Mission on

Activity 3.7: Metadata strategy – I

DoS, 24-29 August 2014

Mogens Grosen Nielsen, Statistics Denmark, mgn@dst.dk
Flemming Dannevang, Statistics Denmark, fda@dst.dk

DANMARKS STATISTIK

Agenda

1. Objectives of mission and methodology
2. Input from DoS part 1
3. Metadata - introduction, definition and functions, principles and users
4. Standards
5. Towards a strategy for the development of a metadata system for DoS

1. Objectives of mission and methodology

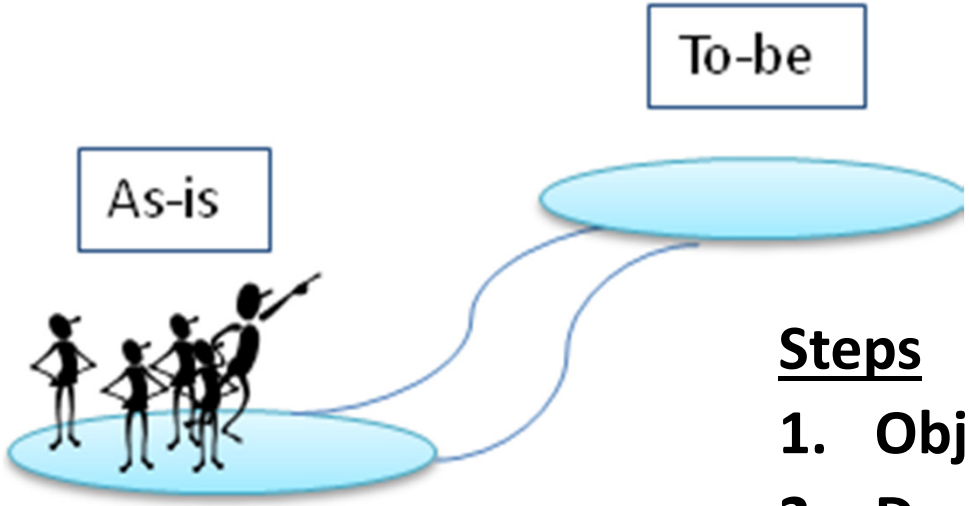
Objectives of the mission: Purpose (ToR)

- Assessment on current use of metadata systems incl. IT
- Presentation of different metadata systems
- Discussion on options for Jordan
- Discussion of the content of a metadata strategy

Output from the mission (ToR)

- Knowledge gained on current metadata system
- Recommendations prepared for a metadata strategy based on international standards
- Transfer of the Danish and in general the European Union, experience in metadata systems
- To decide upon the preferred timing of the next activity (3.8, Suggestion 2. November – 6. November 2014?)
- A lining up of work programme for the next activity (3.8)

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

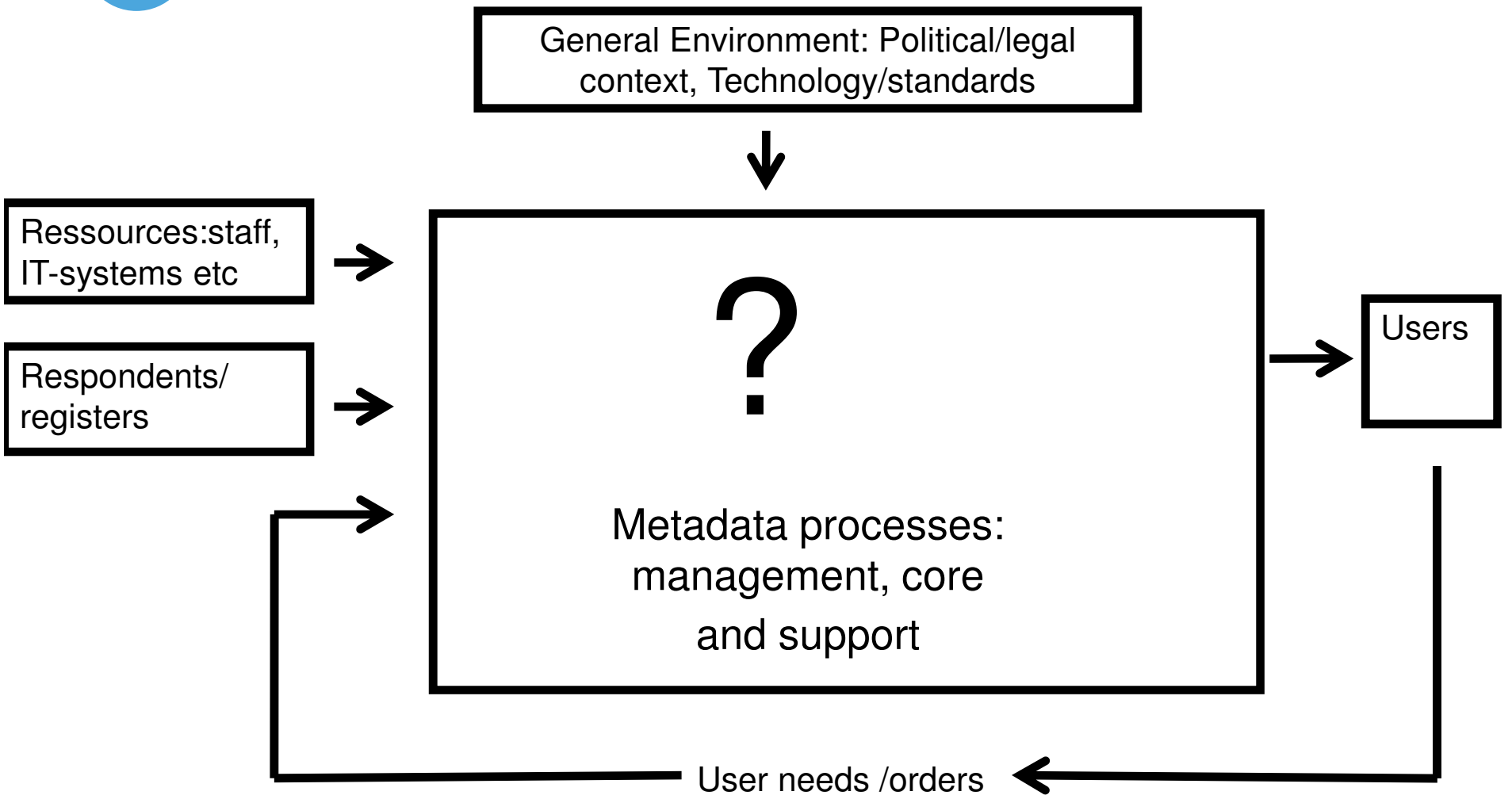


Steps

1. **Objectives for ‘to-be’**
2. **Describe/analyze existing processes**
3. **Decide on prioritization of changes**
4. **Implement changes**

- **Iterate on 1-4**

Business proces perspective.



3 main elements in the analysis

A: The situation today "as-is"

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

B: Objectives / results "to-be"

Objectives:

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

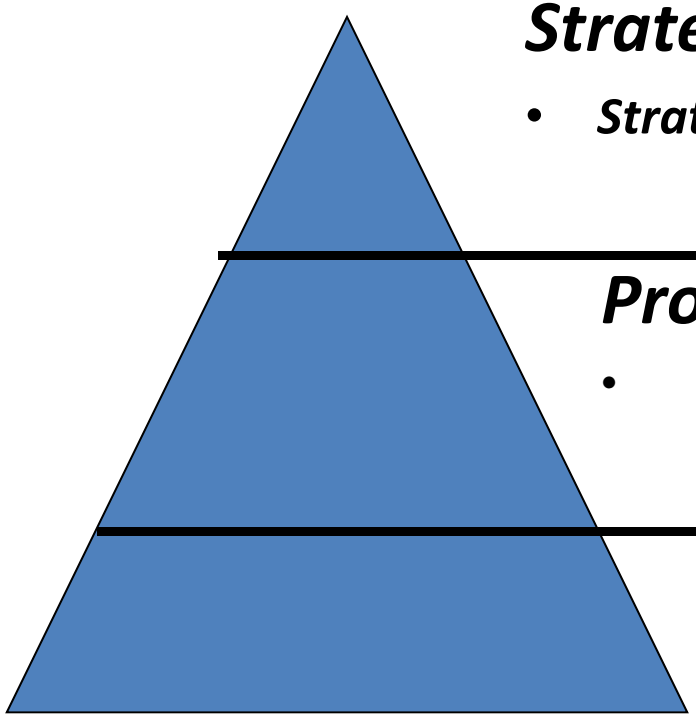
Functions the SMS must support

- See outline on strategy

C: How to get from A to B

Issues and recommendations

Issues at three levels



Strategic issues

- *Strategy, short and long term objectives etc*

Process issues

- *Management-, core and support processes (development and operation)*

Technical issues

- *Standards, IT-architecture etc*

2. Input from DoS part 1

- Relation to other projects quality etc
- As-is: situation today
- To-be: objectives etc.

