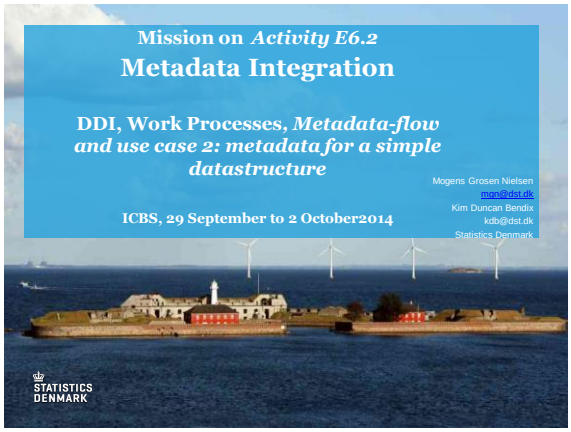


Agenda and credits

Agenda

- Introduction to DDI (and SDMX reference metadata)
- Work processes with focus on metadata mapped to GSBPM
- Metadata- and data-flow mapped to GSBPM
- Introduction to use-cases
- Use case 2: Demonstrate how Colectica can be used to build metadata for a simple data-structure (survey/admin. micro-data)
  - Introduction
  - Demo

Credits: Power Points prepared by Gesis, Leibnitz, Bryan Fitzpatrick and Colectica reused or used for inpiration.

DDI: Data Documentation Initiative

What is it?

Documentation standard, expressed in open XML standard

Many years of experience including use in NSI's

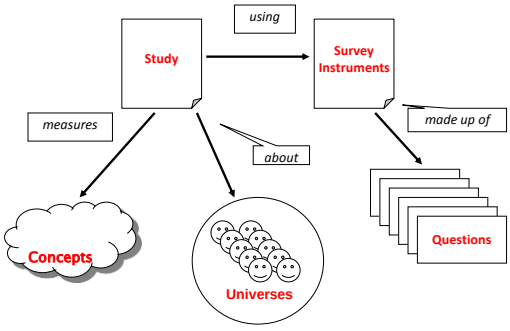
Advantages

Common language and understanding
Integration of concepts, variables, classifications quality

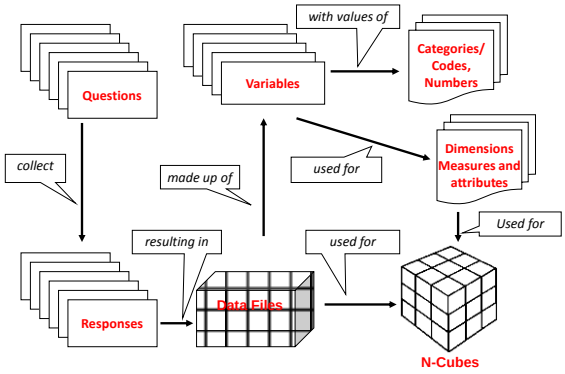
Both for schema and register based statistics
Model currently used in Australia, New Zealand, Canada etc.(together with SDMX)
Tools avaiable

STATISTICS DENMARK

Statistics and DDI in 60 seconds



Statistics and DDI in 60 seconds



Types of metadata – DDI and SDMX

Metadatatypes DDI

- 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")
- N-Cubes ("aggregated data")
- Survey instruments ("questionnaire", "form")
- Questions ("observations")
- Responses

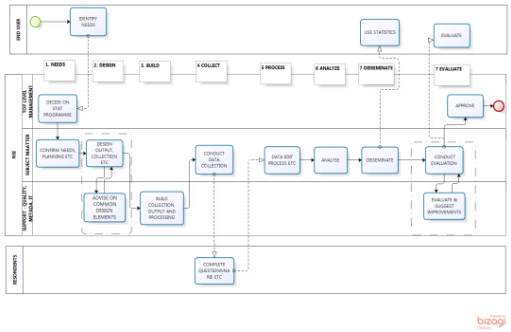
For quality reporting

- SDMX reference metadata (SIMS a general structure. ESMS and ESQRS for reporting)

