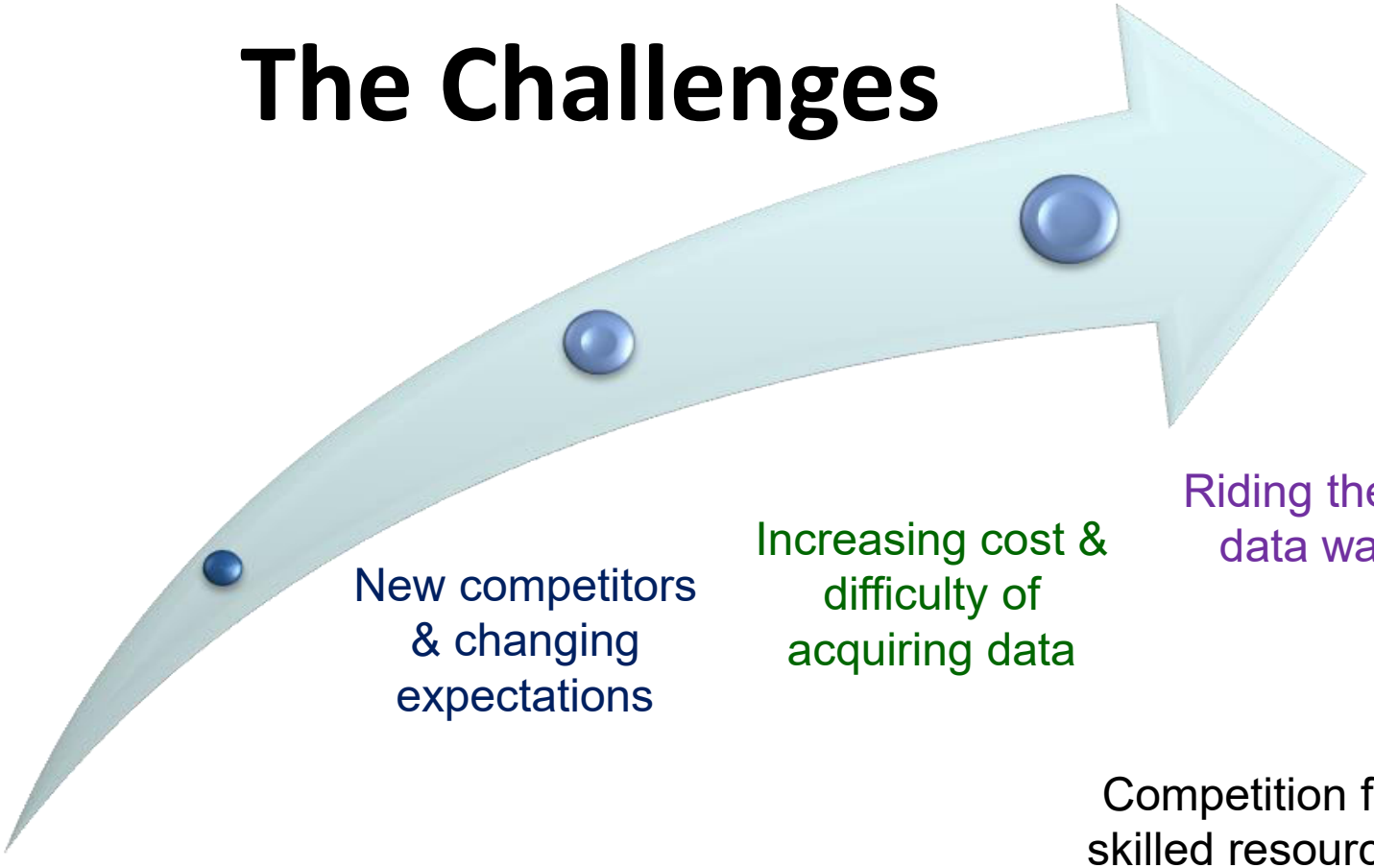
Generic Statistical Information Model (GSIM)

Thérèse Lalor and Steven Vale
United Nations Economic Commission for Europe
(UNECE)

GSIMPROJECT

The Challenges



New competitors & changing expectations

Increasing cost & difficulty of acquiring data

Riding the big data wave

Competition for skilled resources

Reducing budget

Rapid changes in the environment

GSIMPROJECT

These challenges are too big for statistical organisations to tackle on their own.

