Semantics, linked data and interoperability in Statistics

Semantic Days

Oslo, 8 June 2011

Rune Gløersen
IT Director
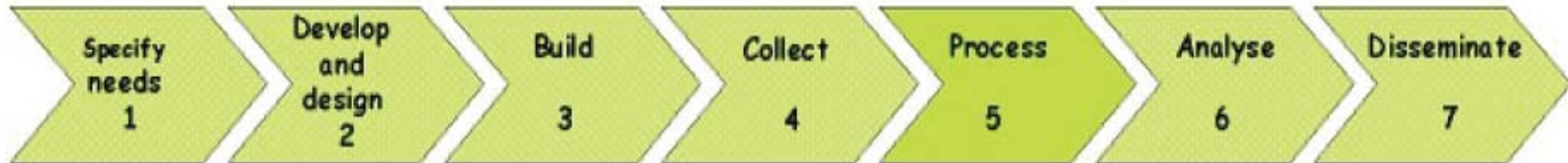Statistics Norway

Contents

- The characteristics of processes and data at NSIs
- Applicable standards for various business processes
- Statistics and linked open data
- The preconditions for increased interoperability
- Industrialisation of statistics

Statistical Business Process Model
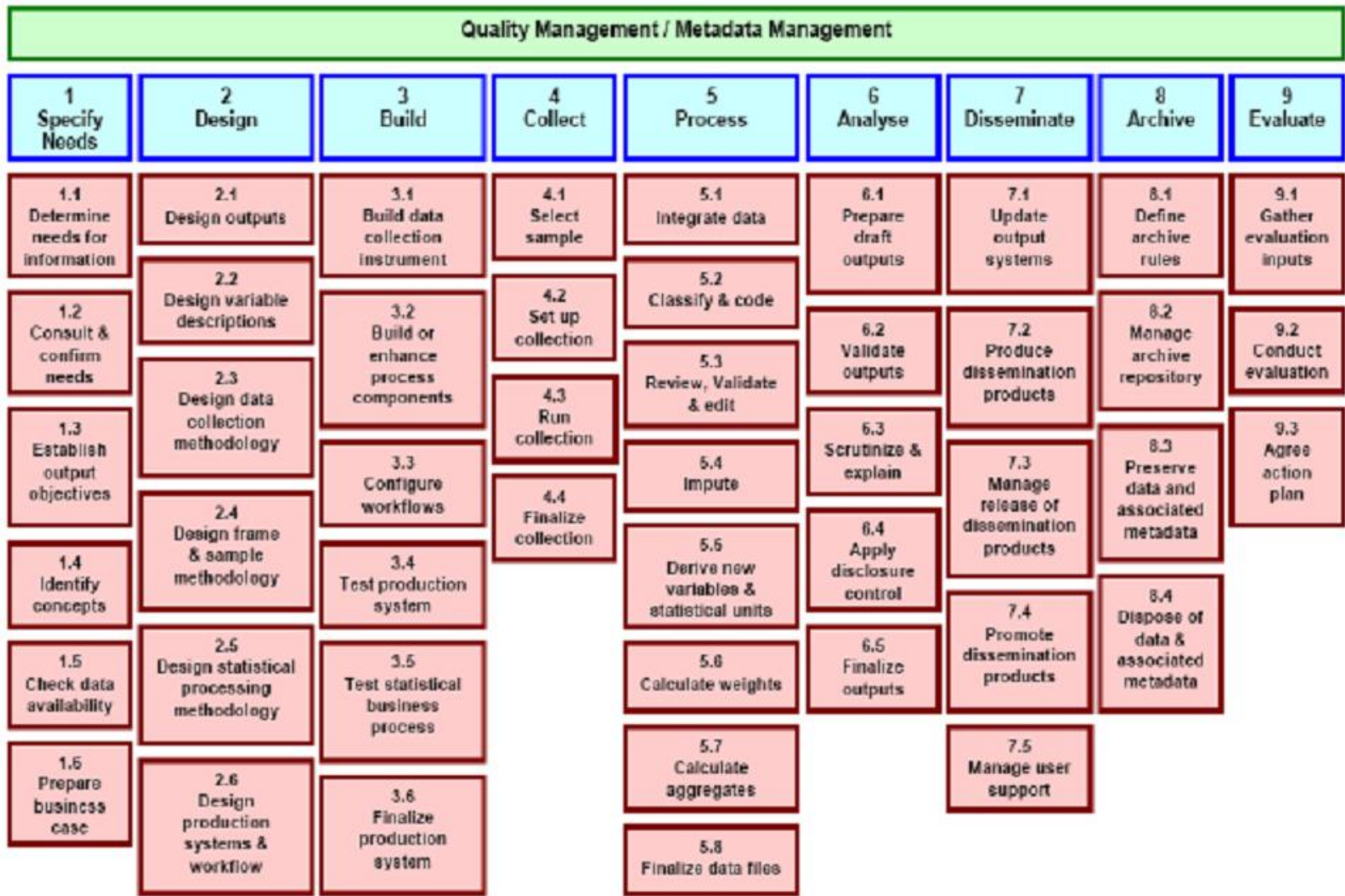## Statistics Norway