STATISTICS DENMARK

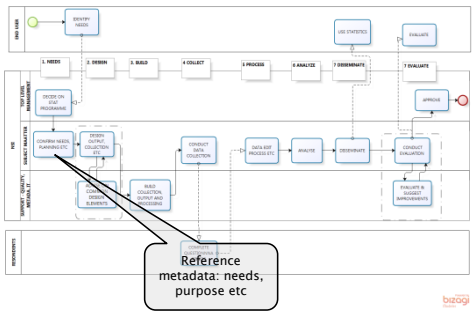
STATISTICS DENMARK

GSBPM and work-processes overall



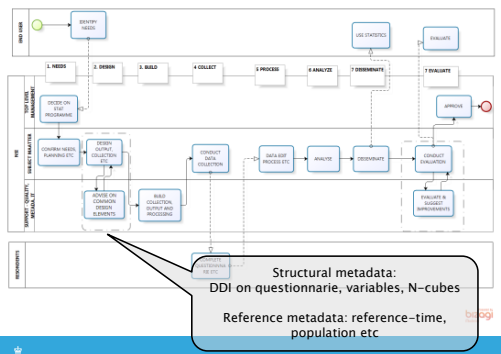
STATISTICS DENMARK 7

GSBPM and work-processes overall



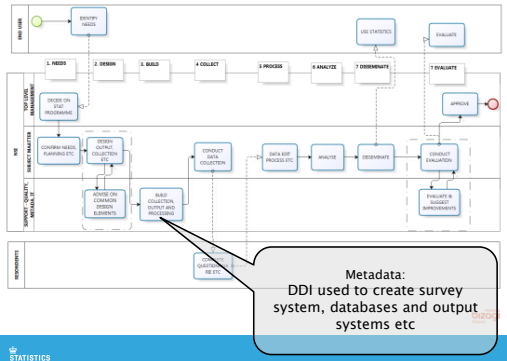
STATISTICS DENMARK 8

GSBPM and work-processes overall

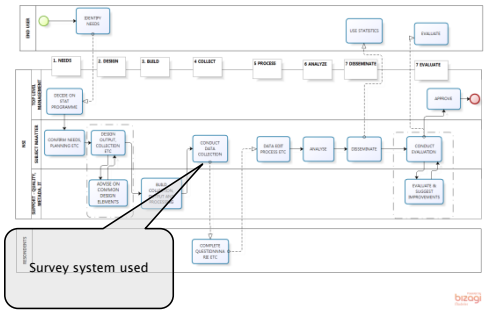


STATISTICS DENMARK 9

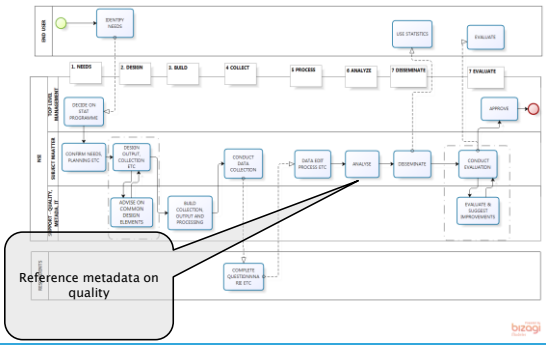
GSBPM and work-processes overall



GSBPM and work-processes overall

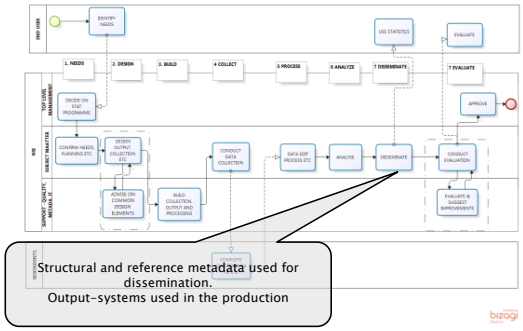


GSBPM and work-processes overall

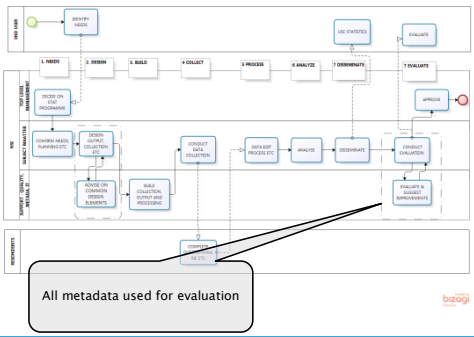


STATISTICS DENMARK

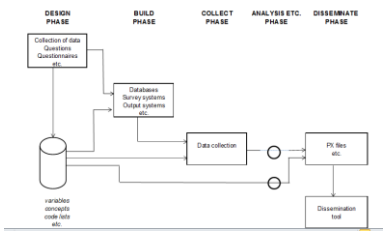
GSBPM and work-processes overall



GSBPM and work-processes overall



Metadata and data-flow



Use cases - introduction

- Use case 1: Demonstrate how Colectica can be used to build metadata for a simple questionnaire with reuse of code-lists and variables
- Use case 2: Demonstrate how Colectica can be used to build metadata for a simple data-structure (survey/admin. micro-data)
- Use case 3: Demonstrate how Colectica can be used to build metadata and data for an aggregated dataset using N-cube, with reuse of variables and codelist from micro-level.
- Use case 4: Demonstrate how Colectica can be used support work on quality-declarations with reuse of statistical concepts and with integration into GSBPM