We need to work together

GSIMPROJECT

Response from Official Statistics

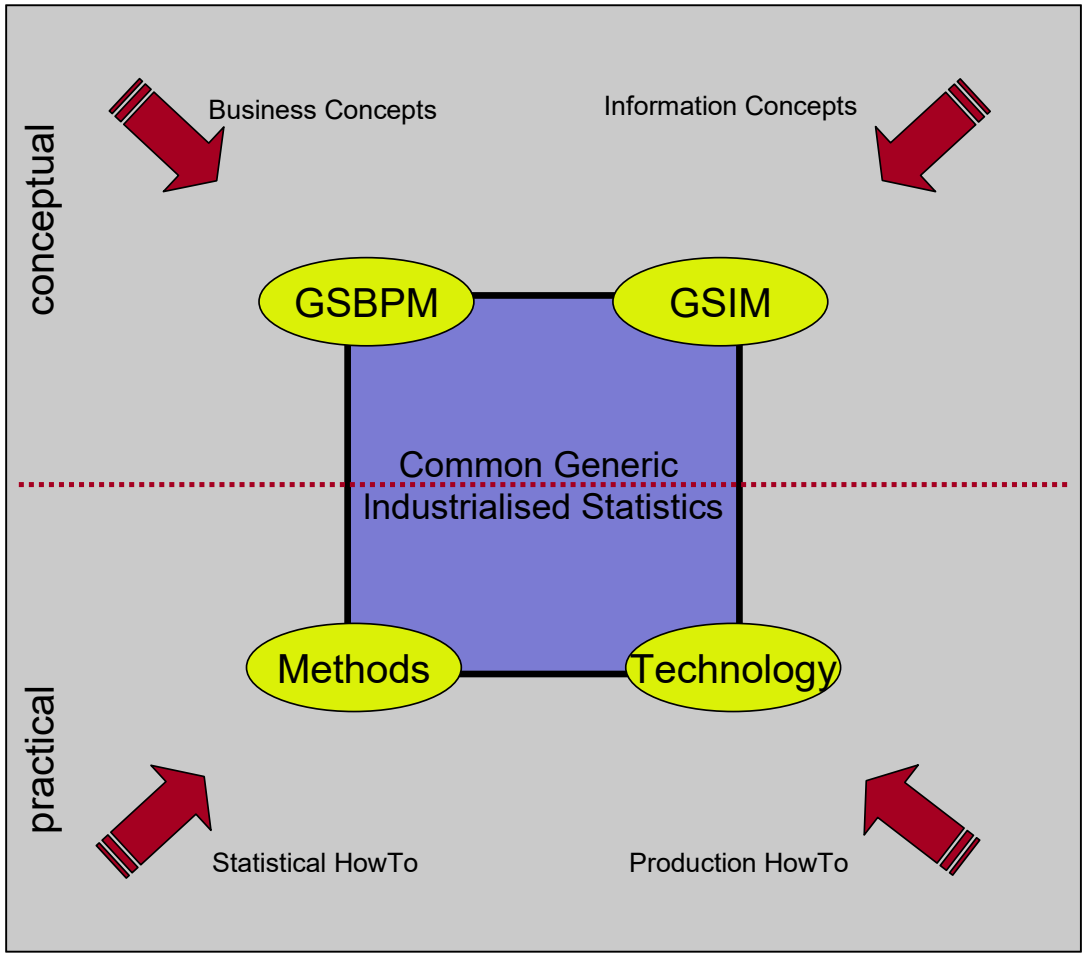
- A High Level Group consisting of 10 heads of national and international statistical organizations was created

