European Commission

CSPA

Common Statistical Production Architecture

GSBPM and GSIM as a framework for describing statistical production process and supporting information

Carlo Vaccari
Istat (vaccari@istat.it)

Eurostat

GSBPM background

Defining and mapping business processes in statistical organisations started at least 10 years ago

Many terms used:

- “Statistical value chain”
- “Survey life-cycle”
- “Statistical process cycle”
- “Business process model”

Now: **Generic Statistical Business Process Model**

Why do we need a model?

- To define, describe and map statistical processes in a coherent way
- To standardize process terminology
- To compare / benchmark processes within and between organizations
- To identify synergies between processes
- To make informed decisions on systems architectures and organization of resources

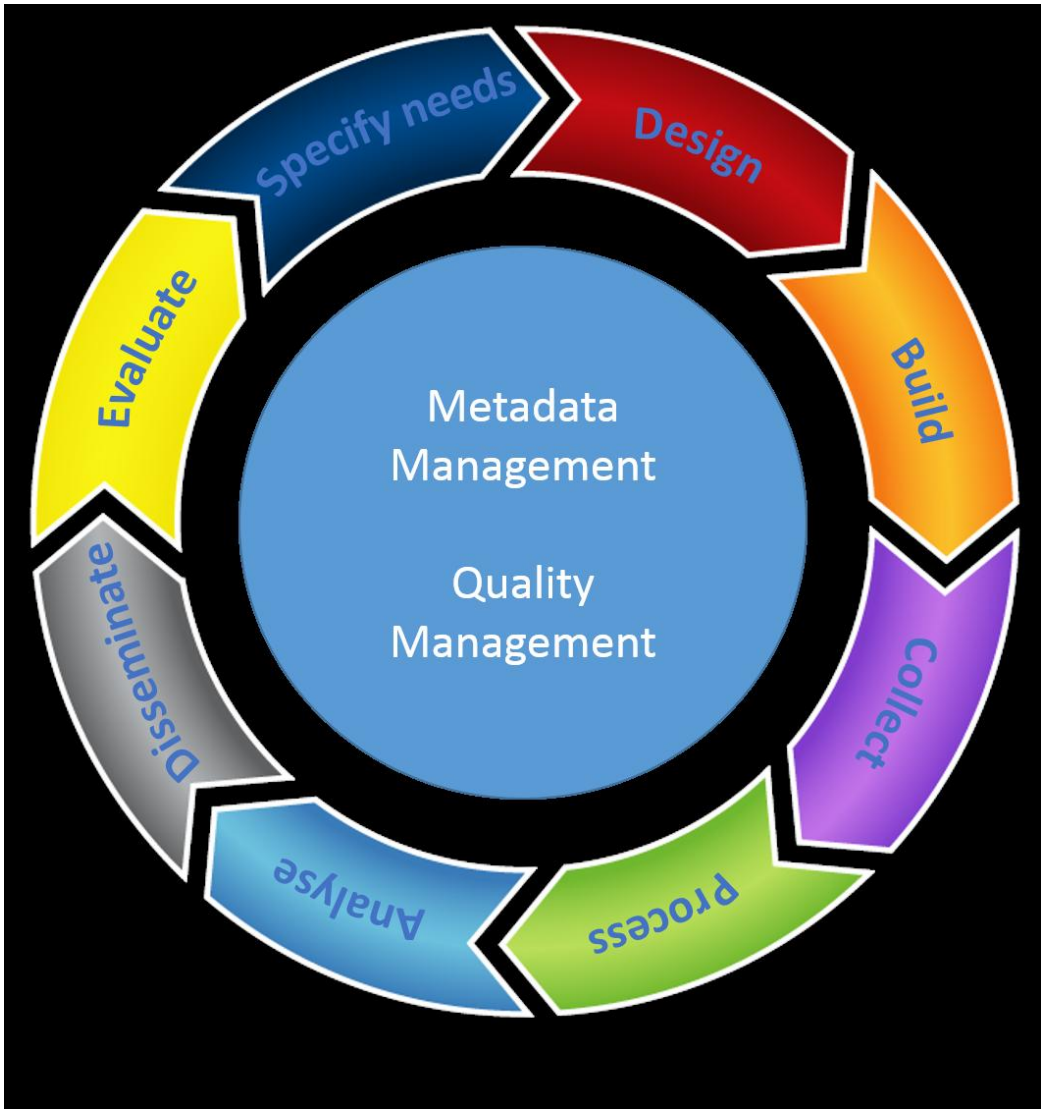
Applicability

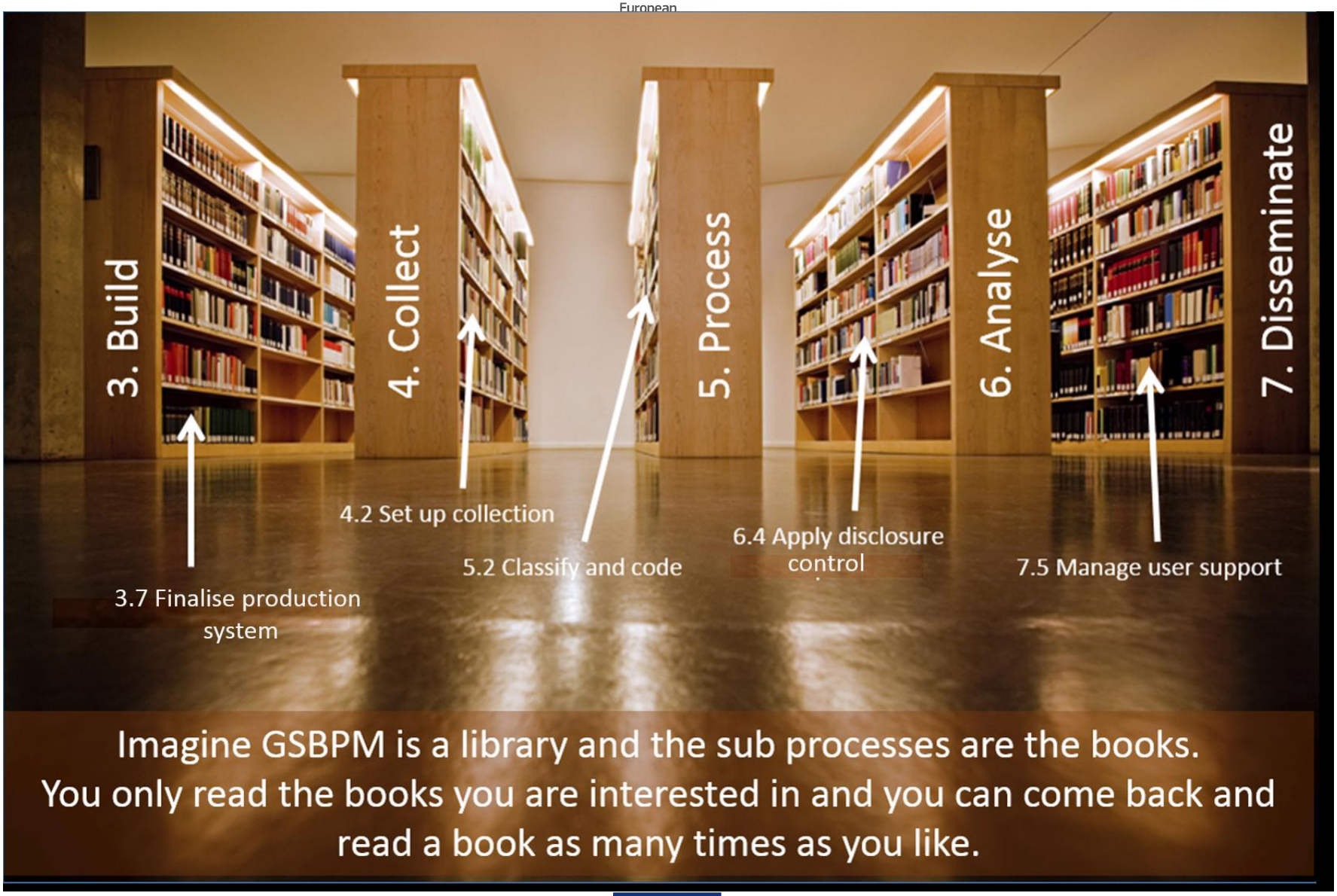
All activities undertaken by producers of official statistics which result in data outputs