Generic Statistical Business Process Model
Internationally agreed reference model

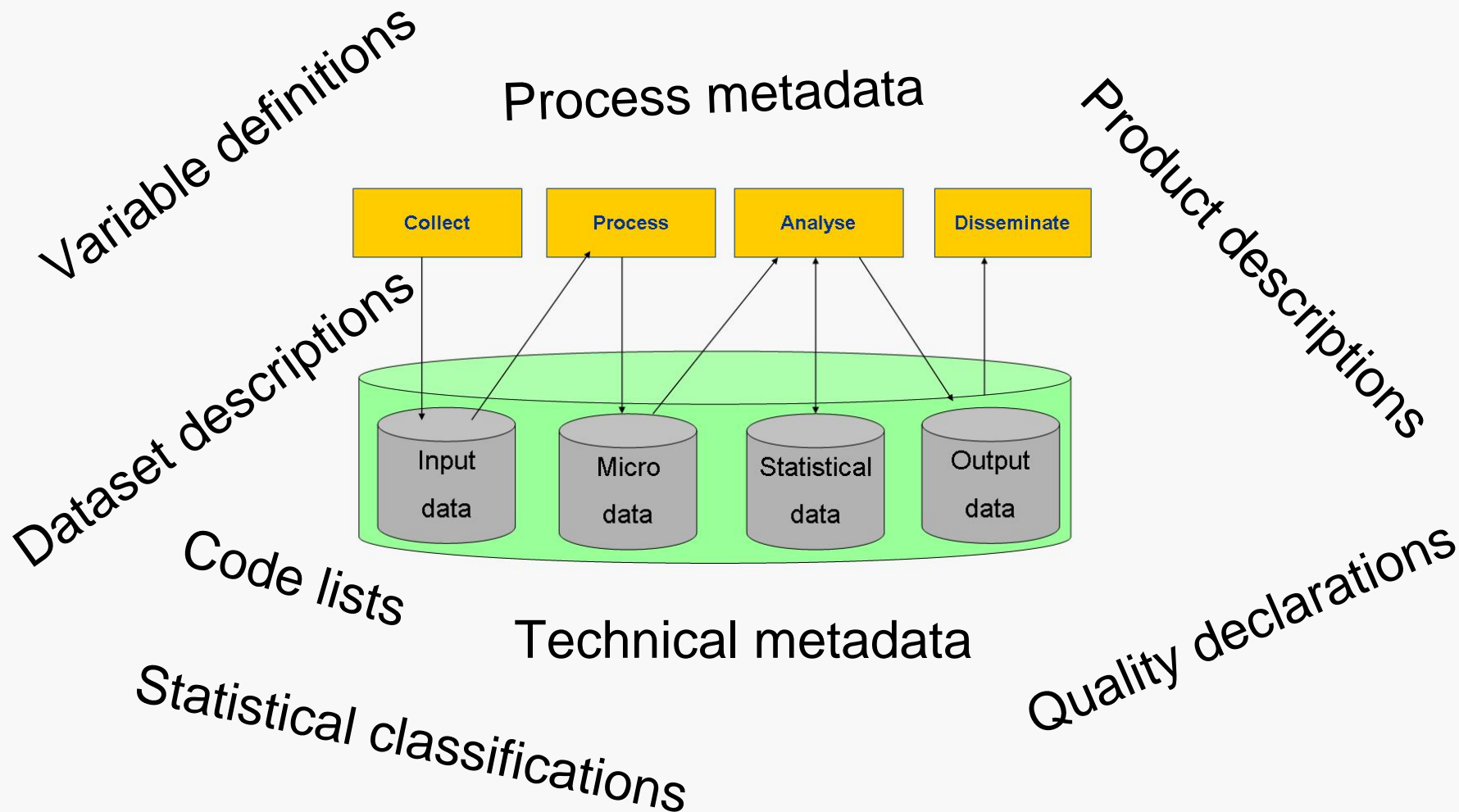| Quality Management / Metadata Management								
1 **Specify Needs**	**2** **Design**	**3** **Build**	**4** **Collect**	**5** **Process**	**6** **Analyse**	**7** **Disseminate**	**8** **Archive**	**9** **Evaluate**
1.1 Determine needs for information	2.1 Design outputs	3.1 Build data collection instrument	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				

