specify Needs | 2 Design | 3 Build | 4 Collect | 5 Process | 6 Analyse | 7 Disseminate | 8 Archive | 9 Evaluate | → **Users** |

Metadata supporting process
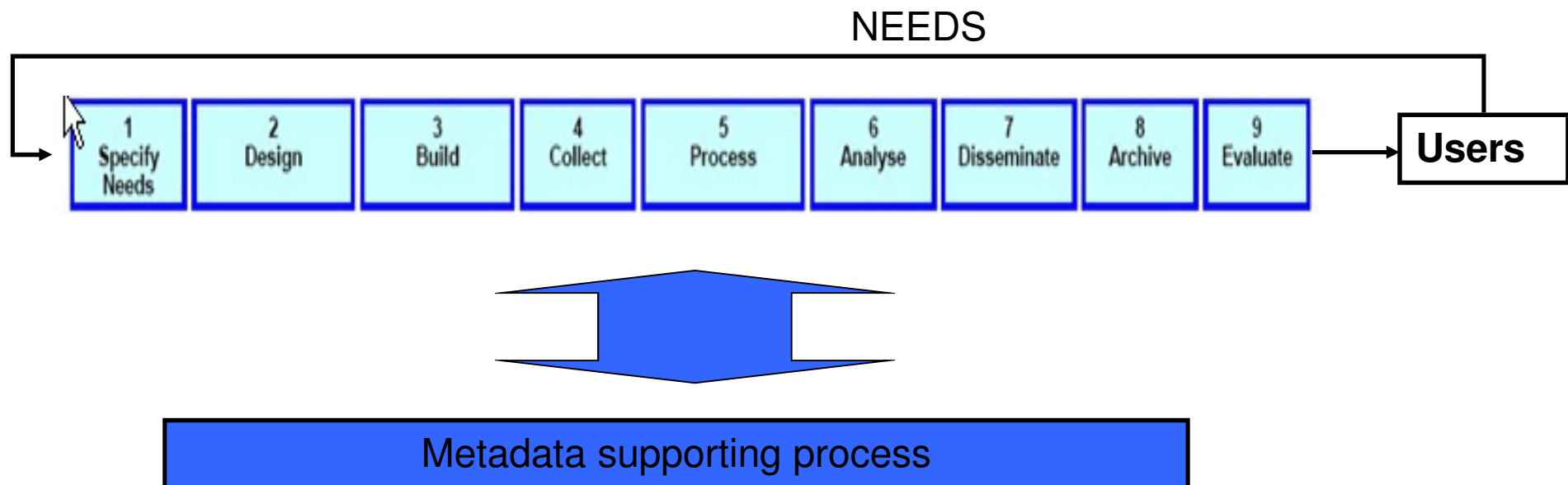
High level IT-architecture

PROCESSES THAT USES METADATA

1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Archive	9 Evaluate
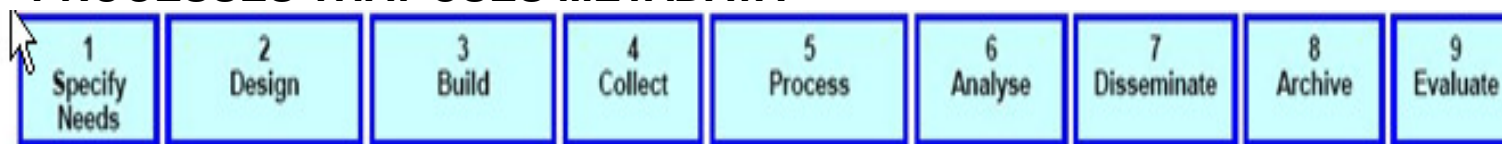
COMMON METADATA-APPLIKATIONS
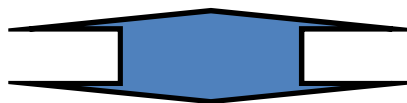
COMMON METADATABASE
Quality content, concepts
variables and classifications