GSIMPROJECT

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

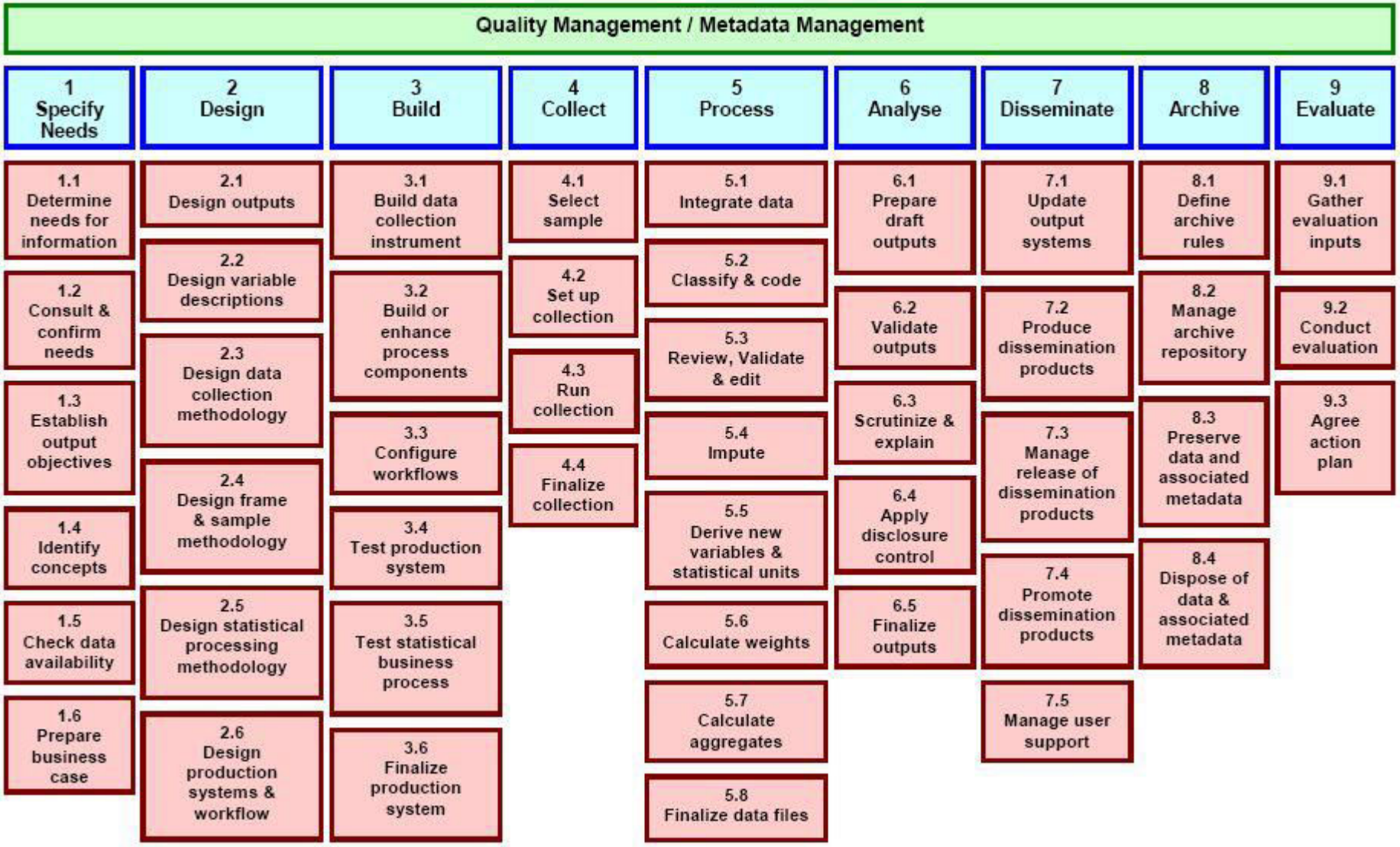
No domain is special!

GSIMPROJECT



GSIMPROJECT

The GSBPM

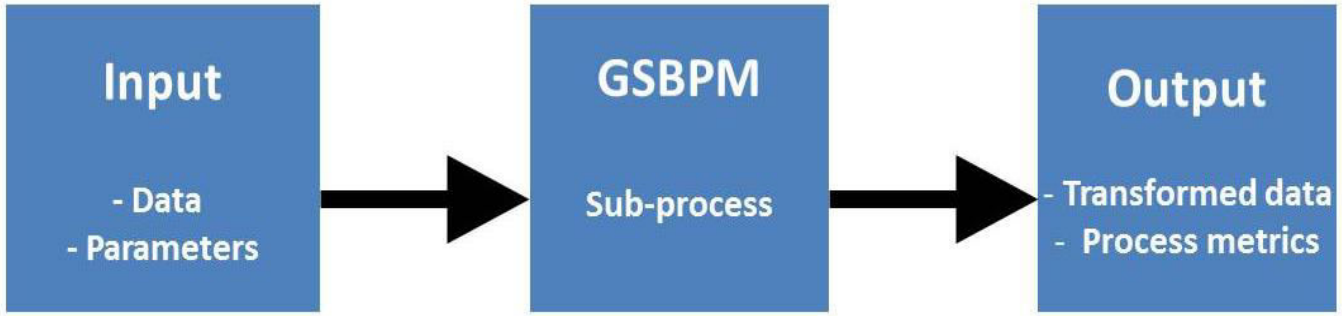

Quality Management / Metadata Management
1 Specify Needs
1.1 Determine needs for information
1.2 Consult & confirm needs
1.3 Establish output objectives
1.4 Identify concepts
1.5 Check data availability
1.6 Prepare business case
2 Design
2.1 Design outputs
2.2 Design variable descriptions
2.3 Design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