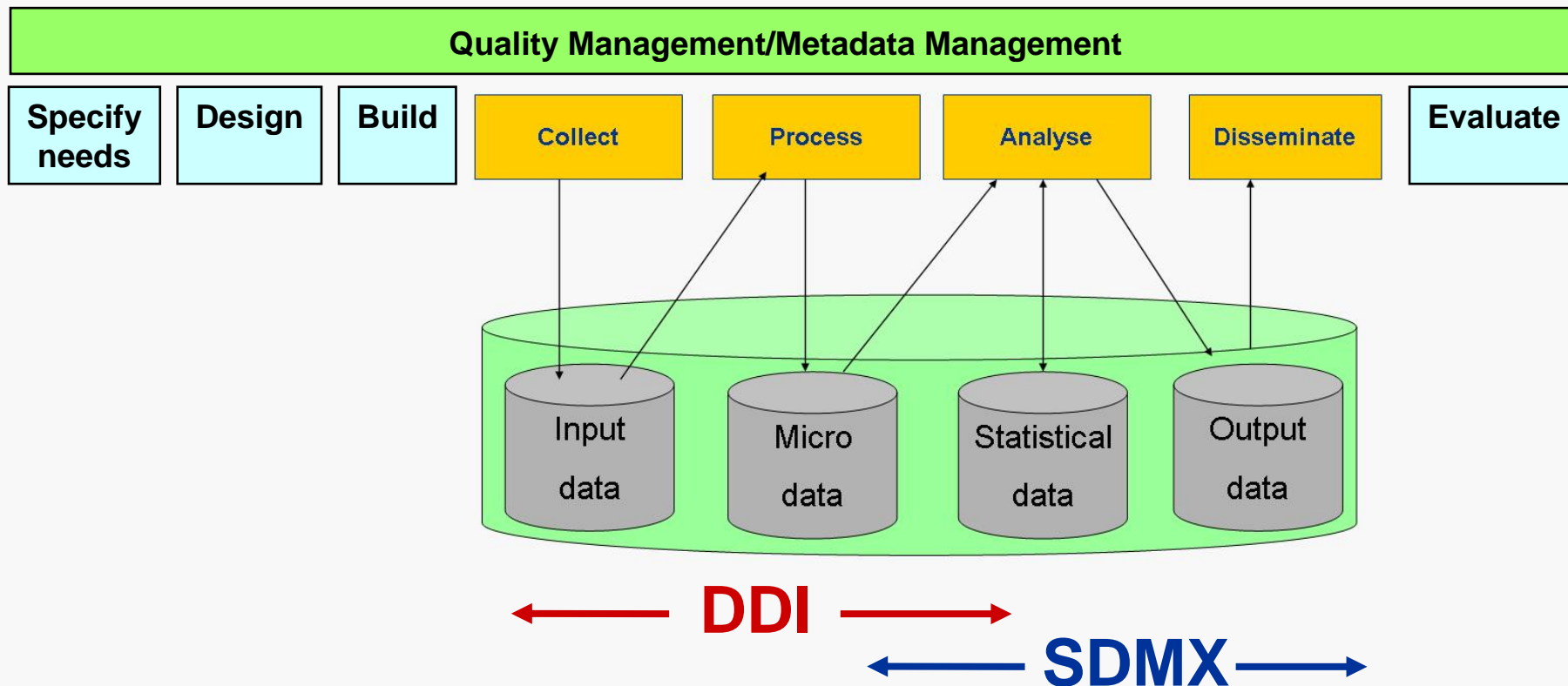# Process stages and data archiving
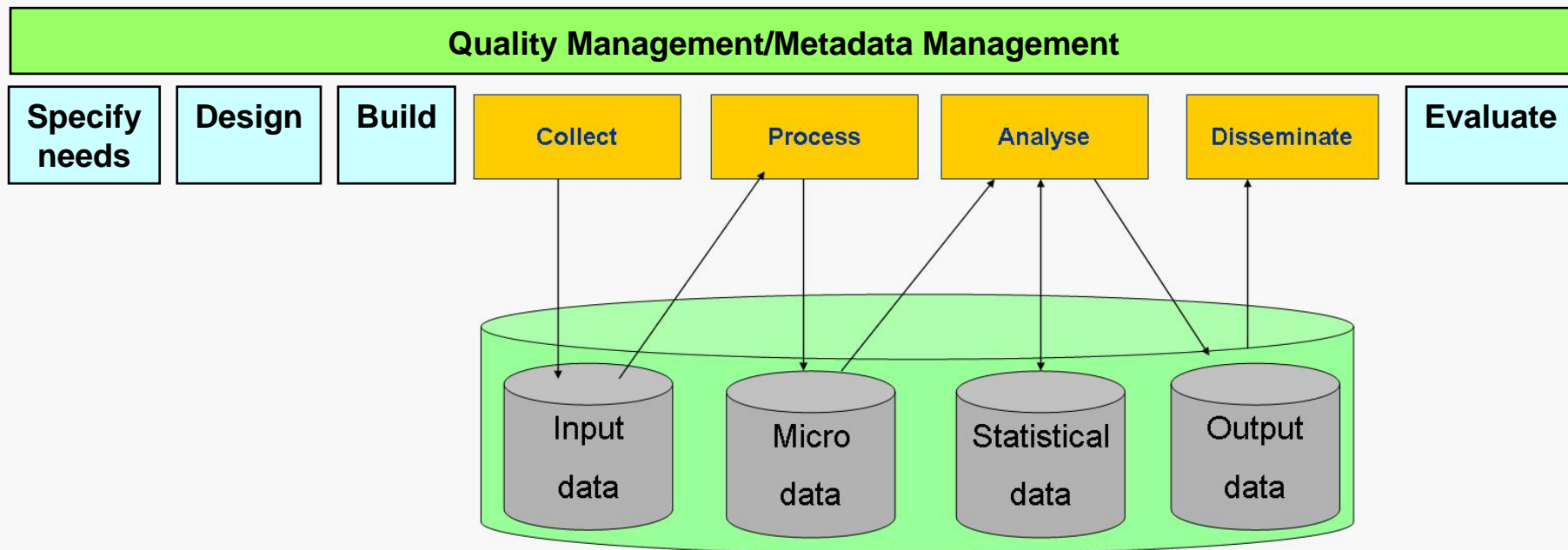
Data archiving spans the 4 main business processes, and comprises 4 steady states of the data life cycle