Why have standards (xml) for metadata

- Sustainable development

- No 'lock in' to software

- Support integration / exchange

- 3 layer model / many tools

- "Division of labor" in the international statistical community

  – National Statistical Organisations should share the work on developing standards and tools

  – Standards tools should flexible and easy to implement

DANMARKS STATISTIK

Global strategic model
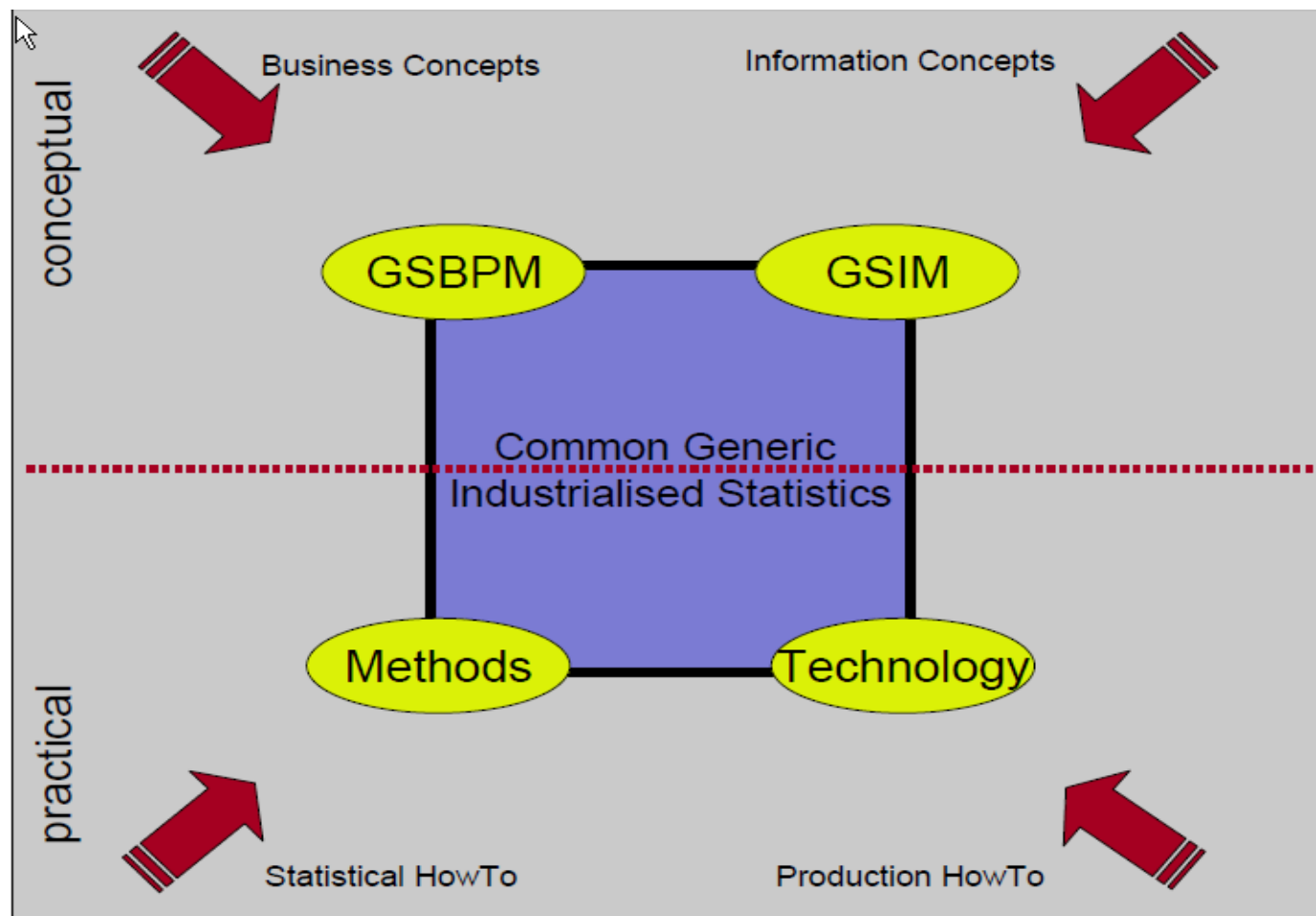
GSBPM and GSIM / SDMX and DDI