From assesment mission

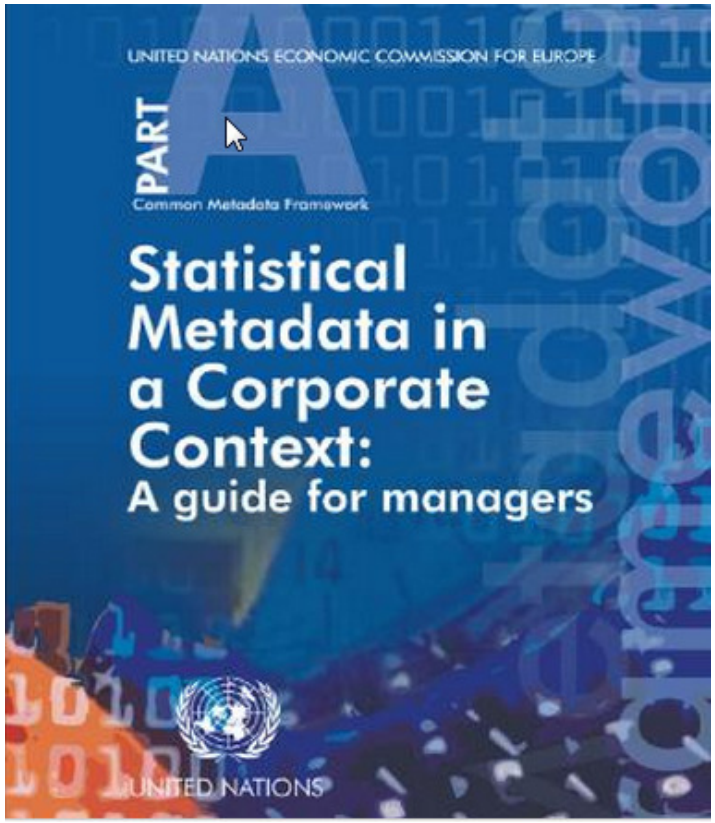
1. Parrallel projects focusing on different aspects of metadata
2. No clear direction
3. No integration with the dissemination platform
4. Metadata inside local applications
5. NADA – an initiative on a common system
6. Different perceptions about what metadata is
7. Different perceptions on which metadata is needed

3. Metadata - introduction, definition and functions

"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

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.

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):

Metadata – definitions #2

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."***

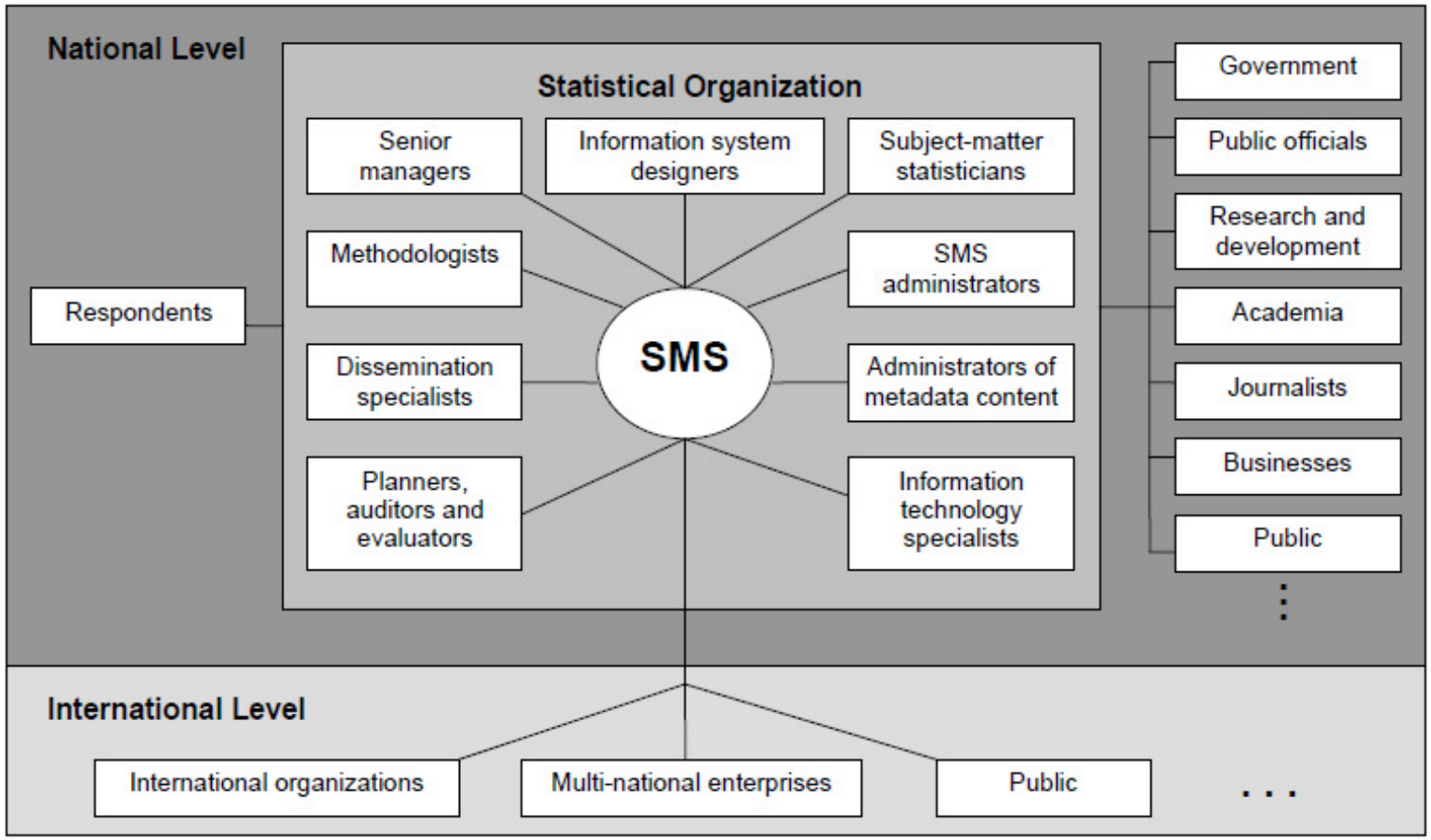
Metadata – definitions #3

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 users



Core principles

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.

SMS functions 1-5

1. Planning, designing, implementing and evaluating statistical production processes.
2. Managing, unifying and standardizing workflows and processes.
3. Documenting data collection, storage, evaluation and dissemination.
4. Managing methodological activities, standardizing and documenting concept definitions and classifications.
5. Managing communication with end-users of statistical outputs and gathering of user feedback.

SMS functions 6-10

6. Improving the quality of statistical data and transparency of methodologies.

7. Managing statistical data sources and cooperation with respondents.

8. Improving discovery and exchange of data between the statistical organization and its users.

9. Improving integration of statistical information systems with other national information systems.

10. Disseminating statistical information to end users – including assistance for post-processing of statistical data

SMS functions 11-15

11. Improving integration between national and international organizations, including harmonization of content

12. Knowledge base on the processes of statistical information systems, to share knowledge among staff

13. Improving administration of statistical information systems

14. Facilitating the evaluation of costs and revenues for the statistical organization

15. Unifying statistical terminology as a vehicle for better communication and understanding between managers, designers, subject-matter statisticians, methodologists, respondents and users of statistical information systems.

Objectives and 12 selected functions (prepared at the mission)

A. Cost efficient production using international standards

- The statistical metadata system must support
- Planning, designing, implementing and evaluating statistical production processes including
- Standardizing of workflows and processes.
- Documenting data collection, storage, evaluation and dissemination.
- Knowledge base on the processes of statistical information systems, to share knowledge among staff
- Improving administration of statistical information systems

Objectives and 12 selected functions (prepared at the mission)

B. Improvement of quality

- Improving the quality of statistical data and transparency of methodologies.
- Managing methodological activities, standardizing and documenting concept definitions and classifications.