The variety of associated metadata

Adopting standards

Quality Management/Metadata Management

| Specify needs | Design | Build | Collect | Process | Analyse | Disseminate | Evaluate |

Input data — Micro data — Statistical data — Output data

DDI ⟶

SDMX ⟶

DDI: Data Documentation Initiative; www.ddialliance.org

SDMX: Statistical Data and Metadata Exchange; www.sdmx.org

The diversity of users, needs and data flows
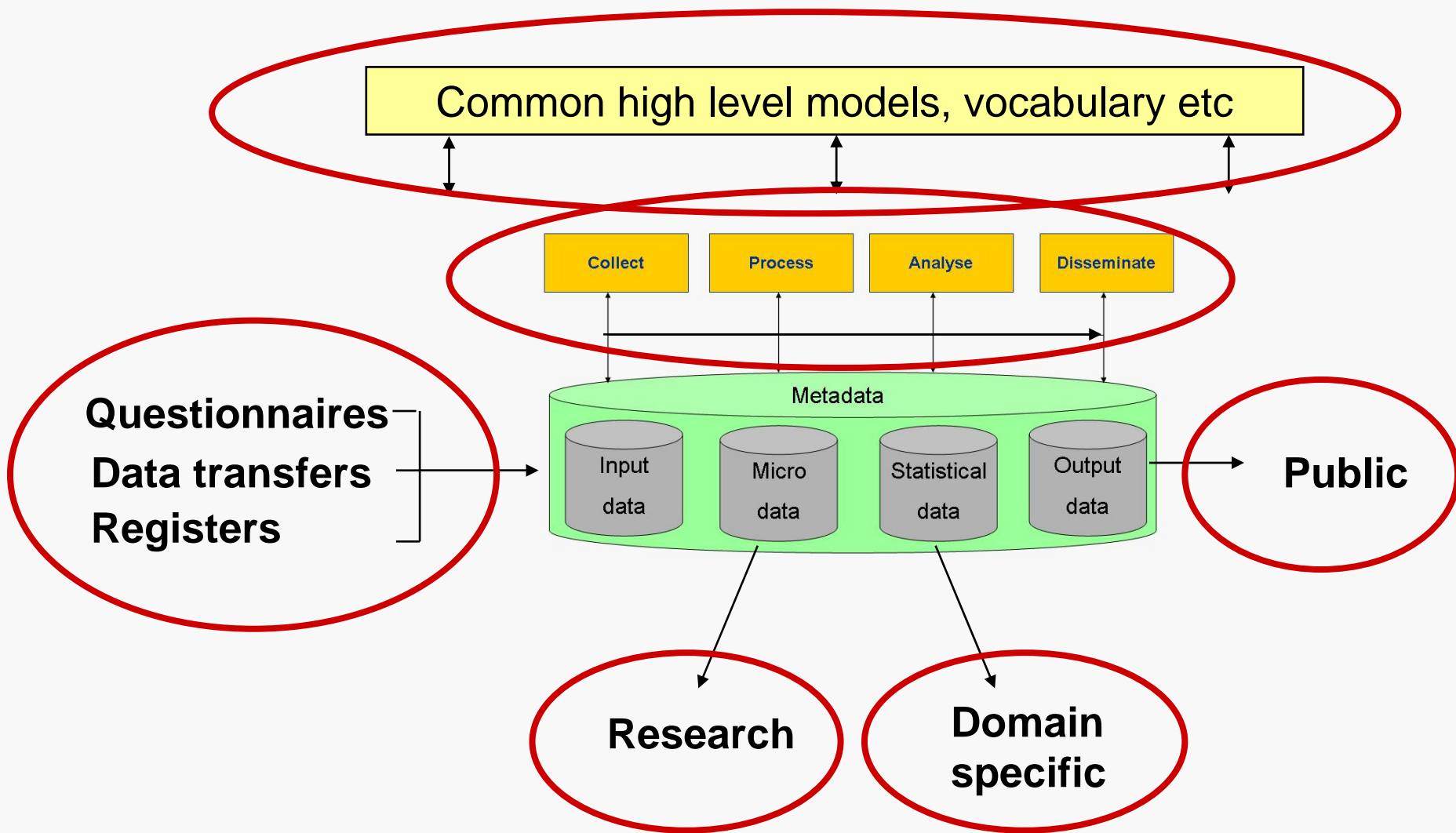
Dissemination of aggregated statistics using SDMX
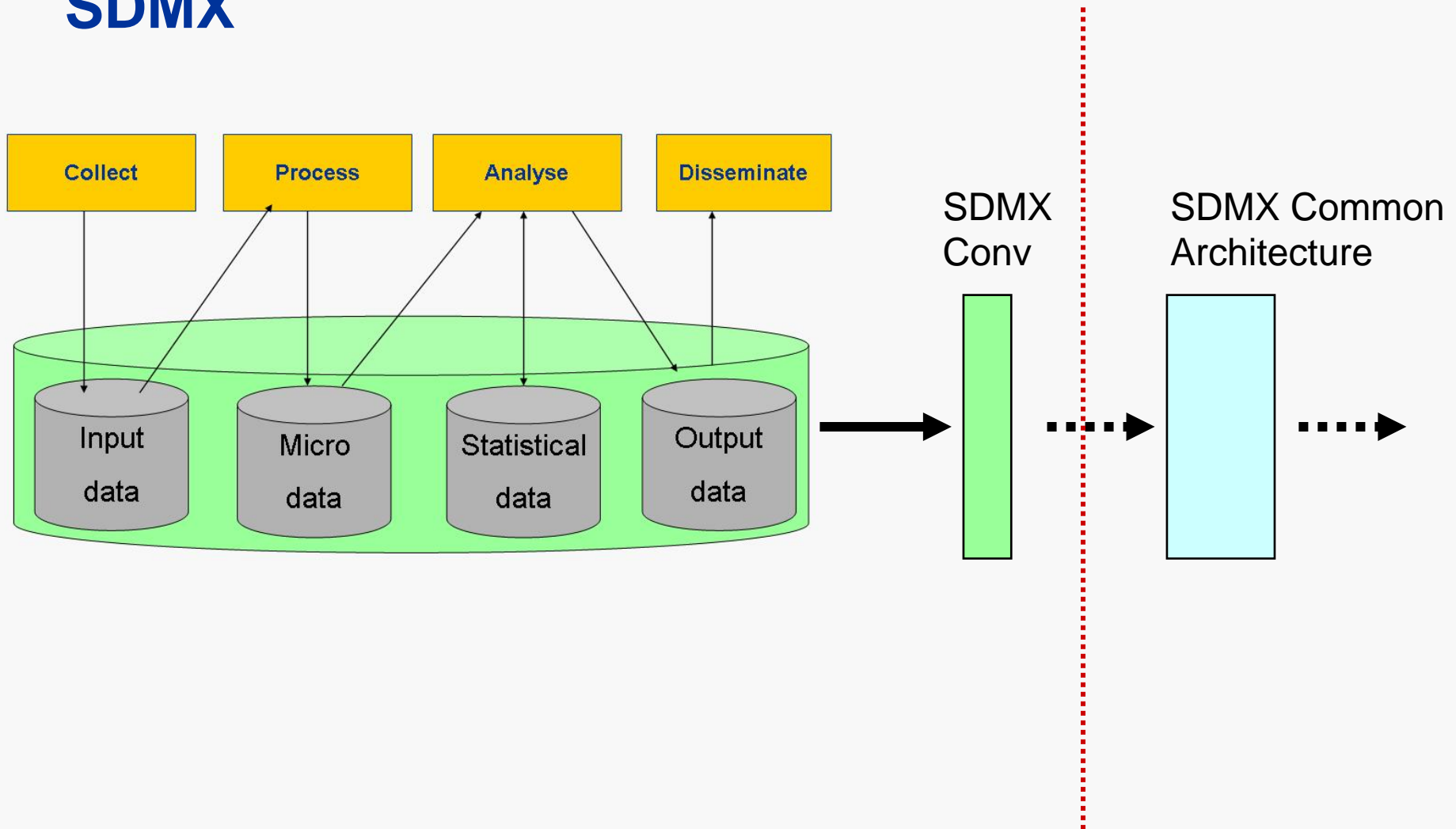
Statistics and Linked Open Data