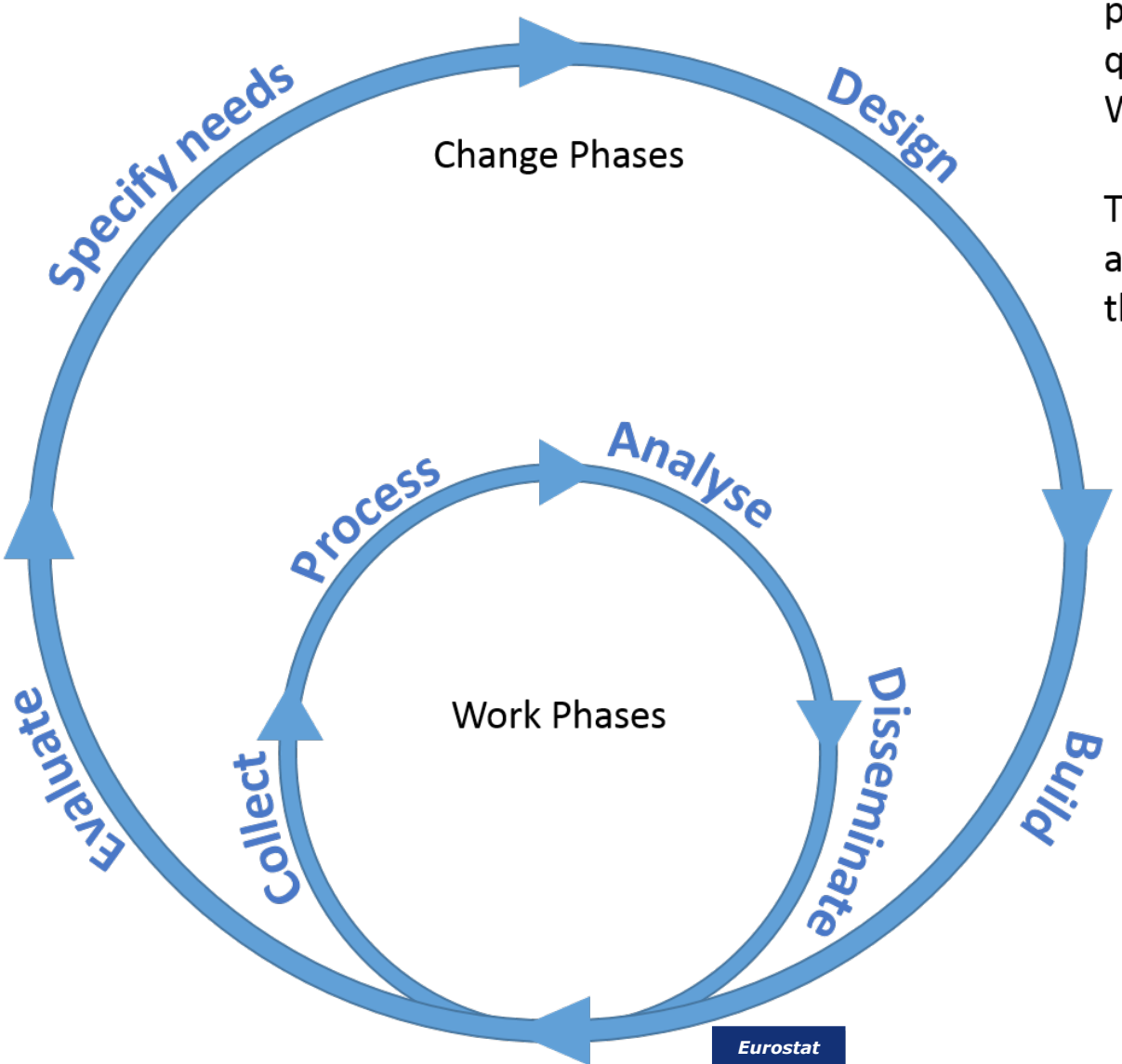
National and international statistical organizations

Independent of data source, can be used for:

- Surveys / censuses
- Administrative sources / register-based statistics
- Mixed sources
- Big Data?


Specify needs
Design
Build
Collect
Process
Analyse
Disseminate
Evaluate
Metadata Management
Quality Management


3. Build
4. Collect
5. Process
6. Analyse
7. Disseminate
3.7 Finalise production system
4.2 Set up collection
5.2 Classify and code
6.4 Apply disclosure control
7.5 Manage user support
Imagine GSBPM is a library and the sub processes are the books.
You only read the books you are interested in and you can come back and read a book as many times as you like.



In GSBPM, there are some phases which are undertaken quickly and frequently – the Work Phases.

There are other phases which are undertaken less often - the Change Phases.