The GSBPM is used by more than 50 statistical organizations worldwide to manage and document statistical production

GSIMPROJECT

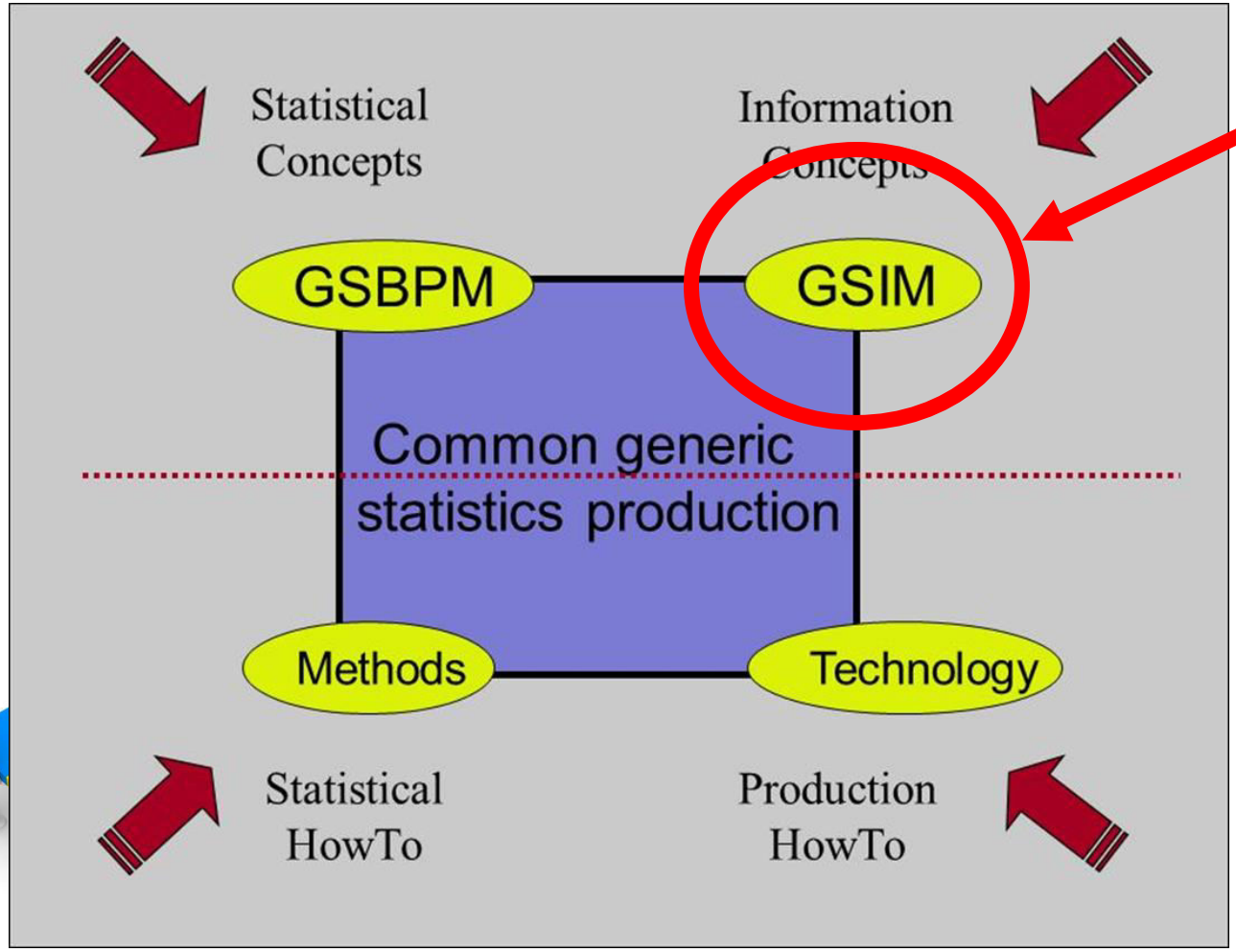
GSIM is complementary to GSBPM

escribe information objects and flows within th



GSIMPROJECT

Introducing the GSIM



You are here

What is GSIM?