SDMX Common Architecture
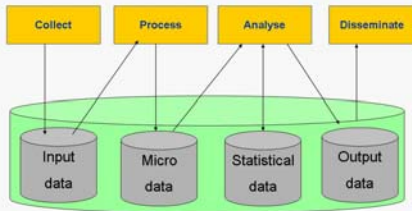
RDF

Linked Open Data
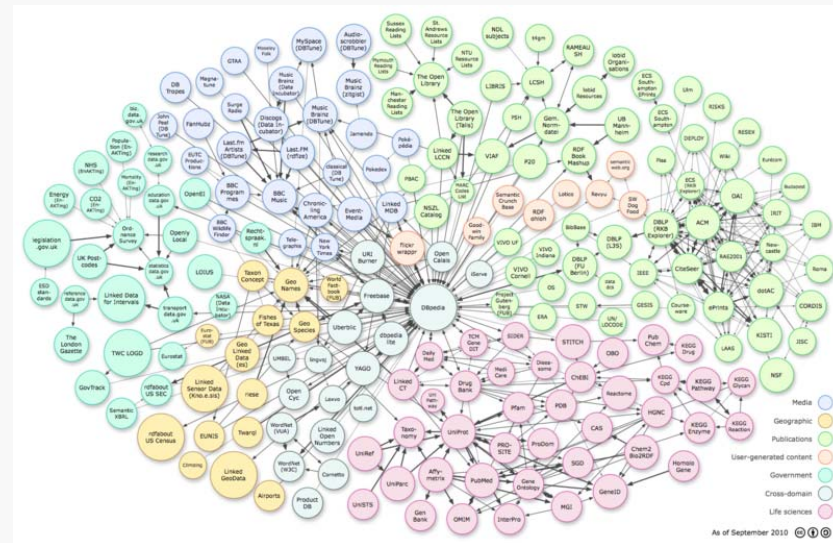
SCOVO

Using SDMX to refine the simple concept of a cube or a table