Objectives and 12 selected functions (prepared at the mission)

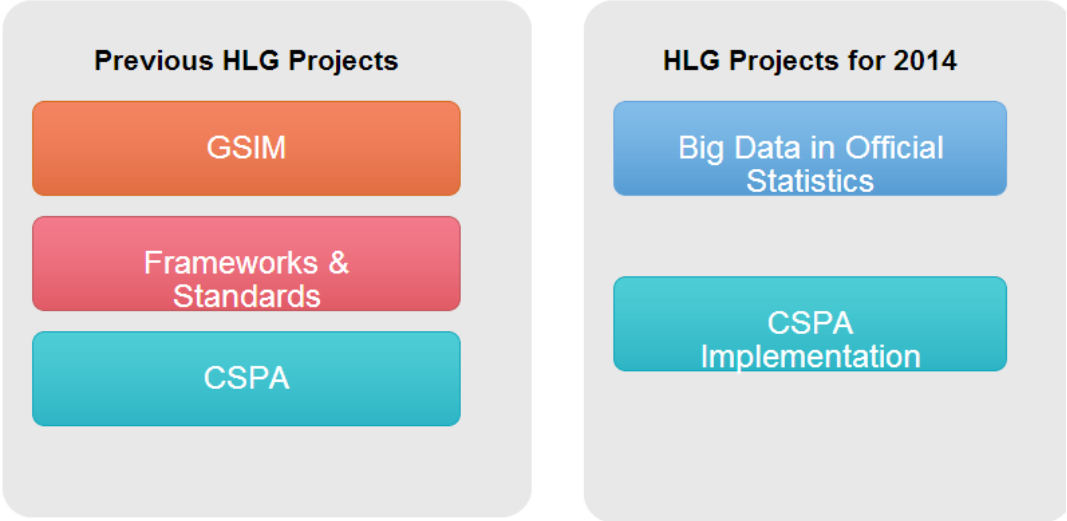
C. Fullfilment of user needs for documentation

- Managing communication with end-users of statistical outputs and gathering of user feedback.
- Improving discovery and exchange of data between the statistical organization and its users
- Improving integration of statistical information systems with other national information systems and
- Disseminating statistical information to end users – including assistance for post-processing of statistical data and
- Improving integration between national and international organizations, including harmonization of content

4. Standards

High Level Group for the Modernisation of Statistical Production and Services

Set up by the Bureau of the Conference of European Statisticians in 2010

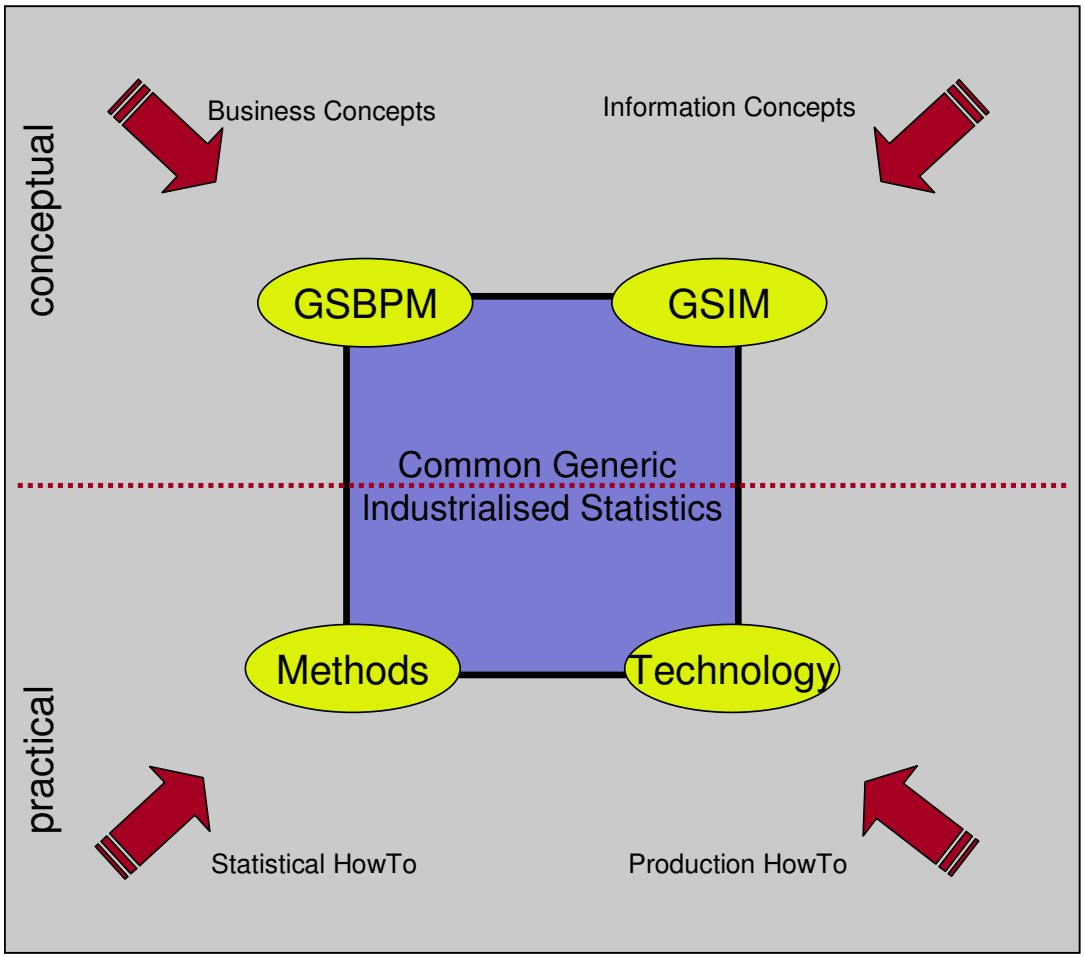


Tasks: Oversee and coordinate international work relating to statistical modernisation.

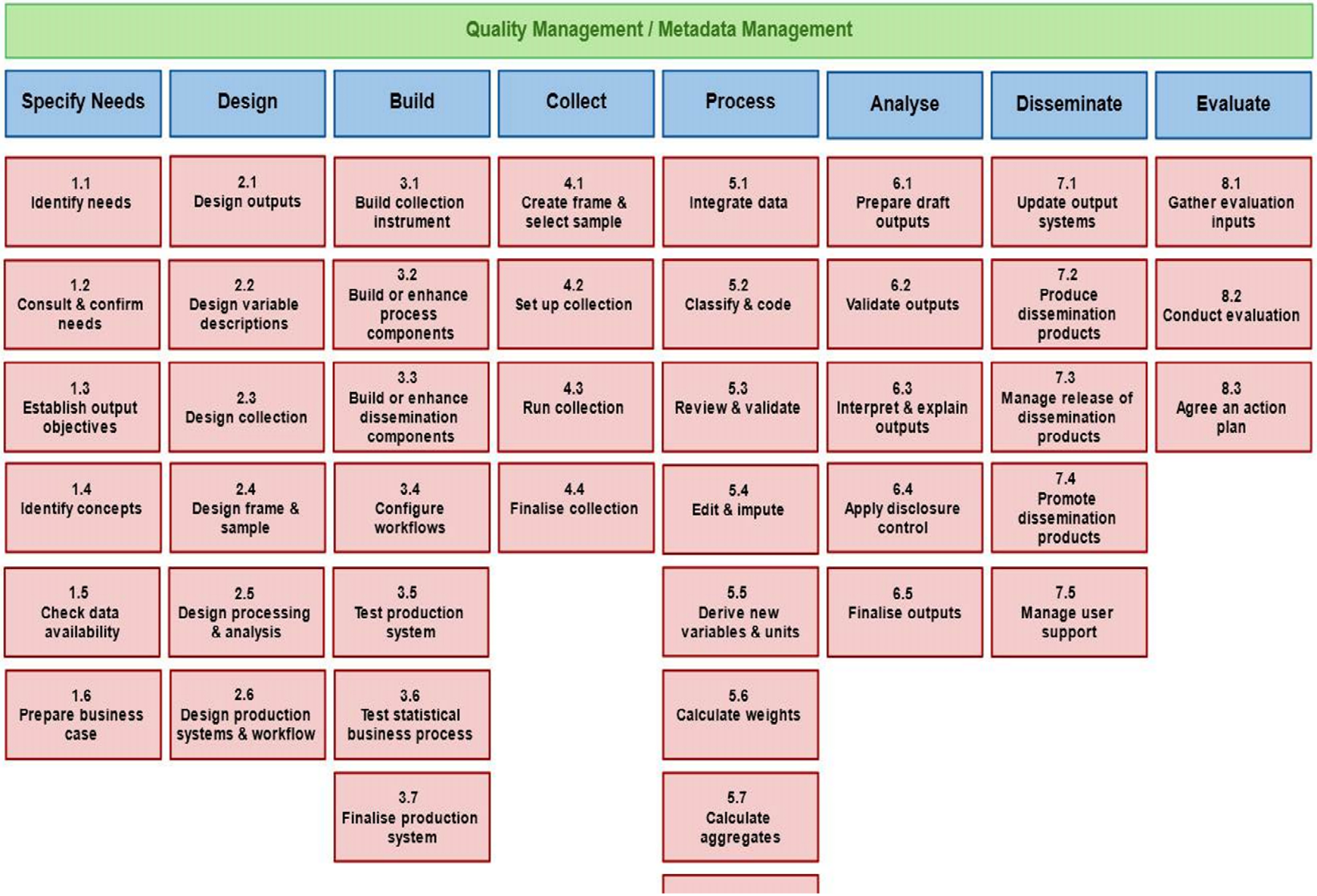
GSIM: Generic Statistical Information Model