Quality Management / Metadata Management

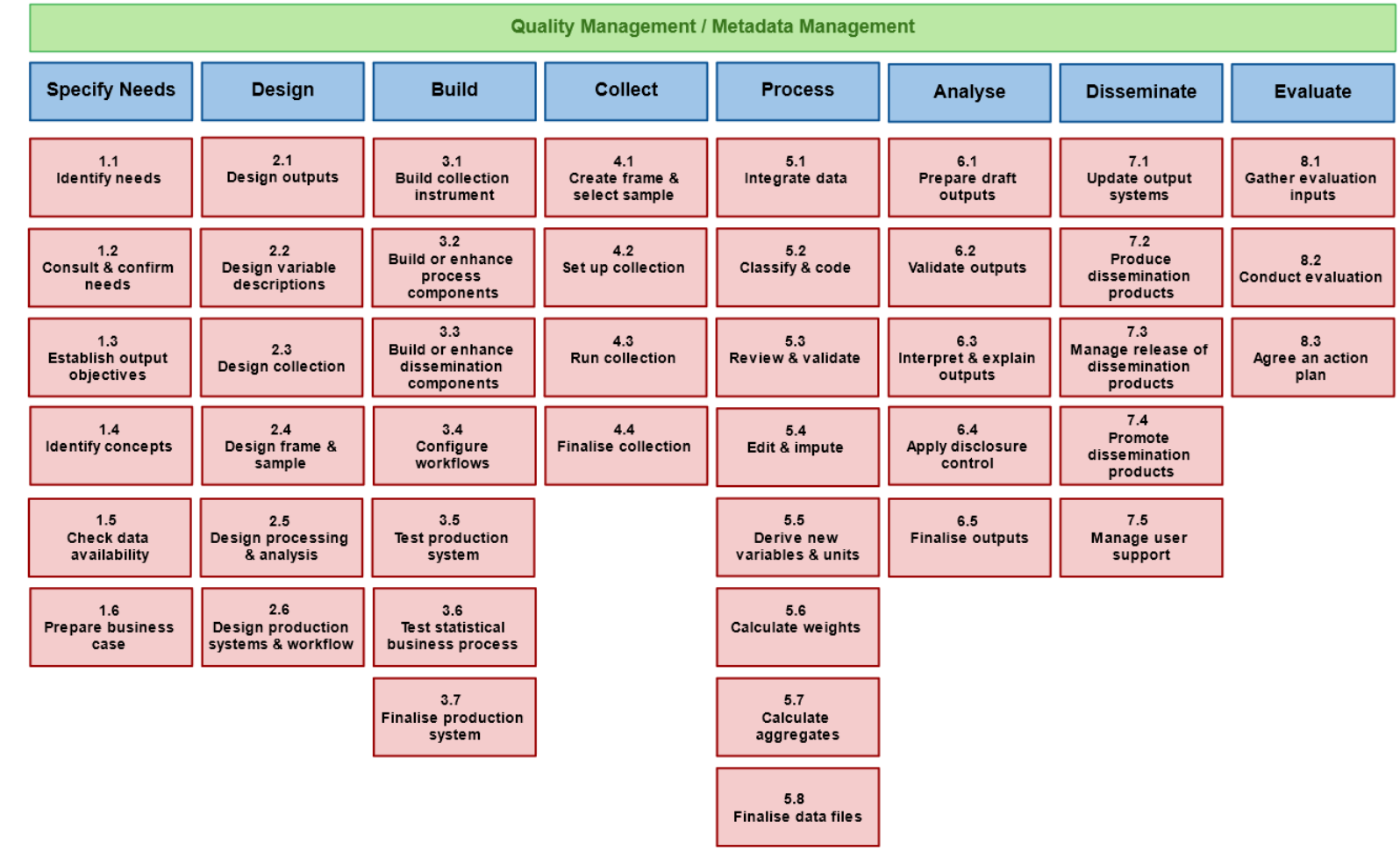
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	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 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Structure of the model

Process

Phases

Sub-processes



Current version

April 2009: version 4.0
December 2013: version 5.0
Now taken over from HLG
Main changes 5.0:

- Phase 8 (Archive) has been removed, and incorporated into the over-arching process
- A new sub-process: "Build or enhance dissemination components" within the "Build"
- Several sub-processes have been re-named
- The descriptions of the sub-processes have been updated and expanded where necessary and now is less survey-centric, in recognition of the growing use of non-survey sources like administrative data, big data ...

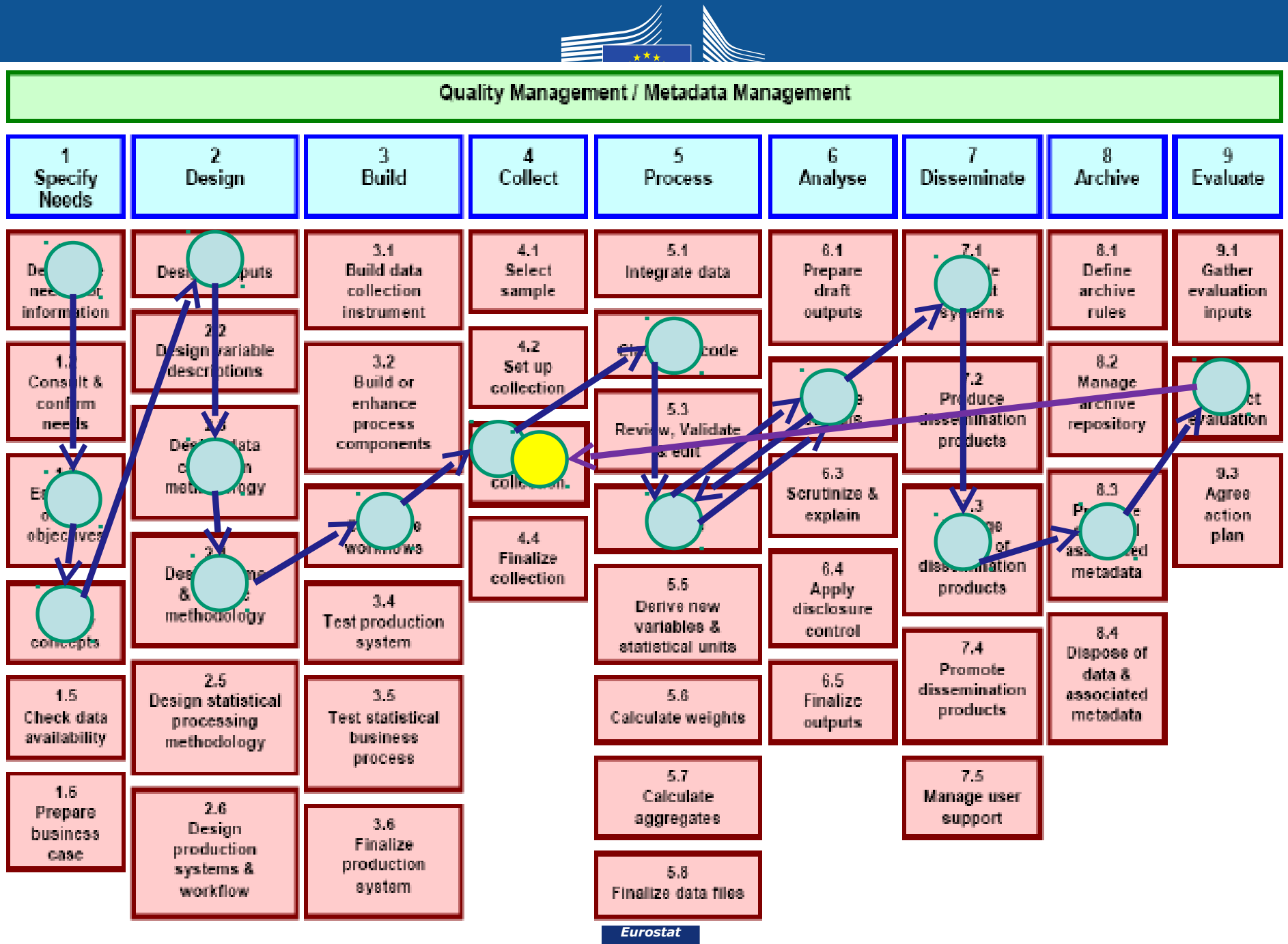
Key features

Not a linear model !