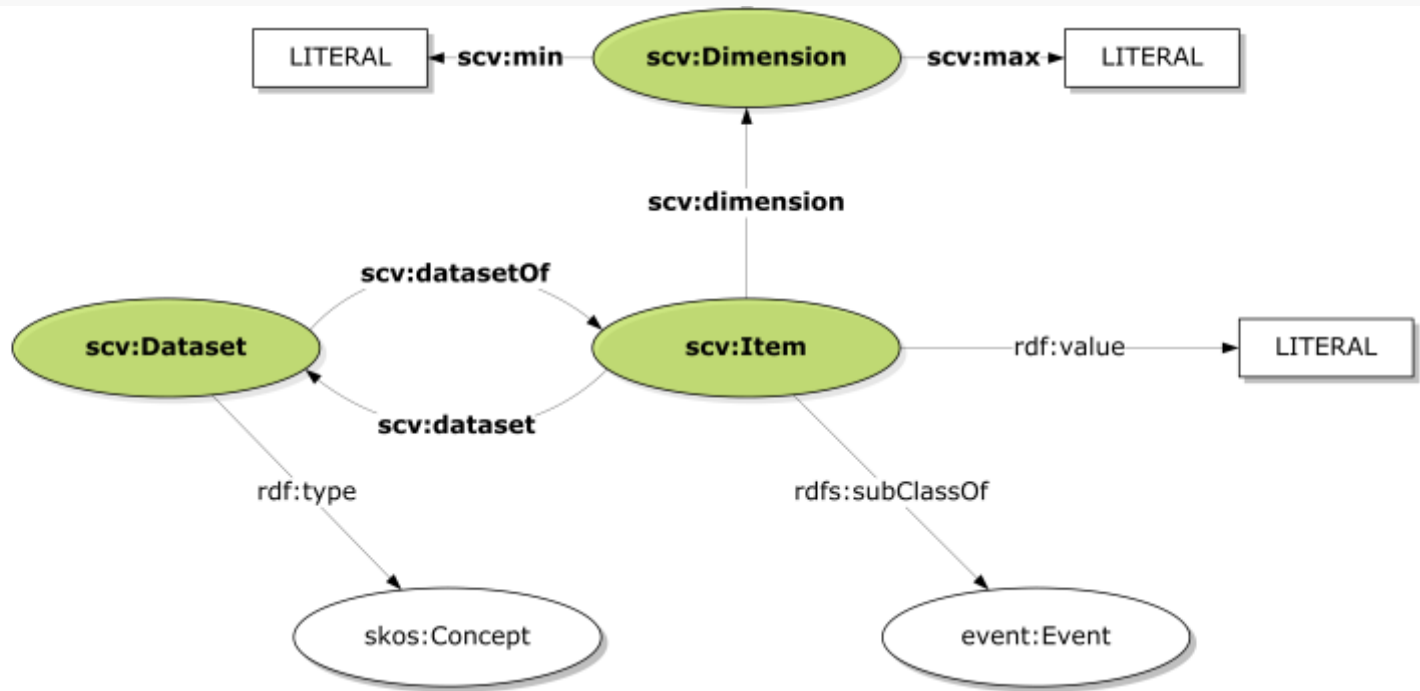
The Statistical Core Vocabulary (scovo)
http://purl.org/NET/scovo
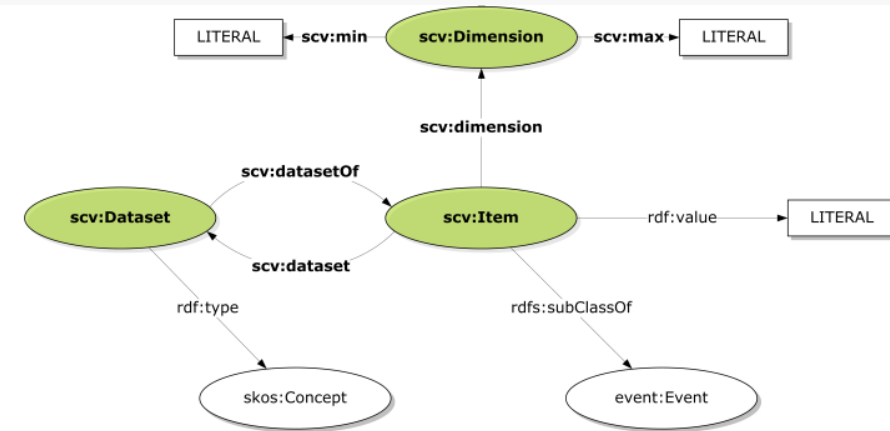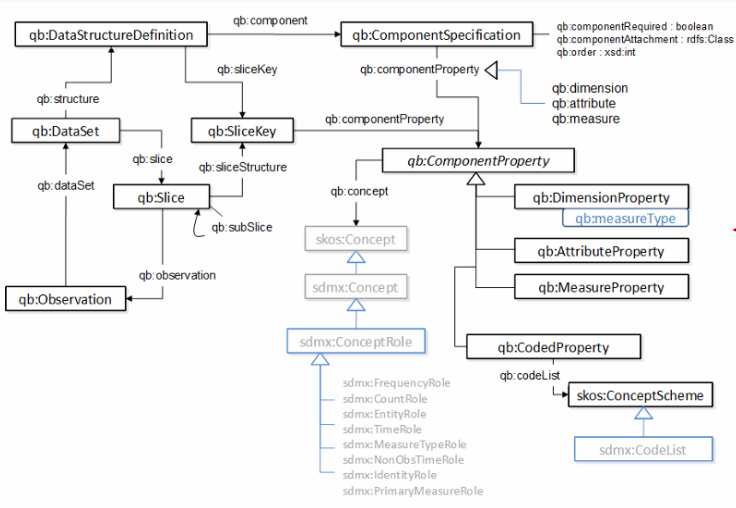v0.3@2008-05-15