CSPA: Common Statistical Production Architecture

Business concepts and informations concepts



The Generic Statistical Process model



Why have a common business proces model

- **Common (generic) description of the workflow in the production of official statistics**
- **International standard - used by many NSI'S**
- **Process orientation is the starting point for**
- Quality models (Code of Practice, Quality Assurance Framework etc)
- Metadata models (DDI, SDMX ETC)
- Processes that handle feedback and knowledge in relation to user
- Project management models

GSIM is complementary to GSBPM

- Things that flow between GSBPM sub-processes
- Things that drive and integrate sub-processes

So what is GSIM?

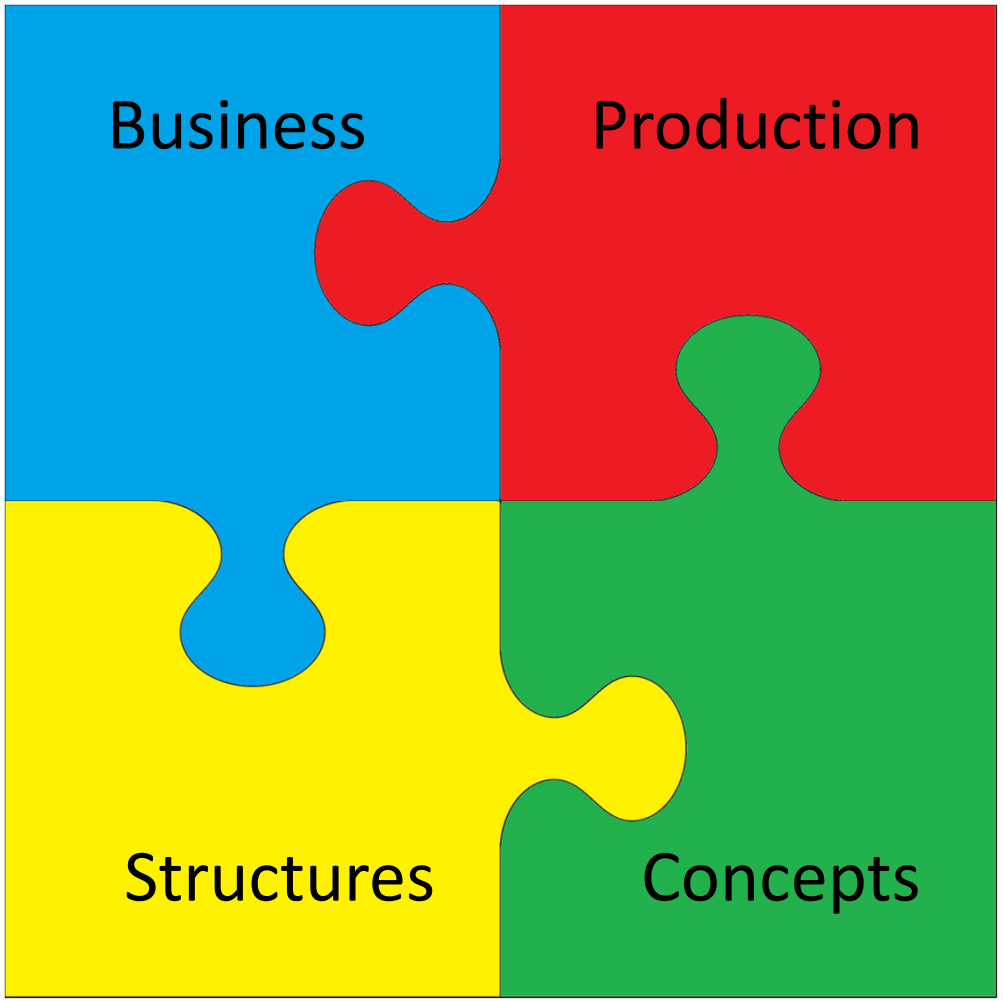
A reference framework of information objects:

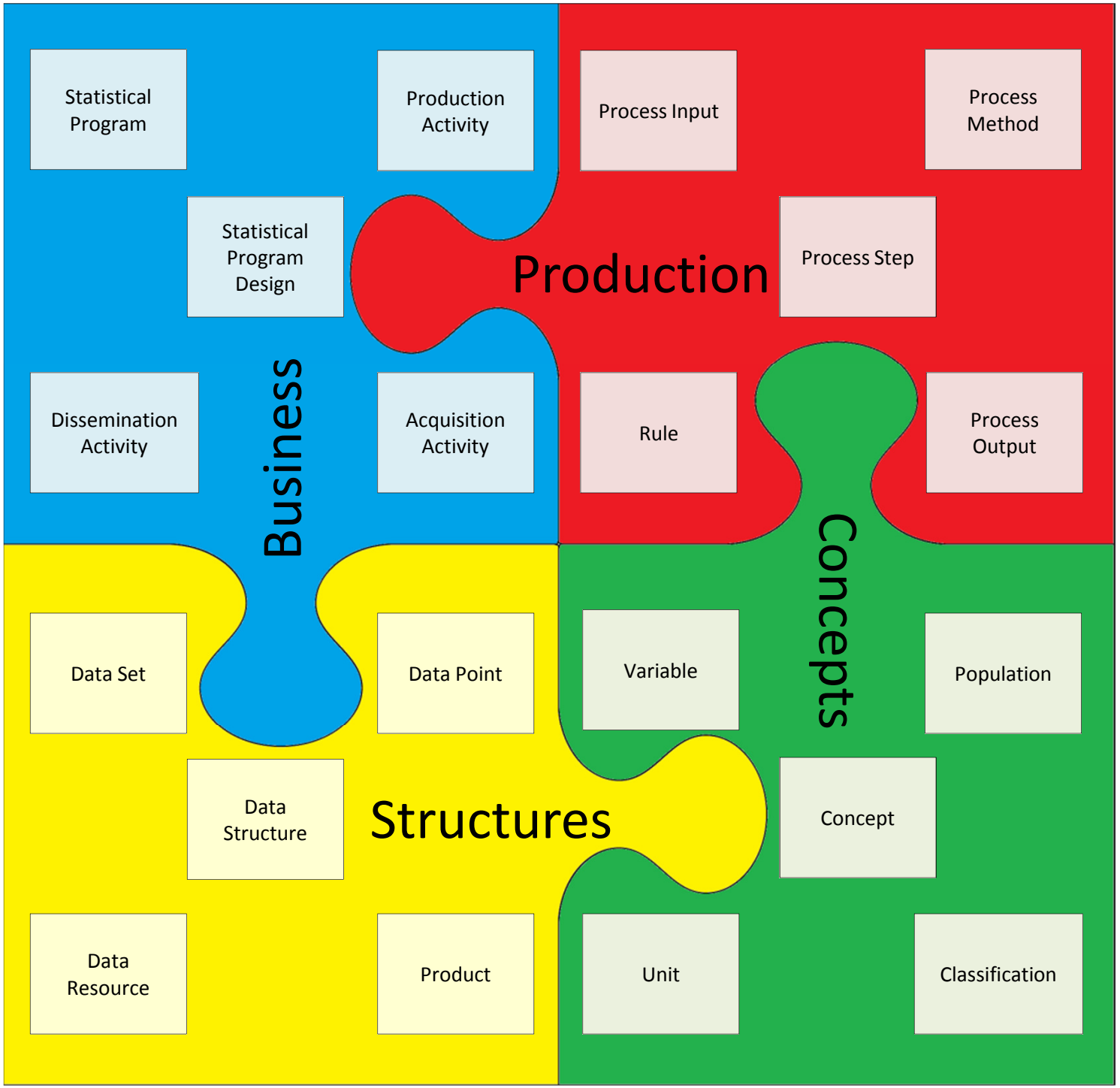
Definitions

Attributes

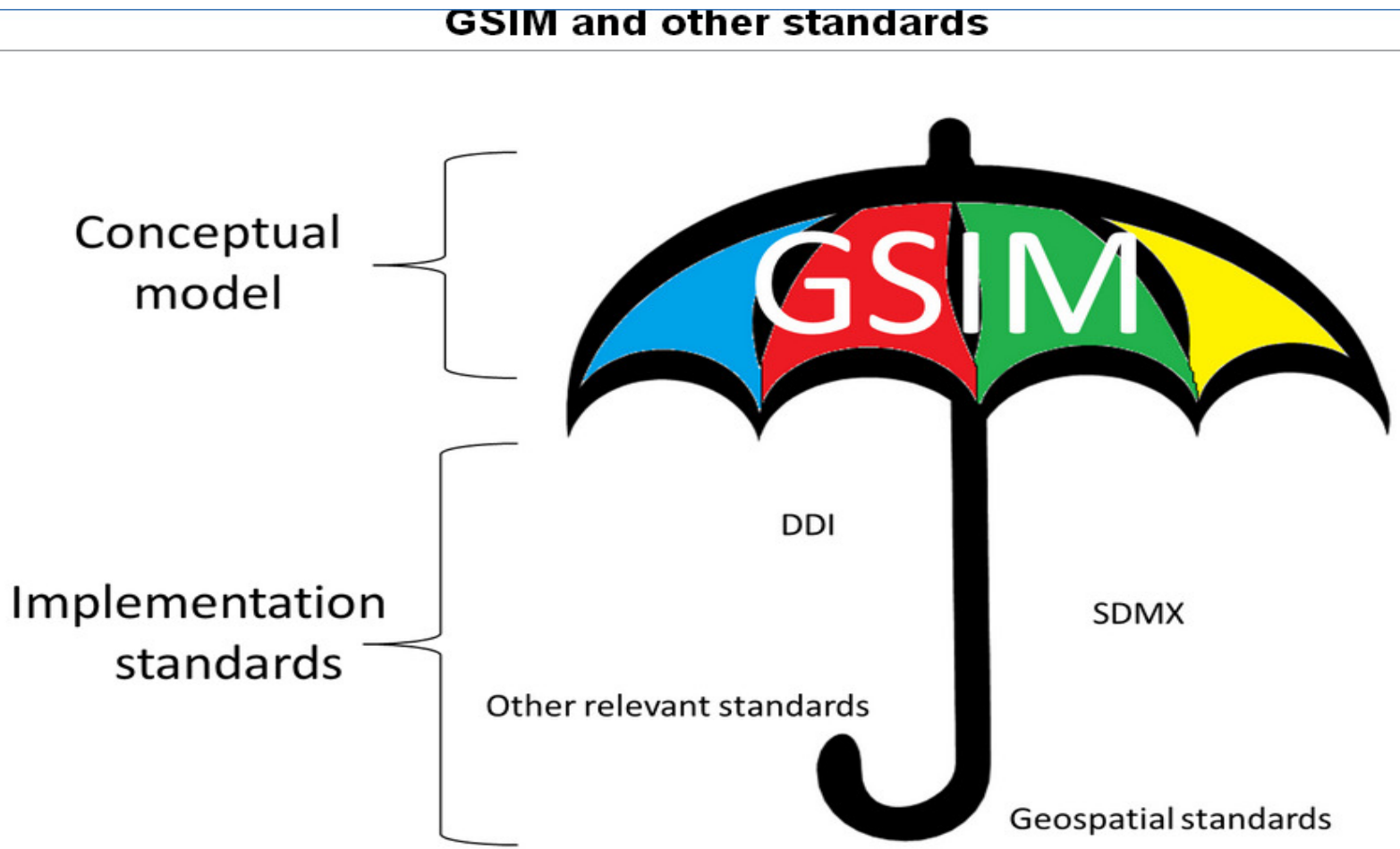
Relationships

GSIM aligns with relevant standards such as DDI and SDMX


Business
Production
Structures
Concepts


Statistical Program
Production Activity
Statistical Program Design
Business
Dissemination Activity
Acquisition Activity
Process Input
Process Method
Production
Process Step
Rule
Process Output
Concepts
Data Set
Data Point
Data Structure
Structures
Data Resource
Product
Variable
Population
Concept
Unit
Classification

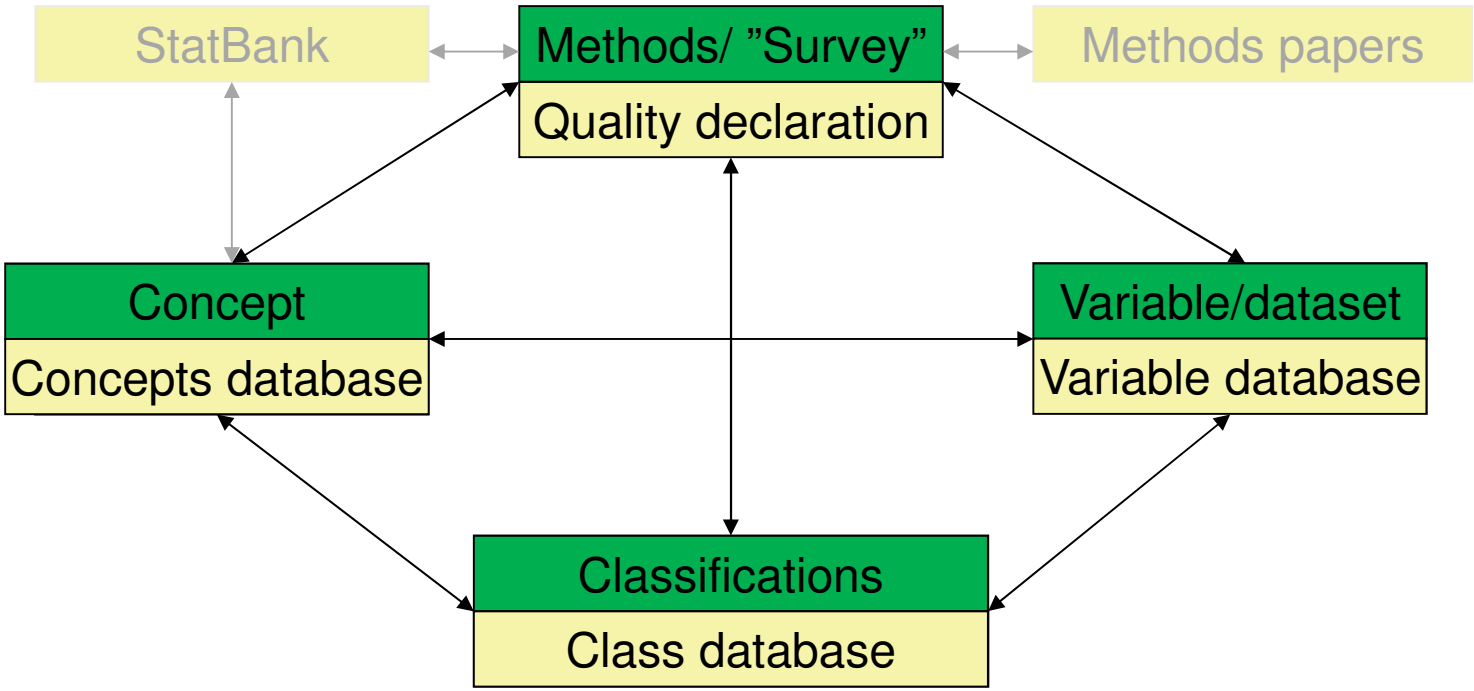
Umbrella perspective



Standards – a starting point

- Abstract level: selected ”classical” metadata elements: methodology, concepts, variables and classifications
- Conceptual level: selected areas in GSIM
- Implementation level: DDI, SDMX and other standards

”Classical” metadata elements implemented using Data Documenation initiative (DDI), SDMX and Colectica



SDMX (Statistical Metadata eXchange)