- *GSBPM (Generic Statistical Business Process Model) and GSIM (Generic Statistical Information Model) supplements each other*

Report from global GSIM working group

- *"Operationalizing GSIM highest priority … for example, a commonly agreed to representation in XML"*

- *"The workgroup identified SDMX and DDI-L (DDI-Lifecycle) as the key starting points in this regard"*

DANMARKS STATISTIK

Standards and tools: different models

- Examples
  - Sweden (metaplus variables and class. integrated)
  - Portugal (4 subsystem)
  - Australia and New Zealand (ddi and sdmx)
  - Denmark (ddi and quality integrated - pilot)
  - Neuchatel (implemented as part systems in Portugal, Sweden)

Sweden METAPLUS
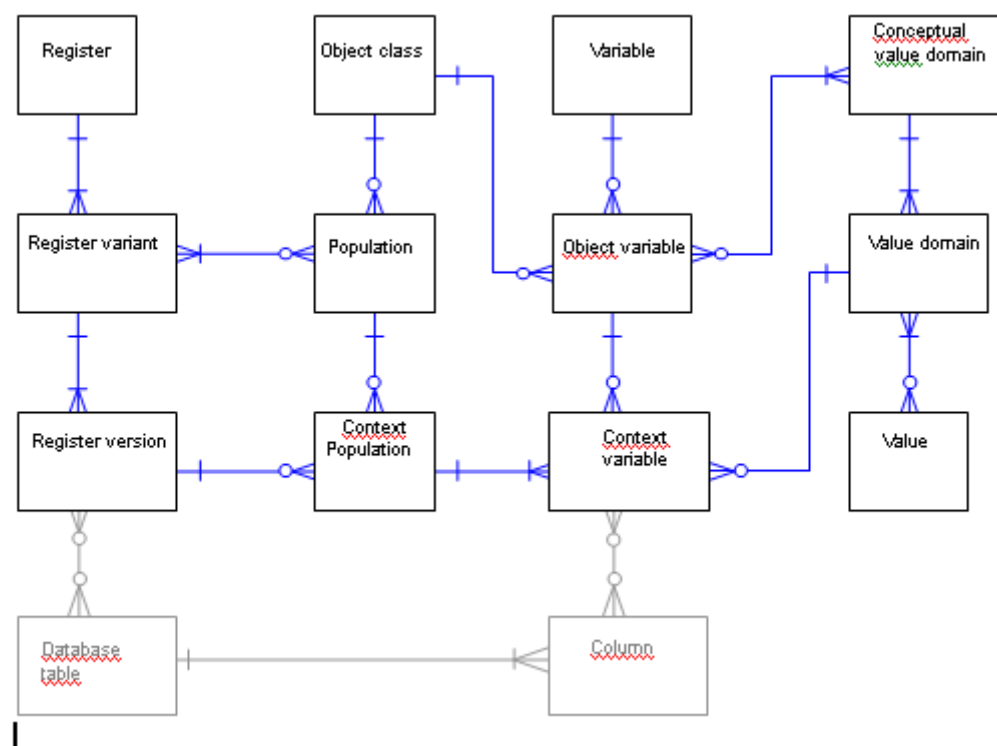
Portugal

- 4 integrated subsystem
  - Variable
  - Methological documents
  - Concepts
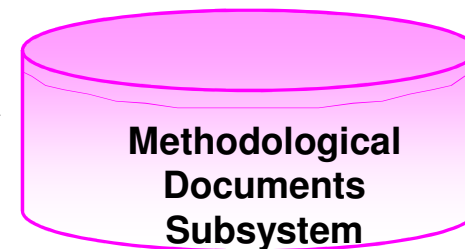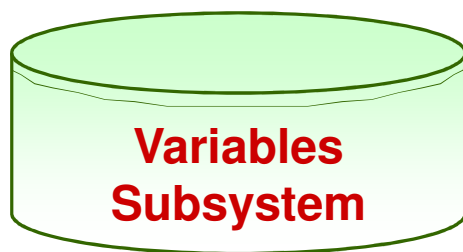 - Classification