A reference framework of information objects

- It sets out definitions, attributes and relationships regarding information objects

- It aligns with relevant standards such as DDI and SDMX

GSIMPROJECT

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

GSIMPROJECT

GSIM is a conceptual model: It is a new way of thinking for statistical organizations

GSIMPROJECT

GSIM enables:

- Communication
- Coordination
- Cooperation
- Collaboration

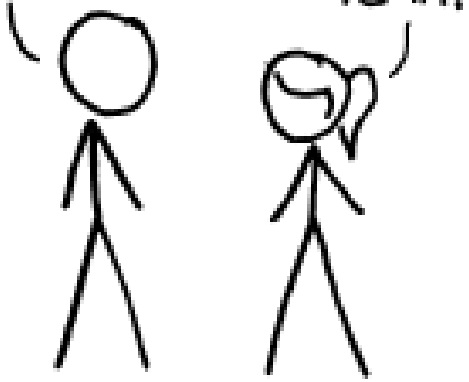
GSIMPROJECT

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.



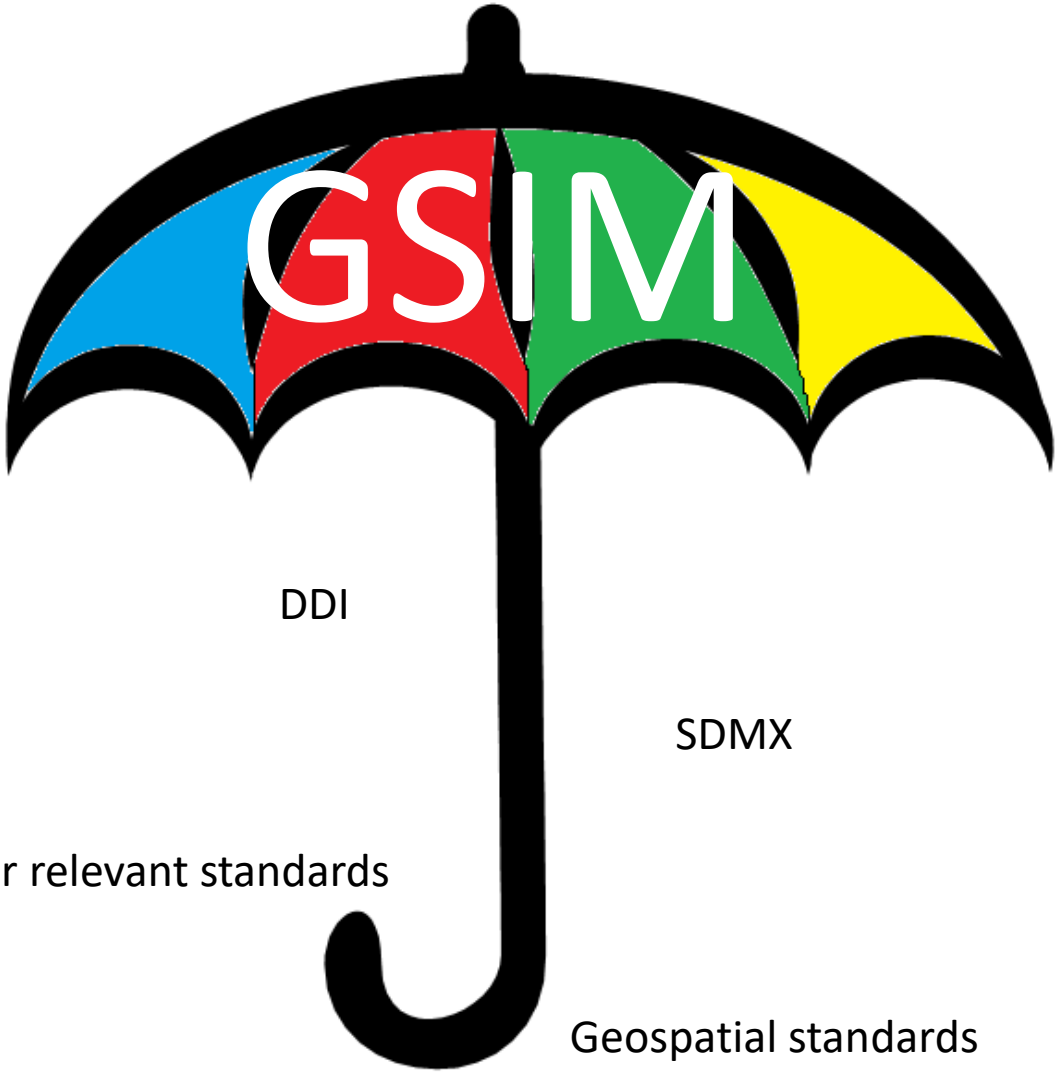
SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