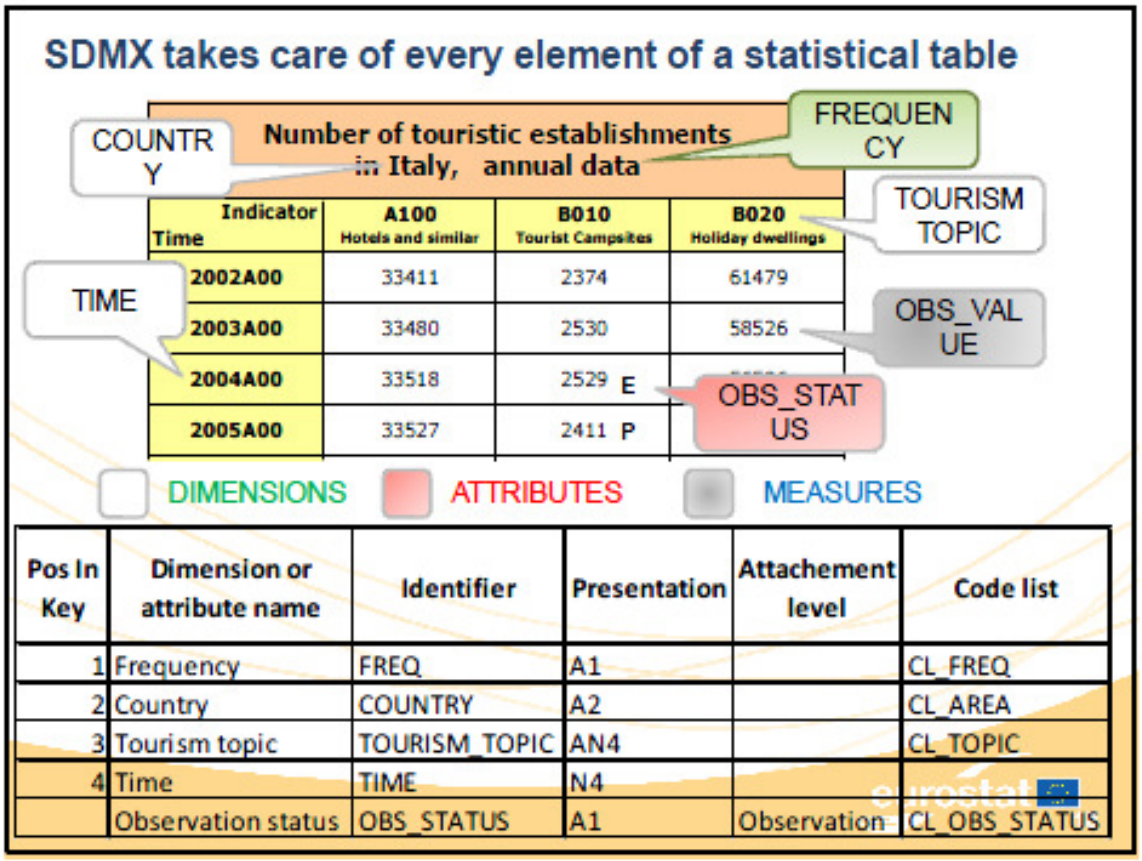
It consists of

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

Advantages – as exchange standard

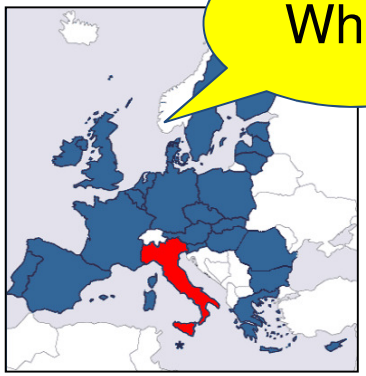
- Common language and understanding
- It is global - BIS, ECB, Eurostat, IMF, OECD, UN joined forces

The SDMX-elements



Pos In Key	Dimension or attribute name	Identifier	Presentation	Attachement level	Code list
1	Frequency	FREQ	A1		CL_FREQ
2	Country	COUNTRY	A2		CL_AREA
3	Tourism topic	TOURISM_TOPIC	AN4		CL_TOPIC
4	Time	TIME	N4		
	Observation status	OBS_STATUS	A1	Observation	CL_OBS_STATUS

Data exchange - including structural and reference metadata



Who?

When?

How?

Where?

Who?

eurostat

What?

Number of touristic establishmen

FREQ: A – Annual
GEO: IT – Italy
TOUR_INDICATOR: A001 – Establishments
UNIT: NBR - Number

Activity / Time	A100 Hotels and similar	B010 Tourist Campsites	B020 Holiday dwellings
2002A00	33518	2529	56586
2003A00	33527	2411	68385
2004A00	33768	2510	68376
2005A00	34058	2587	61810

What?

eurostat EUROPEAN COMMISSION

Capacity of collective tour
establishments, bedroom
Reference Metadata in Euro SDM
Compiling agency: Eurostat, the statistical office of the European Union

For any question on data and metadata, please contact: EUROPEAN STATISTICAL DATA SUPPORT

Download

1. Contact	Top
1.1. Contact organisation	Eurostat, the statistical office of the European Union
1.2. Contact organisation unit	Unit G3 - Short-term statistics; tourism
1.5. Contact mail address	2920 Luxembourg LUXEMBOURG

2. Metadata update	Top
2.1. Metadata last certified	25/09/12
2.2. Metadata last posted	06 March 2013
2.3. Metadata last update	06 March 2013

3. Statistical presentation Top

3.1. Data description

National data

Annual data on number of establishments, bedrooms and bed places in hotels and other collective accommodation establishments.

Regional data

Annual data on number of establishments, bedrooms and bed places in hotels and other collective accommodation establishments at NUTS 3 level.

Please note that for paragraphs where no metadata for regional data has been specified, the regional metadata is identical to the metadata provided for the national data.

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

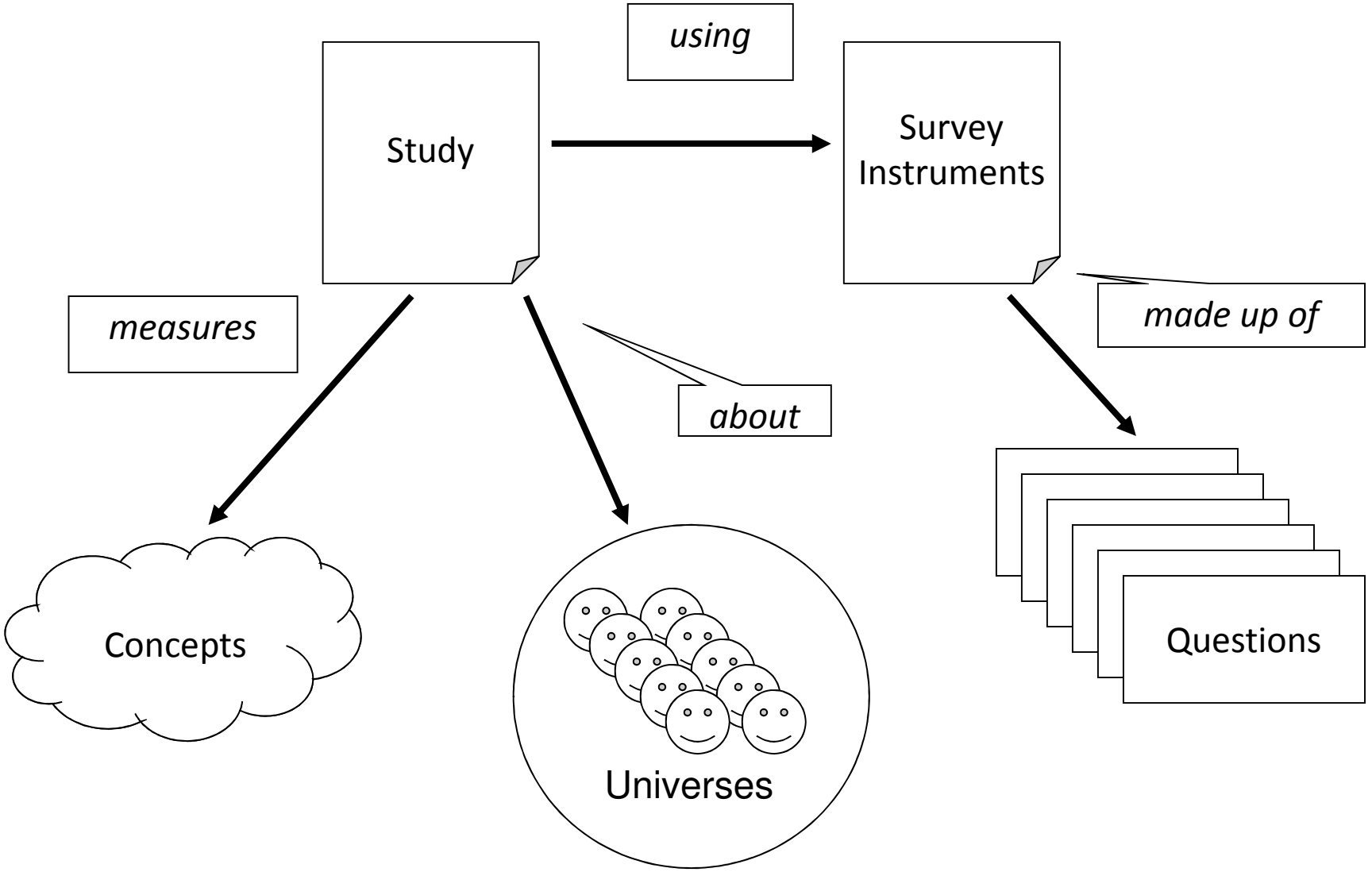
Integration of concepts, variables, classifications quality