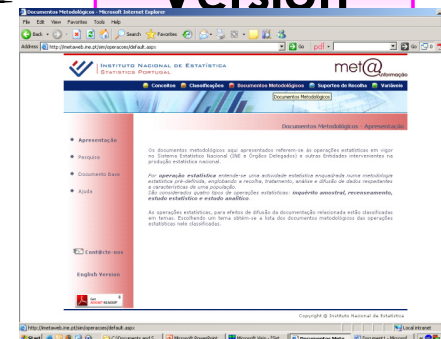# Relationship with other systems

Relationship with other systems

Relationship with other systems

Australia

Plan

- Focus in depth on metadata for a specific statistical business process
  - QBIS used as example
  - Design of ABS Transitional Model (and GSIM) recognises SDMX and DDI as valuable technical standards supporting implementation & interoperability

DANMARKS STATISTIK

Australia: Model (for Proof of Concept) of how vision might be actualised locally

Conceptual

Practical

GSBPM

GSIM Common Reference Model

SEMANTIC REFERENCE MODEL

DDI

SDMX

Simulated Forms Builder

Simulated Registry Search

MRR

SWM

Harmonised Methods

Harmonised Technology

Unresolved discussion in ABS : Where would CORA + CORE constructs be positioned?
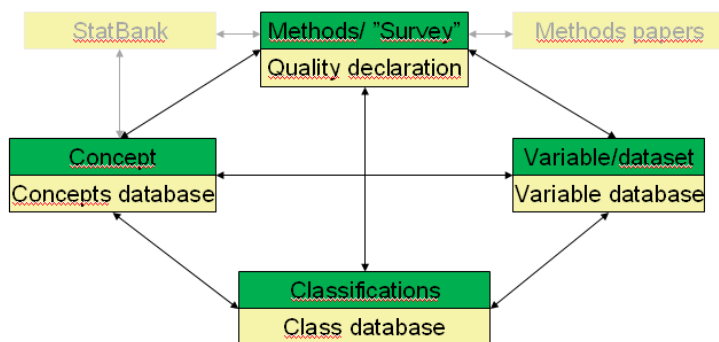
DANMARKS STATISTIK

Issues recorded – what to recommend?

- **Use of standards and tools (DDI, Neuchatel, Metaplus etc) -**
- **Harmonisation of data and systems / metadata (BHAS, RSIS, FIS)**

1. **Processes differ depending on people involved**
2. **All three organization have their own metadata**
3. **Differences in classifications / codelist**
4. **Differences in legislation**
5. **Old and new coding systems**
6. **Data and metadata should go together – how to**
7. **Data in SQL already – how does implementing metadata standard affect our work**
8. **Human resource**
9. **Today excel / word etc. How to arrange data /each survey is separate**
10. **GSBPM how to**

Issues recorded

- 'The qualiy corner'
  1. Framework developed – relation til metadata
  2. purpose is much wider than contributions to metadata
  3. usually published in end of production
  4. What included
  5. What to include in metadata – methodological report and quality report. Depends on user need

StatBank ↔ Methods/ "Survey" / Quality declaration ↔ Methods papers

Concept / Concepts database ↔ Variable/dataset / Variable database

Classifications / Class database

DANMARKS STATISTIK

≫ Plan for work until next mission

- DDI-pilot on SBS

- Draft to plan for improving codelist – tasks, timetable and participiants

- Prepare decision and adoption of GSBPM
  - Translation ready next mission

- Overview of business processes

DANMARKS STATISTIK

Links

- [Metis wiki - GSBPM, GSIM etc](#)
- [Eurostat - quality reporting](#)
- [SDMX](#)
- [DDI alliance](#)

Examples
- [Sweden – Metaplus](#)
- [Portugal - 4 subsystems](#)
- [Denmark - declarations of content](#)

DANMARKS STATISTIK

The End!