Sub-processes do not have to be followed in a strict order

It is a matrix, through which there are many possible paths, including iterative loops within and between phases

Some iterations of a regular process may skip certain sub-processes


Quality Management / Metadata Management
1 Specify Needs
2 Design
3 Build
4 Collect
5 Process
6 Analyse
7 Disseminate
8 Archive
9 Evaluate
information
1.2 Consult & confirm needs
objectives
concepts
1.5 Check data availability
1.6 Prepare business case
2.2 Design variable descriptions
methodology
methodology
2.5 Design statistical processing methodology
2.6 Design production systems & workflow
3.1 Build data collection instrument
3.2 Build or enhance process components
workflows
3.4 Test production system
3.5 Test statistical business process
3.6 Finalize production system
4.1 Select sample
4.2 Set up collection
4.4 Finalize collection
5.1 Integrate data
code
5.3 Review, Validate & edit
5.6 Derive new variables & statistical units
5.6 Calculate weights
5.7 Calculate aggregates
5.8 Finalize data files
6.1 Prepare draft outputs
6.3 Scrutinize & explain
6.4 Apply disclosure control
6.5 Finalize outputs
7.2 Produce dissemination products
products
7.4 Promote dissemination products
7.5 Manage user support
8.1 Define archive rules
8.2 Manage archive repository
metadata
8.4 Dispose of data & associated metadata
9.1 Gather evaluation inputs
9.3 Agree action plan
Eurostat

Key features

Not a linear model !

Sub-processes do not have to be followed in a strict order

It is a matrix, through which there are many possible paths, including iterative loops within and between phases

Some iterations of a regular process may skip certain sub-processes

Mapping with other standards

Mapping with other standards:

- Mapping to the Fundamental Principles of Official Statistics
- Mapping with DDI
- Mapping with SDMX

National implementations etc.

See in the link experiences from Armenia, Australia, Canada, Denmark, Eurostat, Ireland, Korea, Spain, Sweden, Turkey

Other uses:

- Process-related metadata
- Managing Statistical programs
- Tool for Cooperation: Software Inventory

Relations with CSPA

In the context of statistical modernisation, the aim is to align the enterprise architectures of different organisations, creating an “industry architecture” for the whole “official statistics industry”.

This approach is intended to facilitate collaboration, sharing and joint development of the components and services that are needed for the different parts of the statistical business process (defined in relation to the GSBPM).

The result is the Common Statistical Production Architecture (CSPA), first released at the end of 2013.

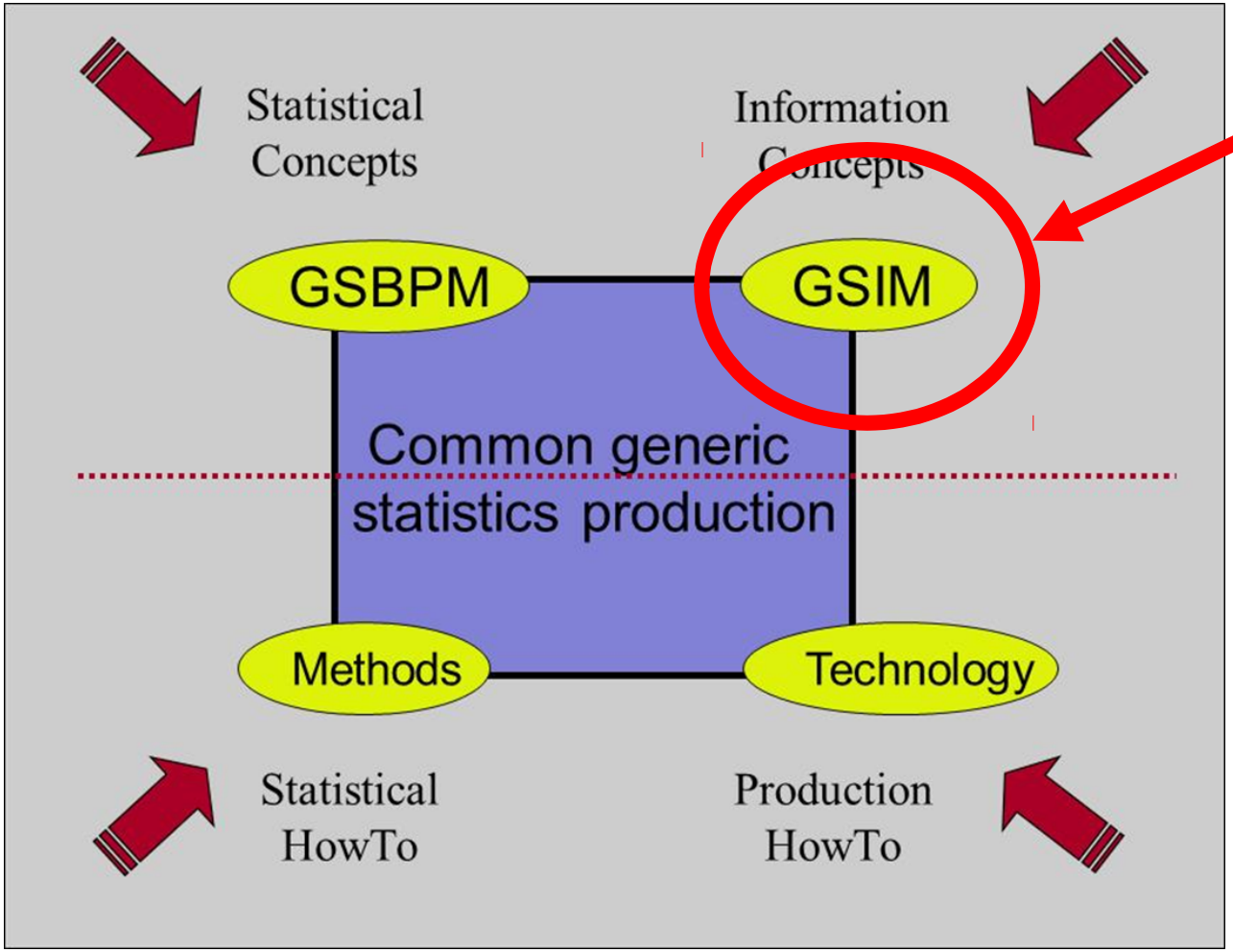
GSIM is complementary to GSBPM

Another model is needed to describe information objects and flows within the statistical business process