Both for schema and register based statistics

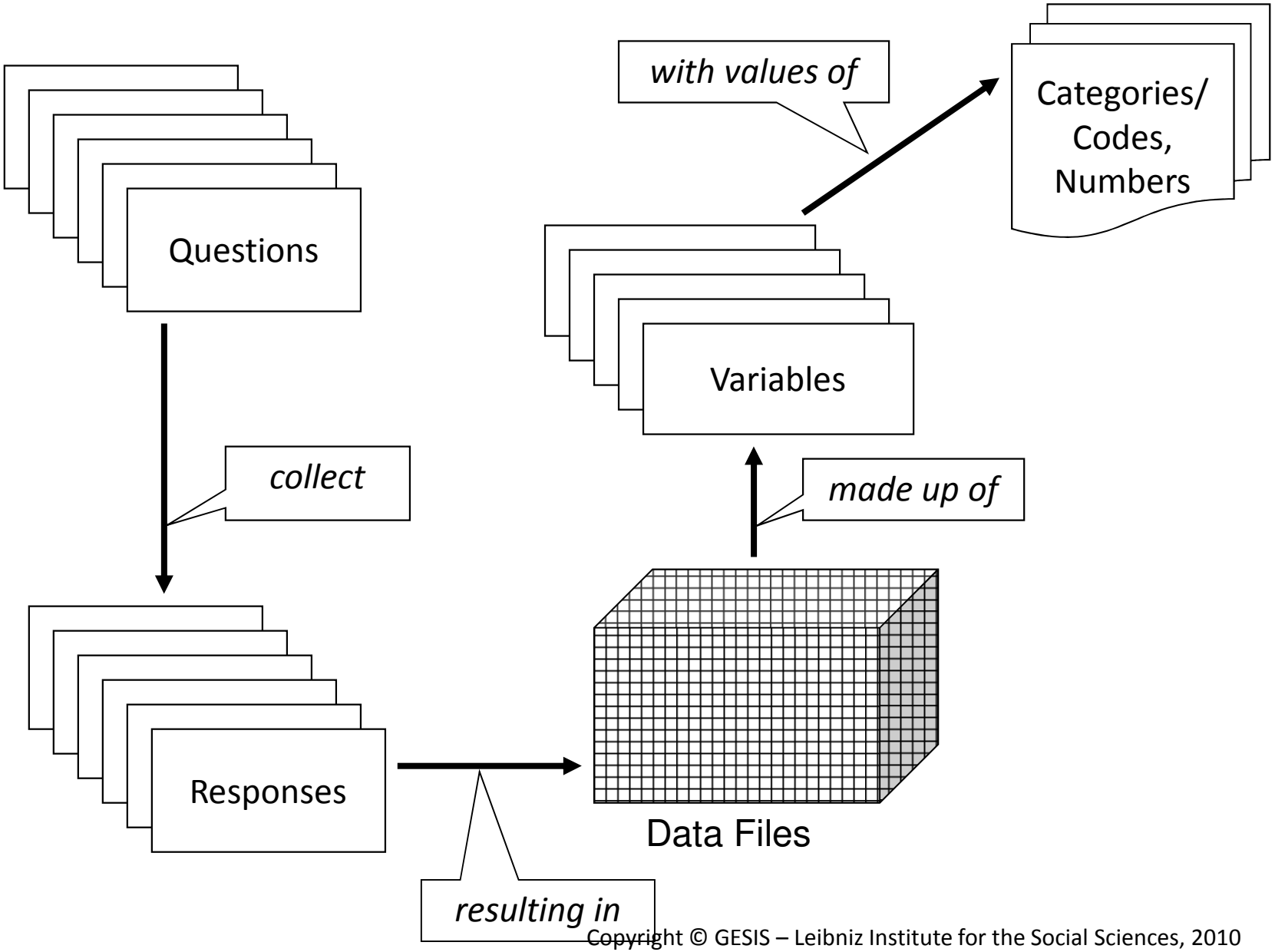
Model currently used in Australia, New Zealand, Canada etc.(together with SDMX)

Tools avaiable

Statistics and DDI in 60 seconds



Statistics and DDI in 60 seconds



DDI – types of metadata

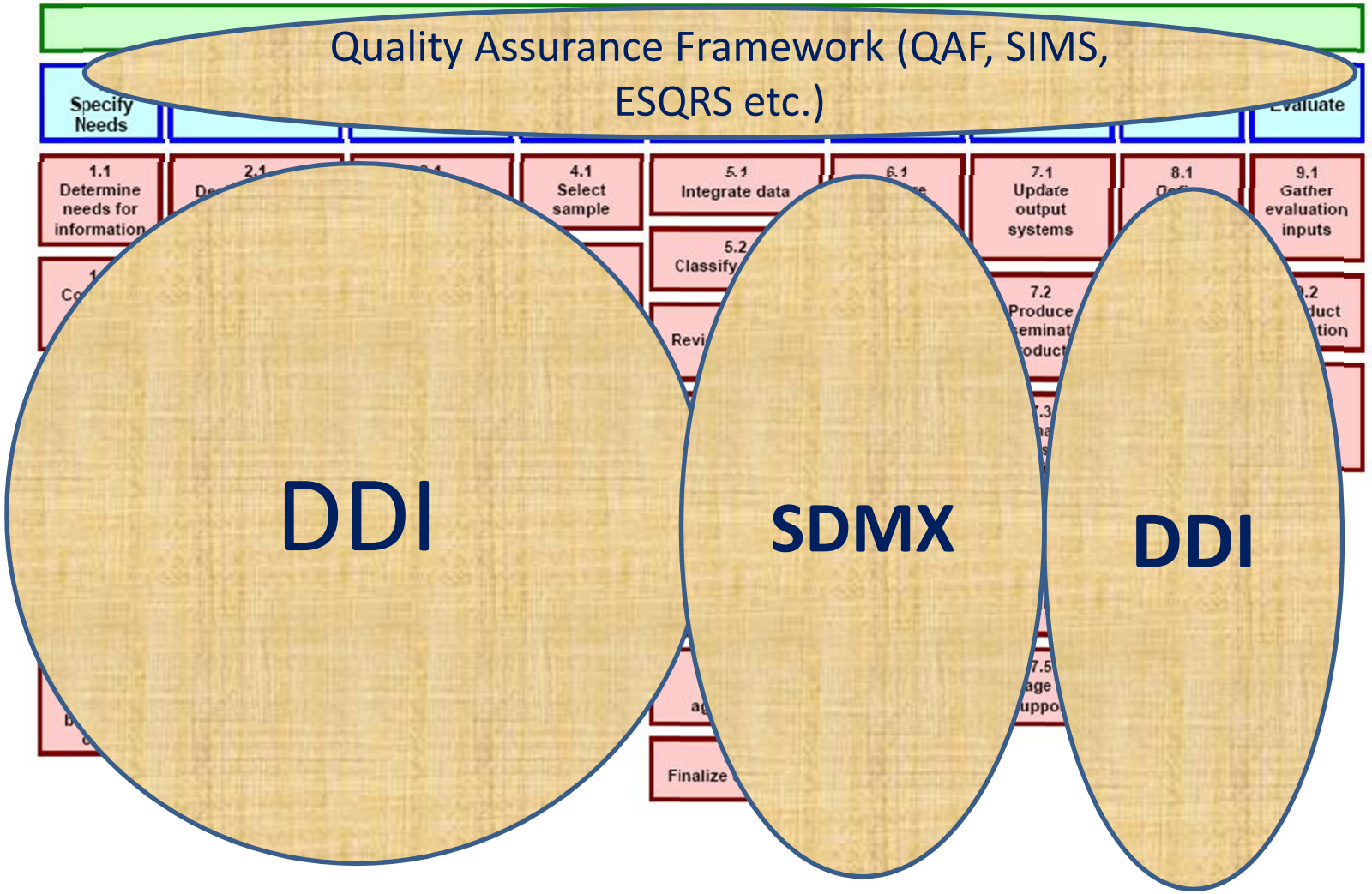
Metadatatypes:

- 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”)

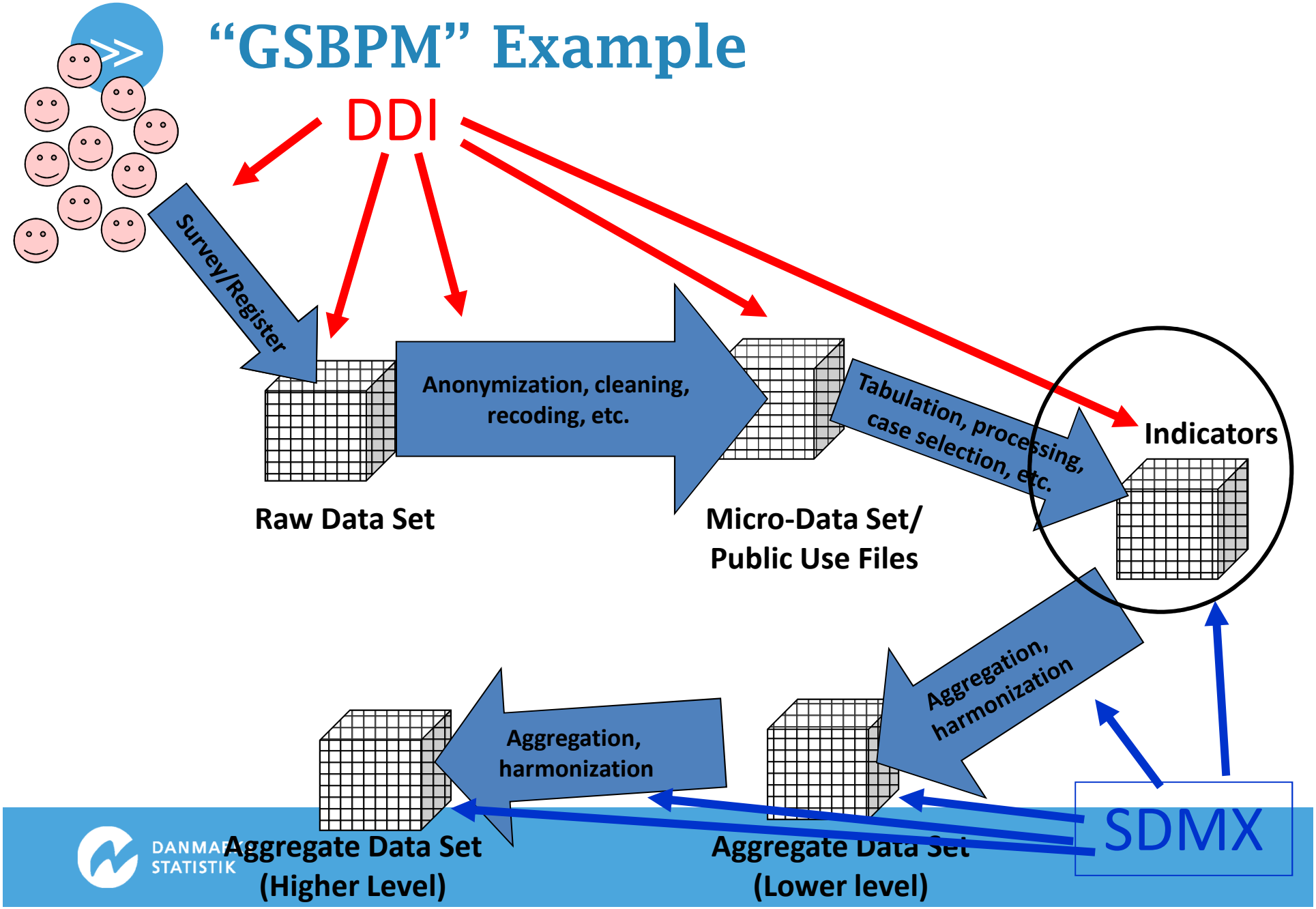
For surveys:

- Survey instruments (“questionnaire”, “form”)
- Questions (“observations”)
- Responses

GSBPM combined with Quality Frameworks, DDI and SDMX

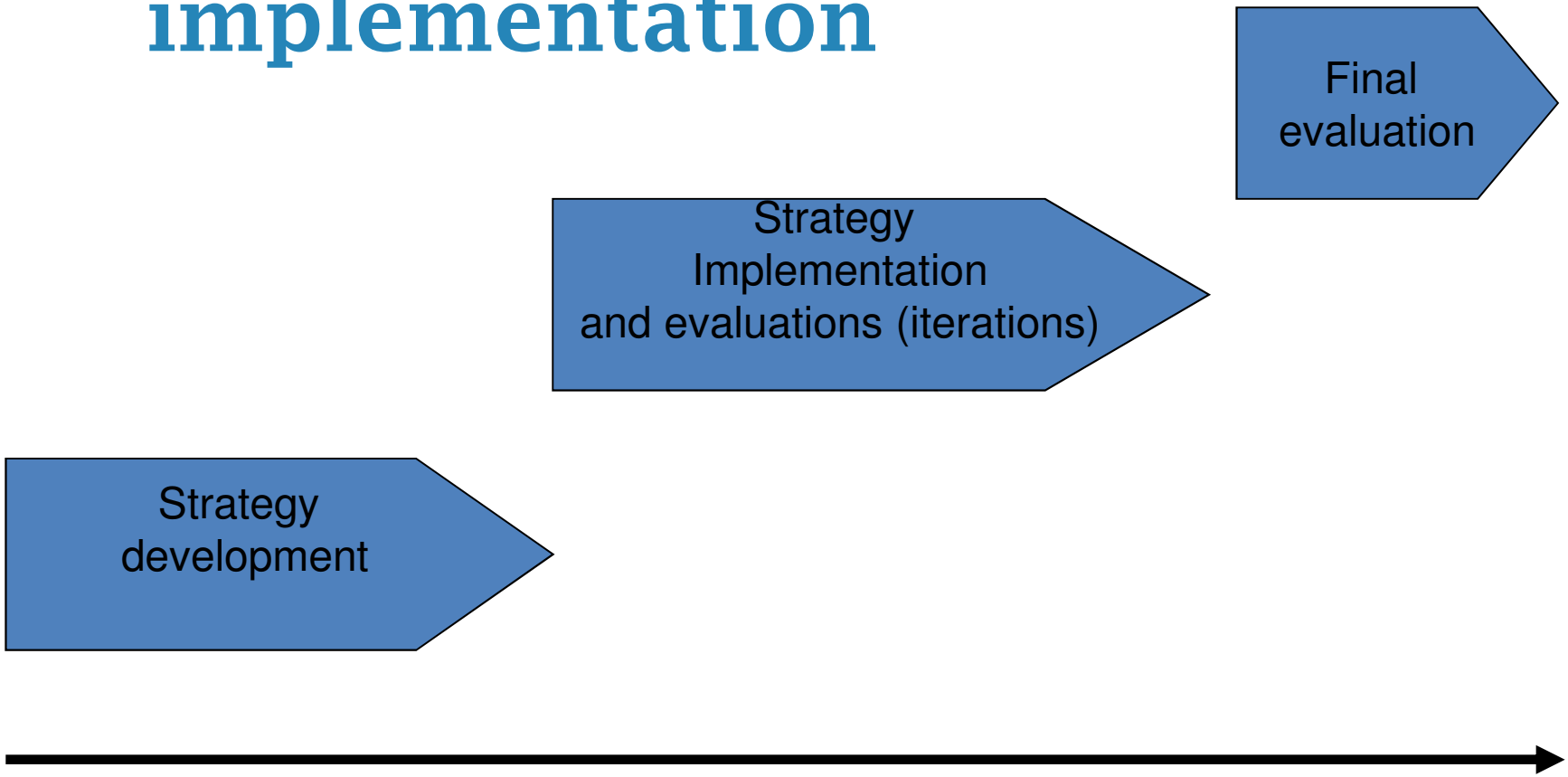


"GSBPM" Example



5. Towards a strategy for the development of a metadata system for DoS

Metadata strategy – development and implementation



Strategy outline

1. Introduction, including situation today
2. International standards, definition of metadata and roles of a Statistical Metadata System
3. Objectives and results
4. Tasks and time-table
5. Organisation

Annex 1. Guidelines for production and use of metadata

Annex 2. IT-solution

1. High level architecture
2. Applications and databases


The End!