STATISTICS DENMARK 16

- Use case 1: Demonstrate how Colectica can be used to build metadata for a simple questionnaire with reuse of code-lists and variables

STATISTICS DENMARK 17

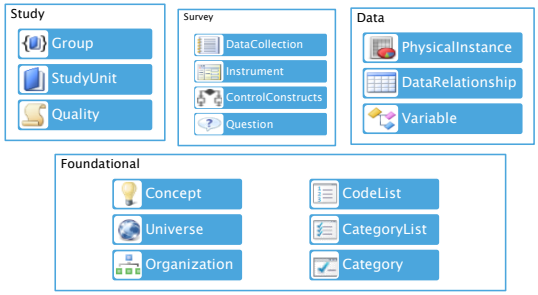
Metadata is in a DDI Instance file (other structures possible)

- DDI instance
  - Group
  - Concepts
  - Code Schemes and Category Schemes
  - Questions
  - Control Constructs
  - Instruments
  - Variables
  - Logical Records
  - N-cubes
    - Group (statistic)
    - Name, purpose etc.
    - Study Unit (wave)
    - Physical Instance

STATISTICS DENMARK 18

STATISTICS DENMARK

DDI in Colectica - a Glance



Classifications

- DDI holds Classifications as linked Code Schemes and Category Schemes
  - a Category Scheme is a list of Categories
    - flat list of multi-lingual names and descriptions
    - eg, Country names, Occupation names, etc
  - a Code Schemes selects Categories from Category Schemes, assigns a Code (not multi-lingual), and may specify a hierarchy
    - a Code Scheme may select Categories from multiple Category Schemes
    - multiple Code Schemes may select the same Categories

STATISTICS DENMARK

Code Schemes and Category Schemes

- Used for
  - Classifications
    - a Classification is a Code Scheme
  - Controlled Vocabularies
    - lists of standardised terms
    - defined by DDI, an organisation, a local area

STATISTICS DENMARK

STATISTICS DENMARK

Code Schemes and Category Schemes

- Used for
  - Response Domains for Questions
    - Code Domains and Category Domains
    - Category domains are not much use in a multi-lingual environment
      - Categories have different names in different languages
        - with no unique handle except a meaningless Id
  - Representations for Variables
    - Code Representation

STATISTICS DENMARK

Variables

- A Variable is a container that will hold a data value
  - has a Name and Description (both multi-lingual)
  - can be linked to a single Concept
    - to indicate what the data represents
  - can be linked to multiple Questions
    - to indicate where the data comes might come from
  - can have a Representation
    - Code, Date/Time, Numeric, Text
      - with constraints on values
  - can identify a Response Unit and an Analysis Unit
    - a population that it can apply to

STATISTICS DENMARK

Logical Record

- A Logical record consists of a sequence of Variables
  - groups data values for a purpose
  - data from a questionnaire goes into one or more Logical Records
  - Logical Records can be linked
  - eg, Households and Persons
  - Logical Records are independent of any storage or stored format

STATISTICS DENMARK

STATISTICS DENMARK

Record Layouts and Physical Structures

- Map a Logical record to a physical record and an actual stored file format
- Can support a very wide range of structures and storage formats
  - CSV, Binary file, XML, database
  - multiple record types, linkages of many kinds
- POC does not actually use this
  - Simple CSV file maps directly from Logical record

STATISTICS DENMARK

Physical Instance

- Holds information about actual data sets produced
  - links to Physical Structures, Record Layouts, and Logical records
  - provides a central management of data from a collection
  - POC uses Physical Instance to manage data
    - POC 2.3.3 builds on this POC to show how to use SDMX and DDI metadata together
      - produces tables from SDMX DSD using data collected with DDI
      - uses the Physical Instance information to find the datasets

STATISTICS DENMARK

DEMO

STATISTICS DENMARK 27

STATISTICS DENMARK

STATISTICS DENMARK