GSBPM is a business process model

GSIM models information on the process

GSIM



You are here

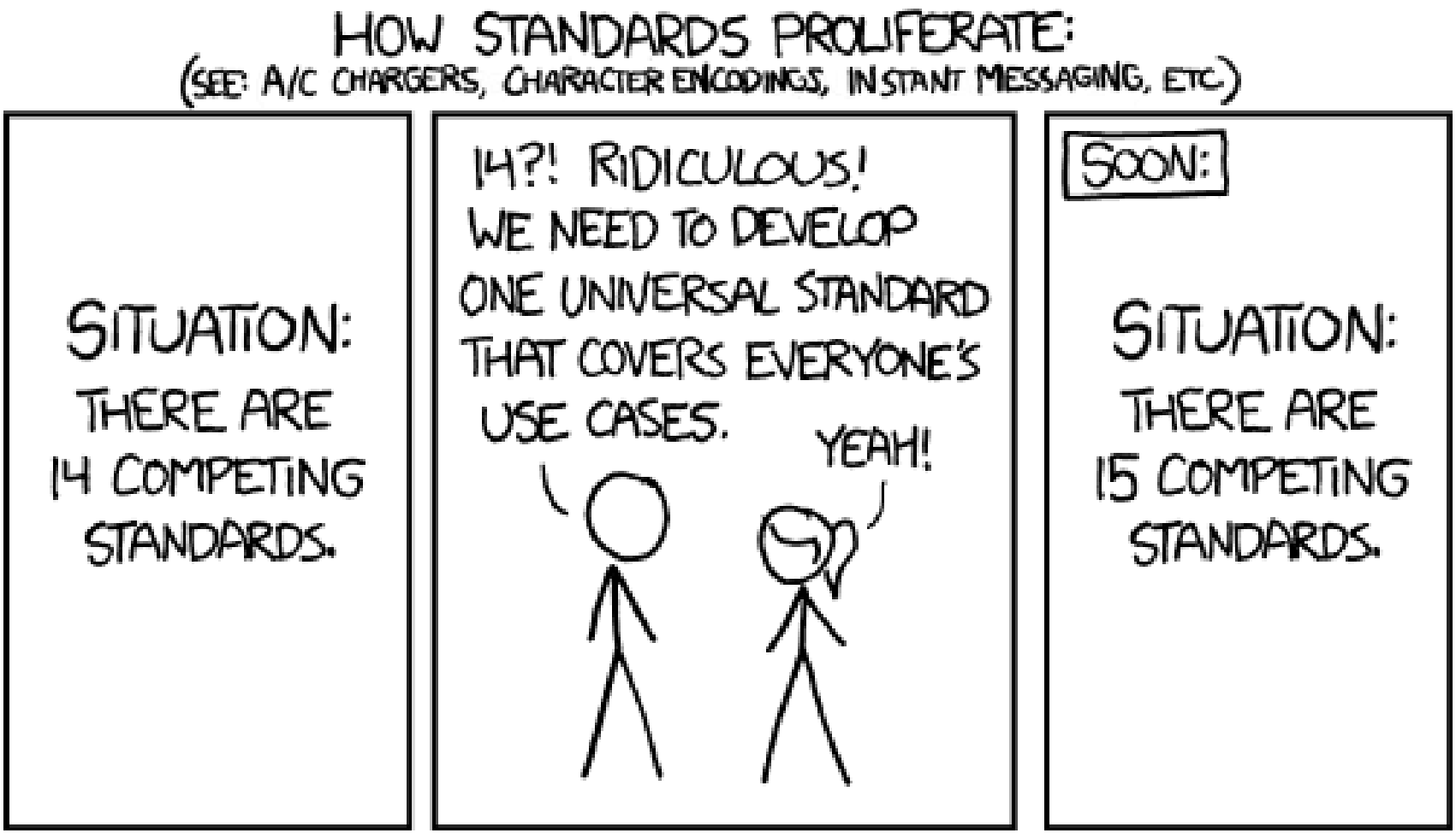
GSIM

GSIM is a reference framework of internationally agreed definitions, attributes and relationships that describe the pieces of information that are used in the production of official statistics (information objects)
This framework enables descriptions of the definition, management and use of data and metadata throughout the statistical information process