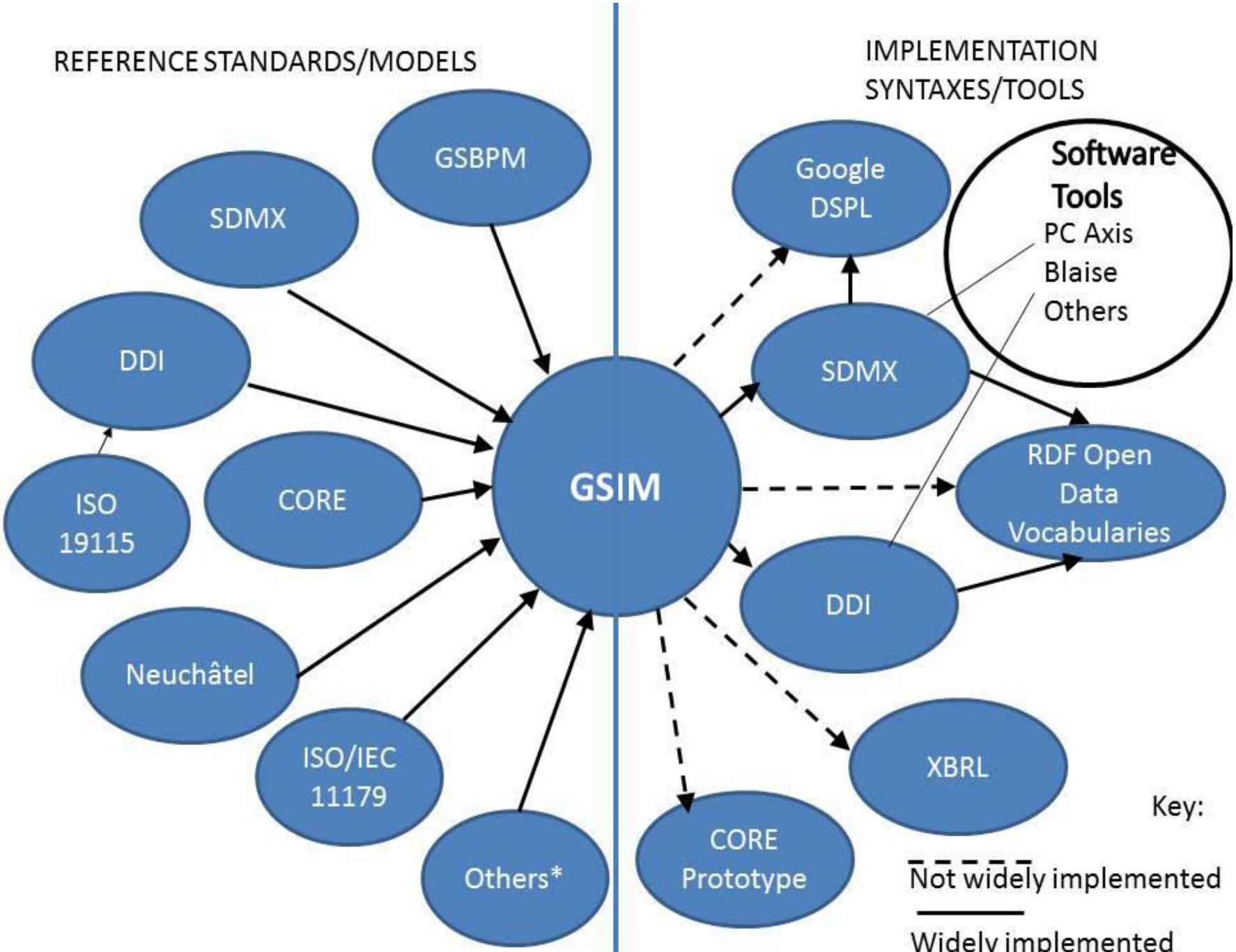
GSIMPROJECT

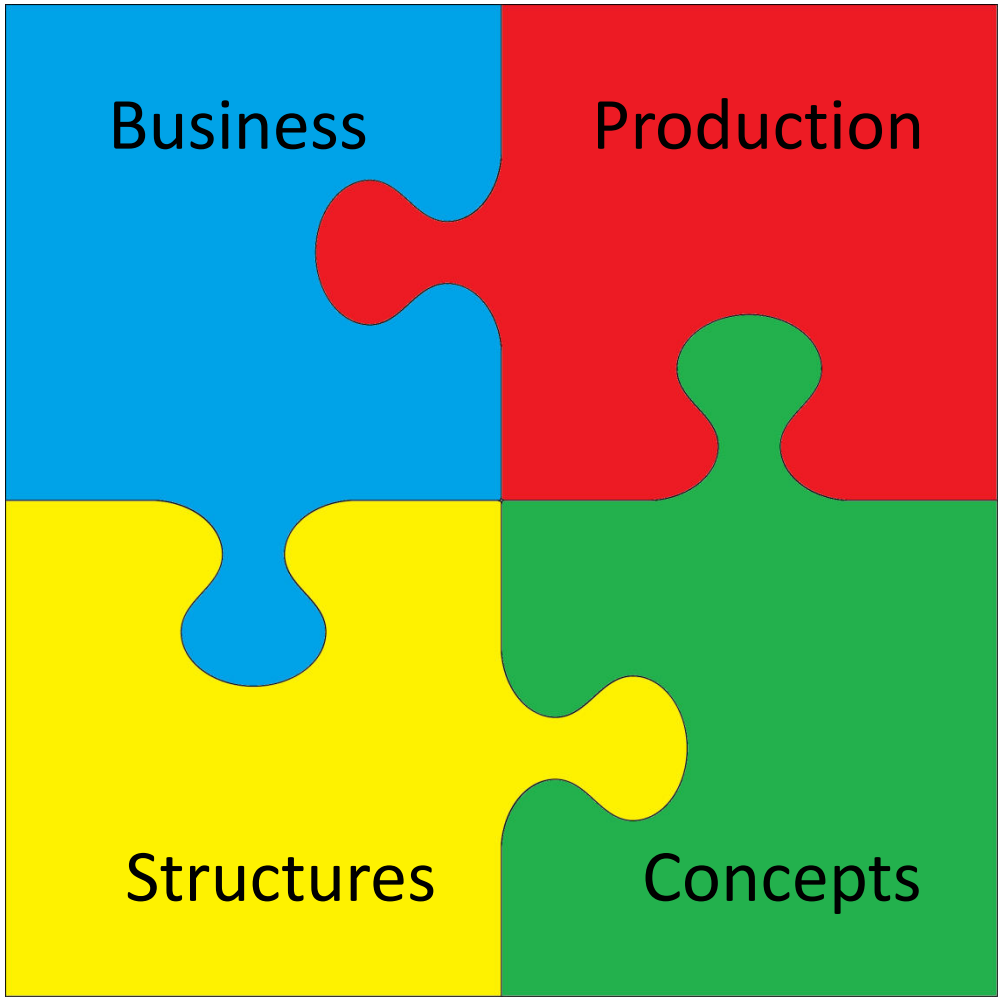
Conceptual model

Implementation standards

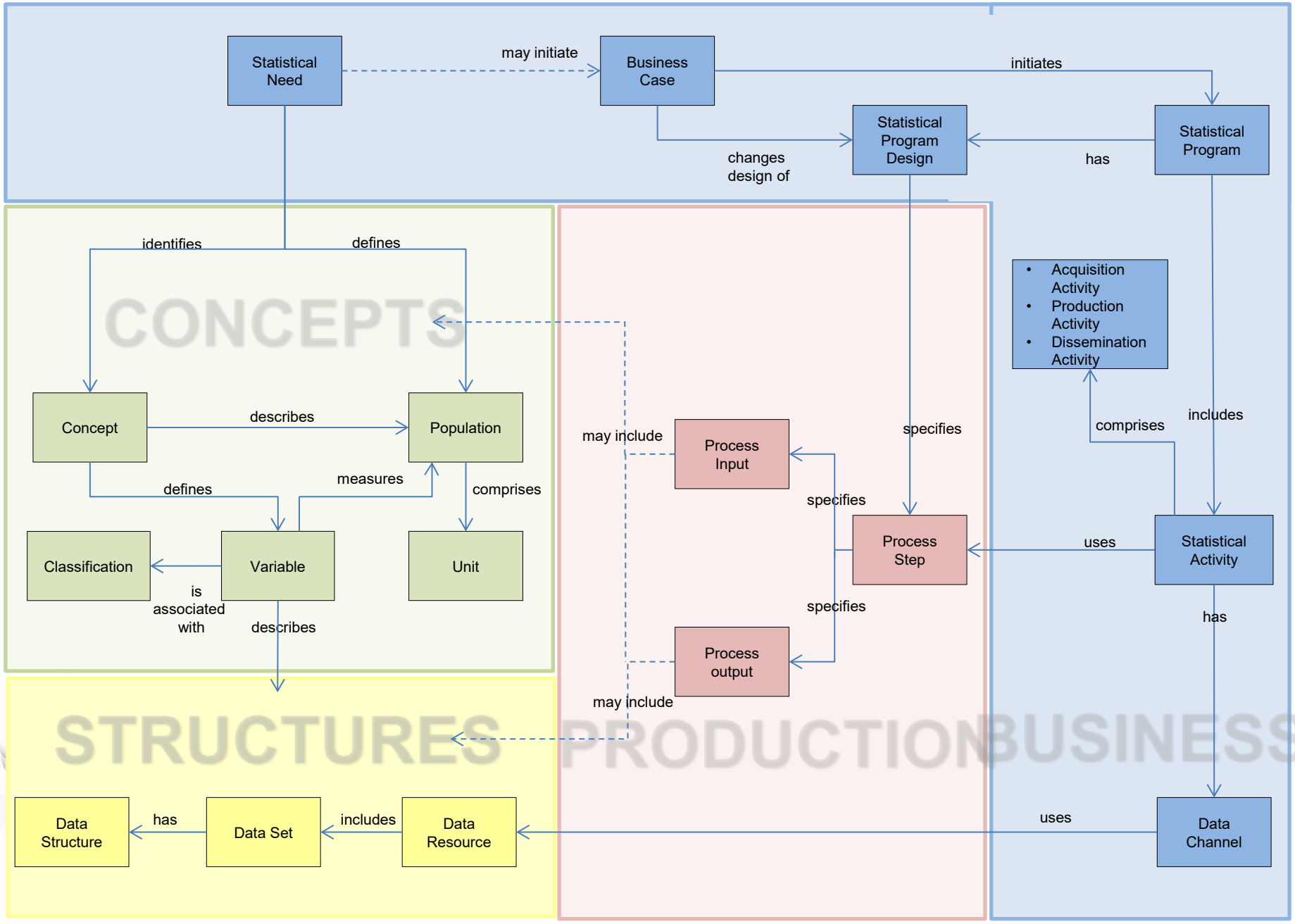


GSIMPROJECT


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



GSIMPROJECT


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
Acquisition Activity
Production Activity
Dissemination Activity
comprises
includes
specifies
may include
Process Input
specifies
Process Step
uses
Statistical Activity
specifies
has
Process output
may include
STRUCTURES
PRODUCTION
BUSINESS
Data Structure
has
Data Set
includes
Data Resource
uses
Data Channel

GSIM: The “sprint’ approach

- The HLG-BAS decided to accelerate the development of the GSIM
- “Sprints” – 2 week workshops for 10-12 experts (IT, methodology, statistics, ...)
- Sprint 1 – Slovenia, February 2012
- Sprint 2 – Republic of Korea, April 2012
- Integration Workshop, Netherlands, September 2012

GSIMPROJECT

Moving 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

GSIMPROJECT

Moving to GSIM in practice

- 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.

GSIM PROJECT

GSIM v1.0

- Released in December 2012
- We need people to use GSIM “in anger”. Then we will know how best to improve it.

GSIMPROJECT

More information

- GSIM
  - http://www1.unece.org/stat/platform/display/metis/Generic+Statistical+Information+Model+(GSIM)
- Poster Session @ EDDI
  - Today at lunch time

GSIMPROJECT