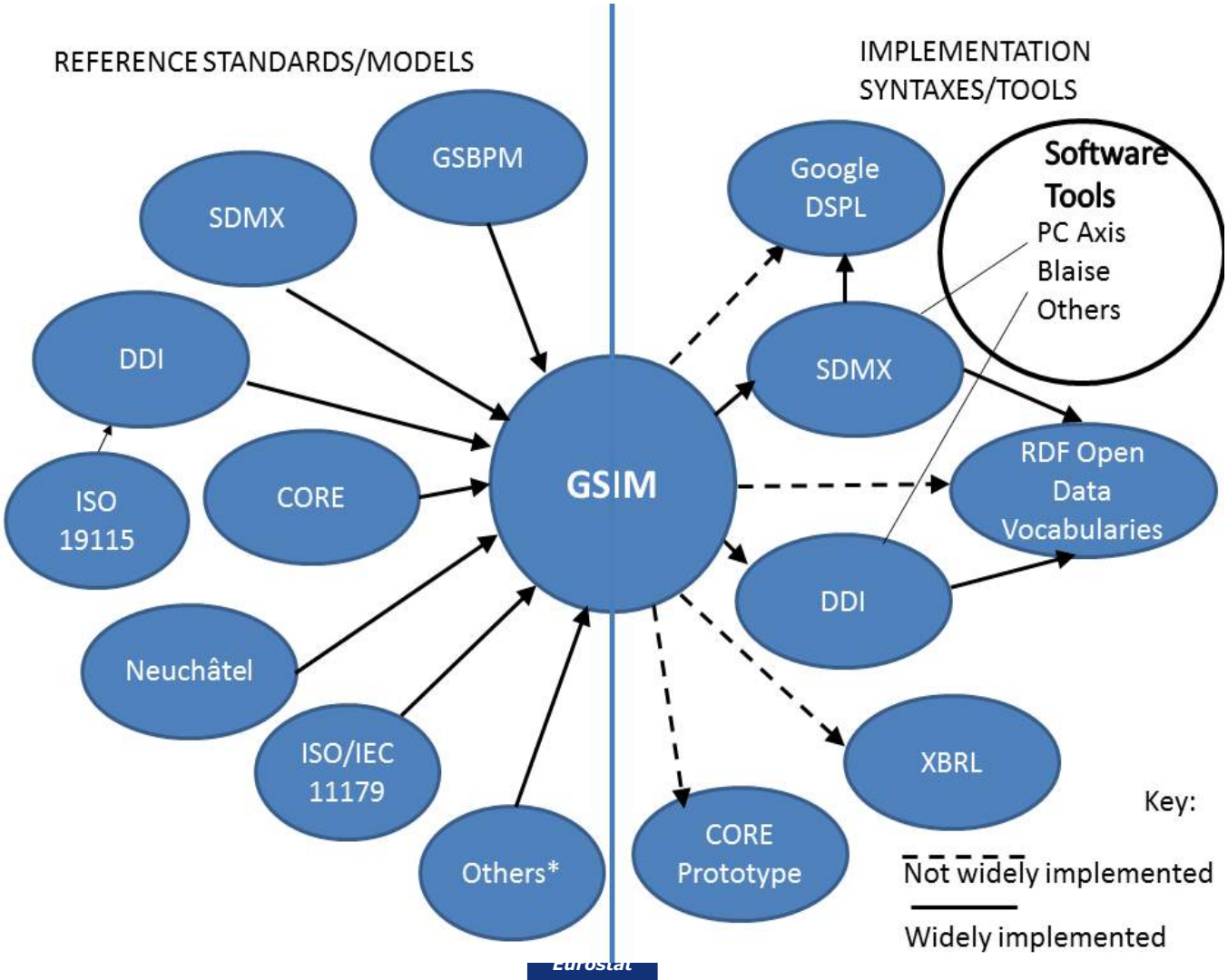
Purposes of GSIM

- Improve communication
- Generate economies of scale
- Enable greater automation
- Provide a basis for flexibility and innovation
- Build staff capability by using GSIM as a teaching aid
- Validate existing information systems

… standards …


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

… standards …


REFERENCE STANDARDS/MODELS
IMPLEMENTATION SYNTAXES/TOOLS
GSBPM
SDMX
DDI
ISO 19115
CORE
Neuchâtel
ISO/IEC 11179
Others*
GSIM
Google DSPL
Software Tools
PC Axis
Blaise
Others
SDMX
RDF Open Data Vocabularies
DDI
XBRL
CORE Prototype
Key:
Not widely implemented
Widely implemented

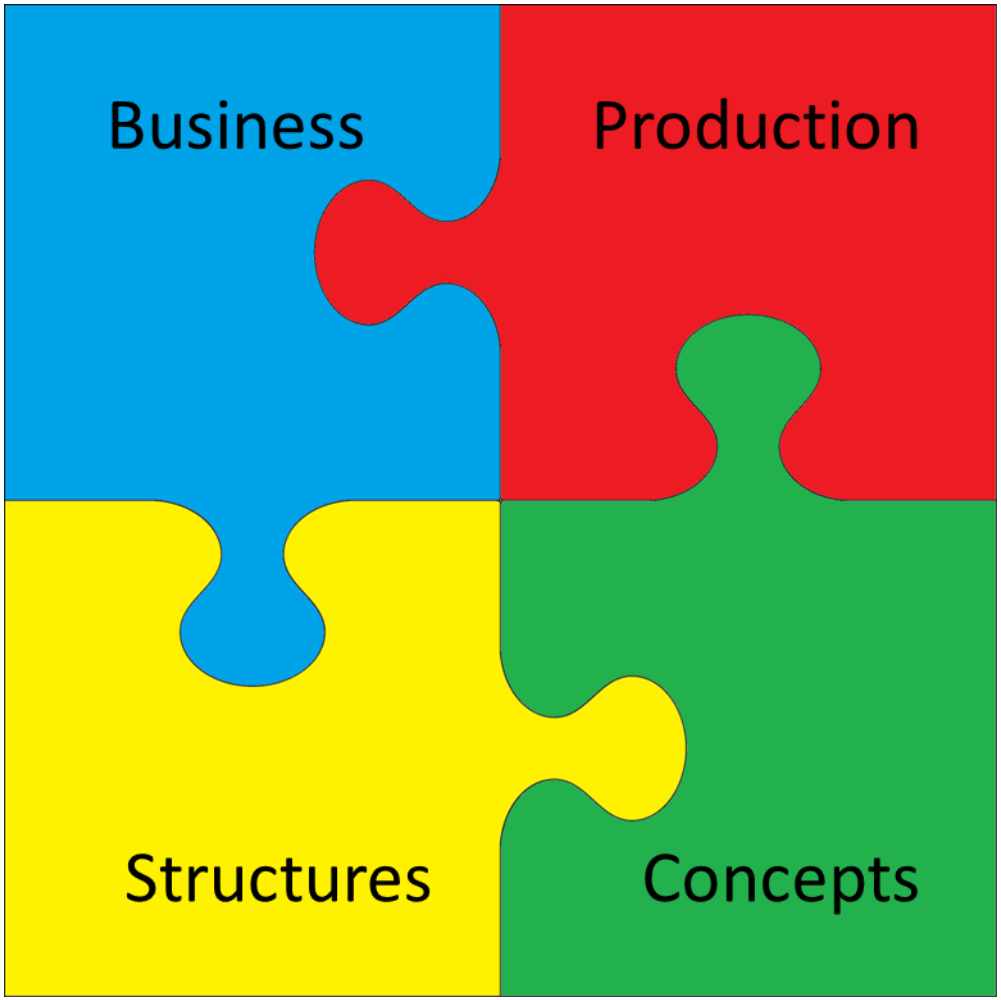
… standards …

Using common standards, statistics can be produced in a more efficient way

No domain is special!

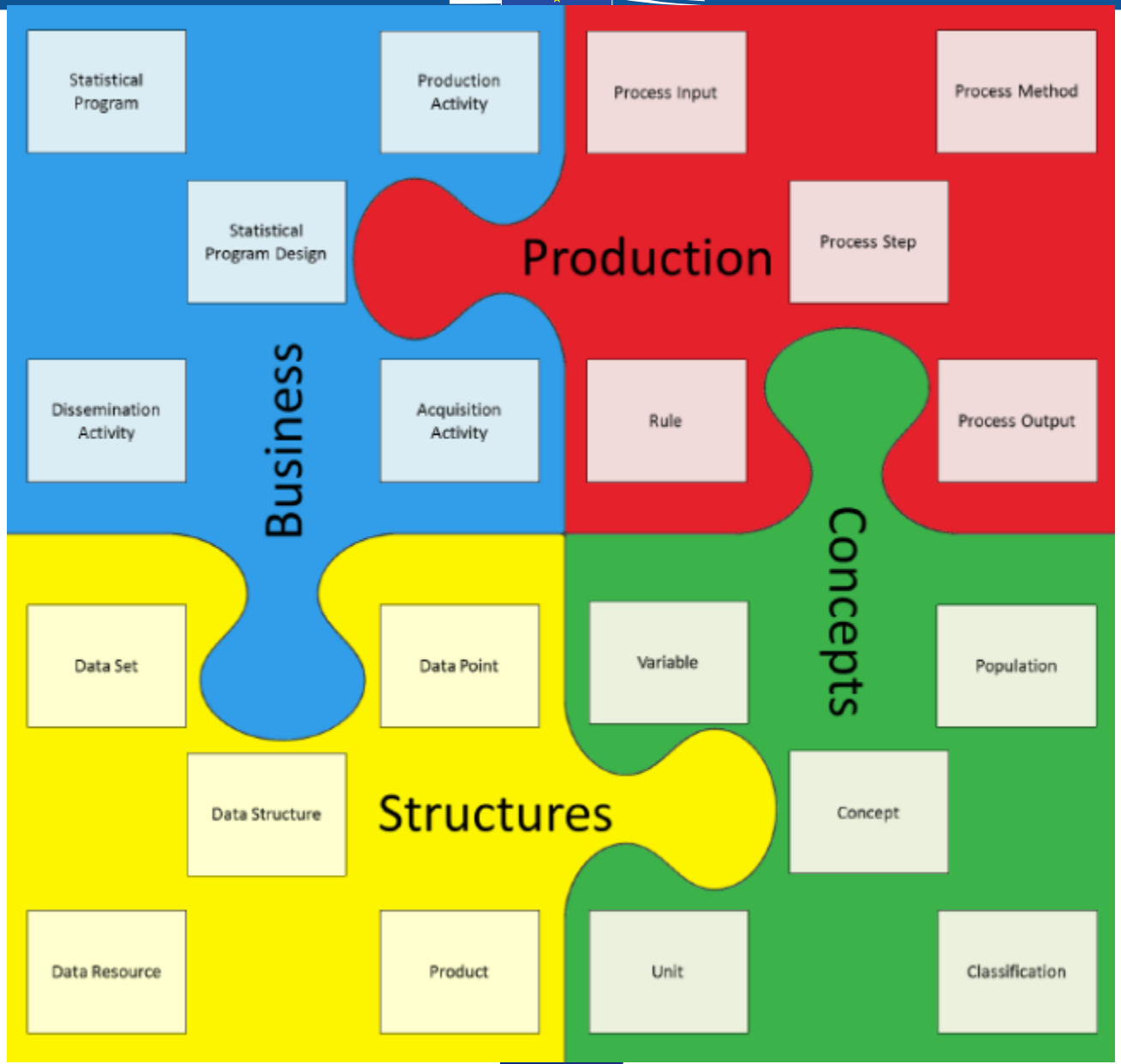
European Commission

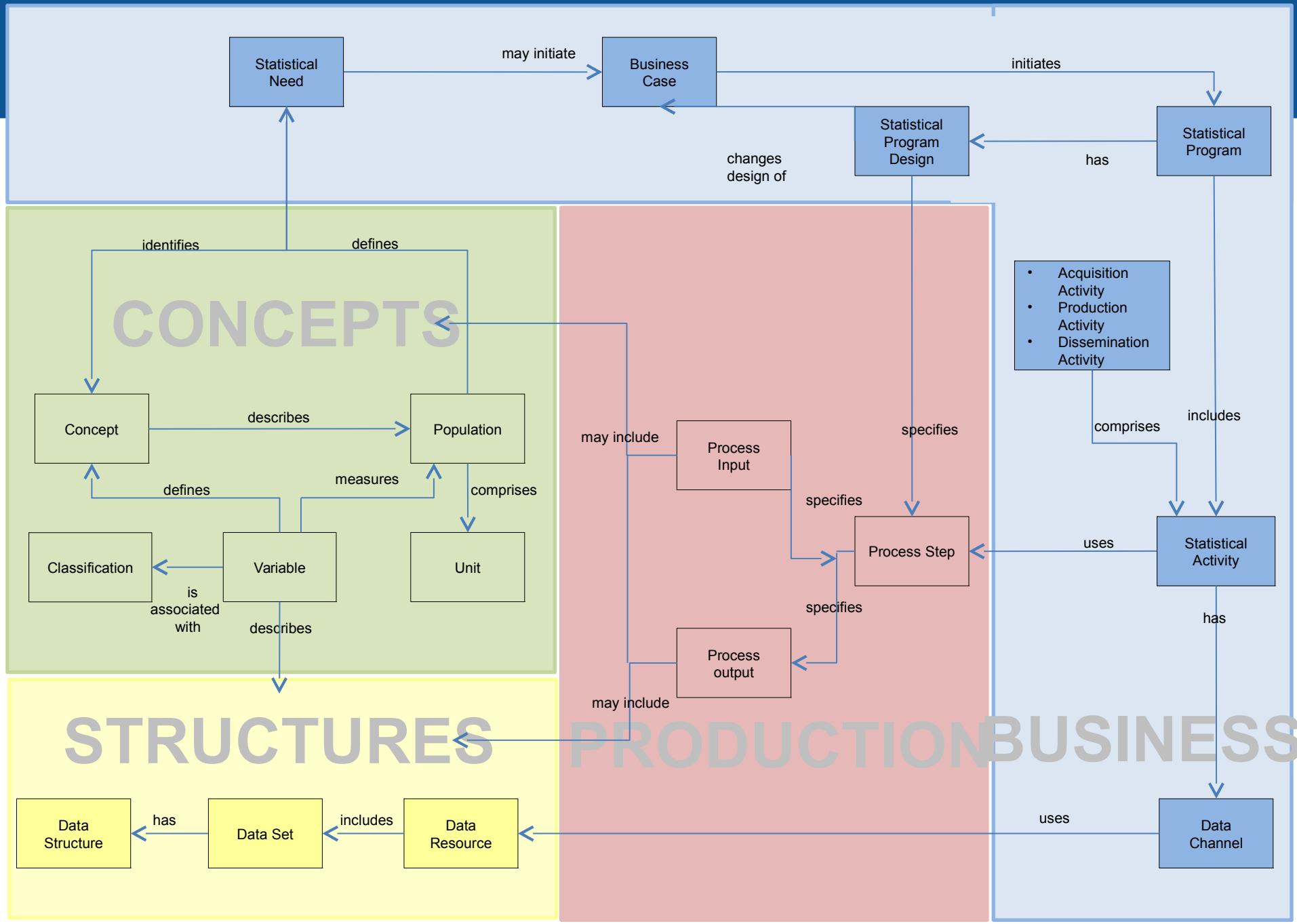
>100 objects grouped in four broad categories

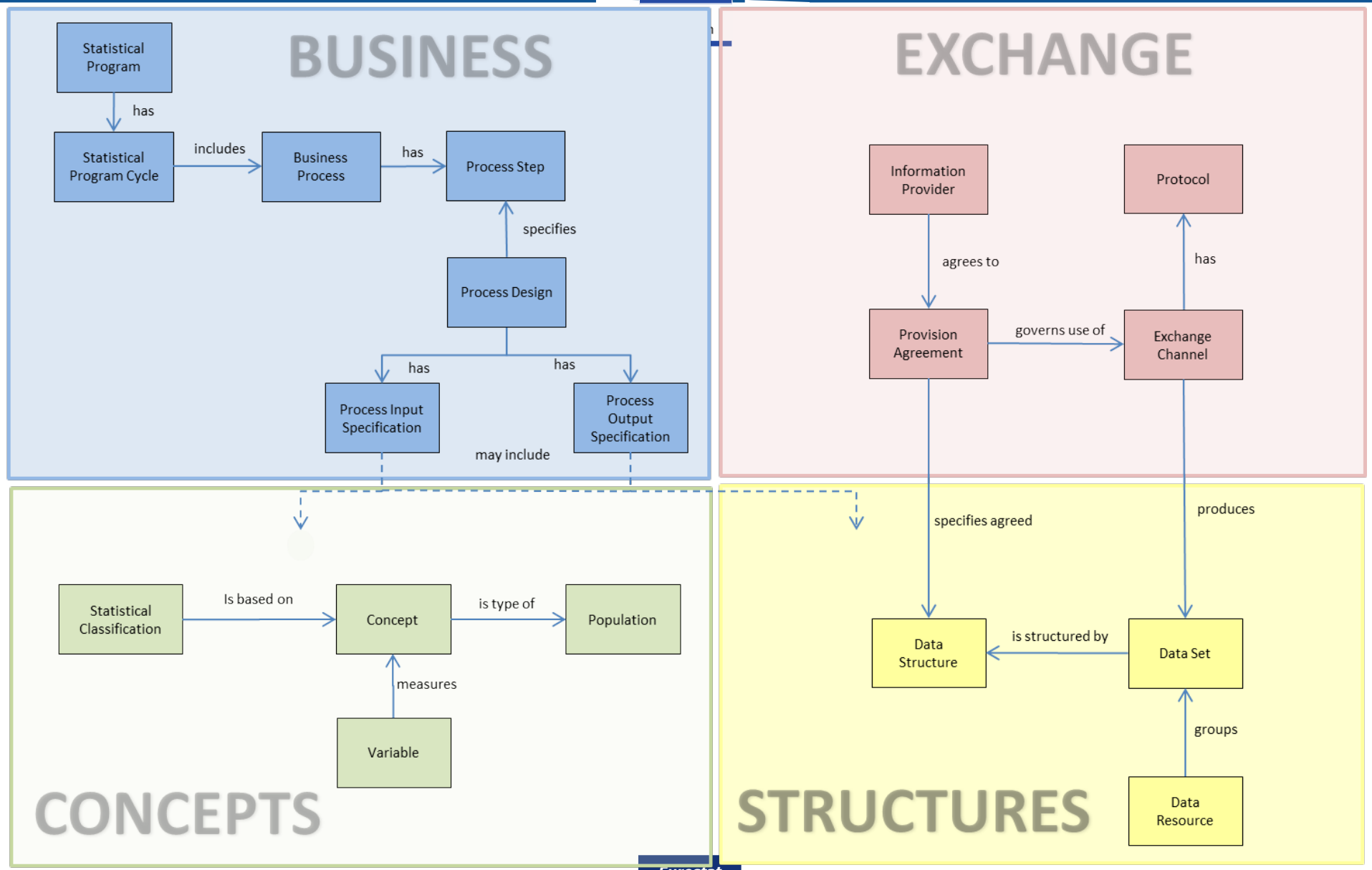


Top-level groups

- The **Business** group is includes the designs and plans of statistical programs i.e. the identification of a Statistical Need, the Business Processes and their evaluations
- The **Exchange** group is used to catalogue the information that comes in and out of a statistical organization via Exchange Channels. It includes objects about the collection and dissemination of information
- The **Concepts** group is used to define the meaning of data, providing an understanding of what the data are measuring
- The **Structures** group is used to describe and define the terms used in relation to information and its structure


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


Statistical Need
may initiate
Business Case
initiates
Statistical Program Design
changes design of
has
Statistical Program
identifies
defines
CONCEPTS
Concept
describes
Population
defines
measures
comprises
Classification
Variable
Unit
is associated with
describes
may include
Process Input
specifies
specifies
Process Step
specifies
Process output
may include
PRODUCTION
Acquisition Activity
Production Activity
Dissemination Activity
comprises
includes
uses
Statistical Activity
has
BUSINESS
STRUCTURES
Data Structure
has
Data Set
includes
Data Resource
uses
Data Channel


BUSINESS
Statistical Program
has
Statistical Program Cycle
includes
Business Process
has
Process Step
specifies
Process Design
has
has
Process Input Specification
Process Output Specification
may include
EXCHANGE
Information Provider
Protocol
agrees to
has
Provision Agreement
governs use of
Exchange Channel
specifies agreed
produces
CONCEPTS
Statistical Classification
Is based on
Concept
is type of
Population
measures
Variable
STRUCTURES
Data Structure
is structured by
Data Set
groups
Data Resource

GSIM long term benefits

- GSIM provides a set of standardized information objects, inputs and outputs in the design and production of statistics, regardless of subject matter. So GSIM enables statistical organizations to rethink how their business could be more efficiently organized
- GSIM could be used to direct future investment towards areas of statistical production where the need is greatest
- It could also enable some degree of specialization within the international statistical community
- GSIM could:
  - Create an environment prepared for reuse and sharing of methods, components and processes
  - Provide the opportunity to implement rule based processes
  - Facilitate generation of economies of scale through development of common tools

GSIM immediate benefits

- A significant benefit of using GSIM is that it provides a common language to improve communication at different levels:
  - Between different roles in statistical production (business / IT)
  - Between the different statistical subject matter domains
  - Between statistical organizations at national/international levels
- More efficient exchange of data and metadata within and between statistical organizations, and also with external users and suppliers
- GSIM can be used by organizations now to:
  - Build capability among staff by using GSIM as a teaching aid to understand view of complex information and clear definitions
  - Validate existing information systems and compare with emerging international best practice
  - Guide development of international or local standards to ensure they meet the needs of the international statistical community

Move to GSIM in practice

GSIM could lead to:

- A foundation for standardized statistical metadata use throughout systems
- A standardized framework to aid in consistent and coherent design capture
- Increased sharing of system components
- Common terminology across and between statistical agencies
- It allows NSIs and standards bodies, such as SDMX and DDI, to understand and map common statistical information and processes.