scv: <http://purl.org/NET/scovo#>
event: <http://purl.org/NET/c4dm/event.owl#>
skos: <http://www.w3.org/2004/02/skos/core#>
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
rdfs: <http://www.w3.org/2000/01/rdf-schema#>

Providing statistical ontology to the semantic web

SDMX Information Model and Metadata Common Vocabulary

Statistical Core Vocabulary (SCOVO)

Utilizing available data

Google Search Trends
(Distributed according to the NAICS, the statistical classification of retail trade categories within the US)

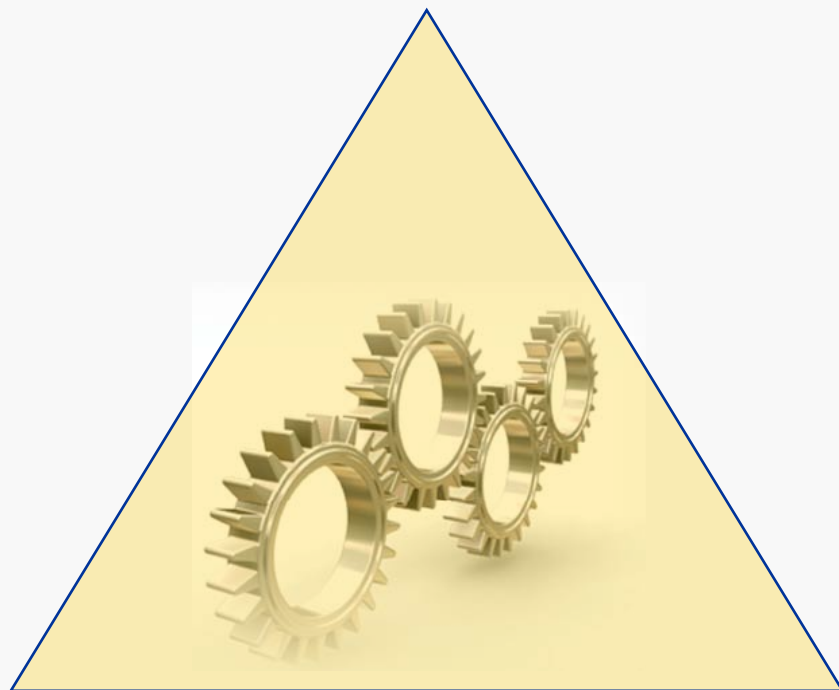Coming soon (?)
Google Price Index
(GPI)

Improved interoperability
Some trends within the statistical community

Organisational
interoperability

Semantical
interoperability

Technological
interoperability

Enterprise Architecture
Coherence and interoperability
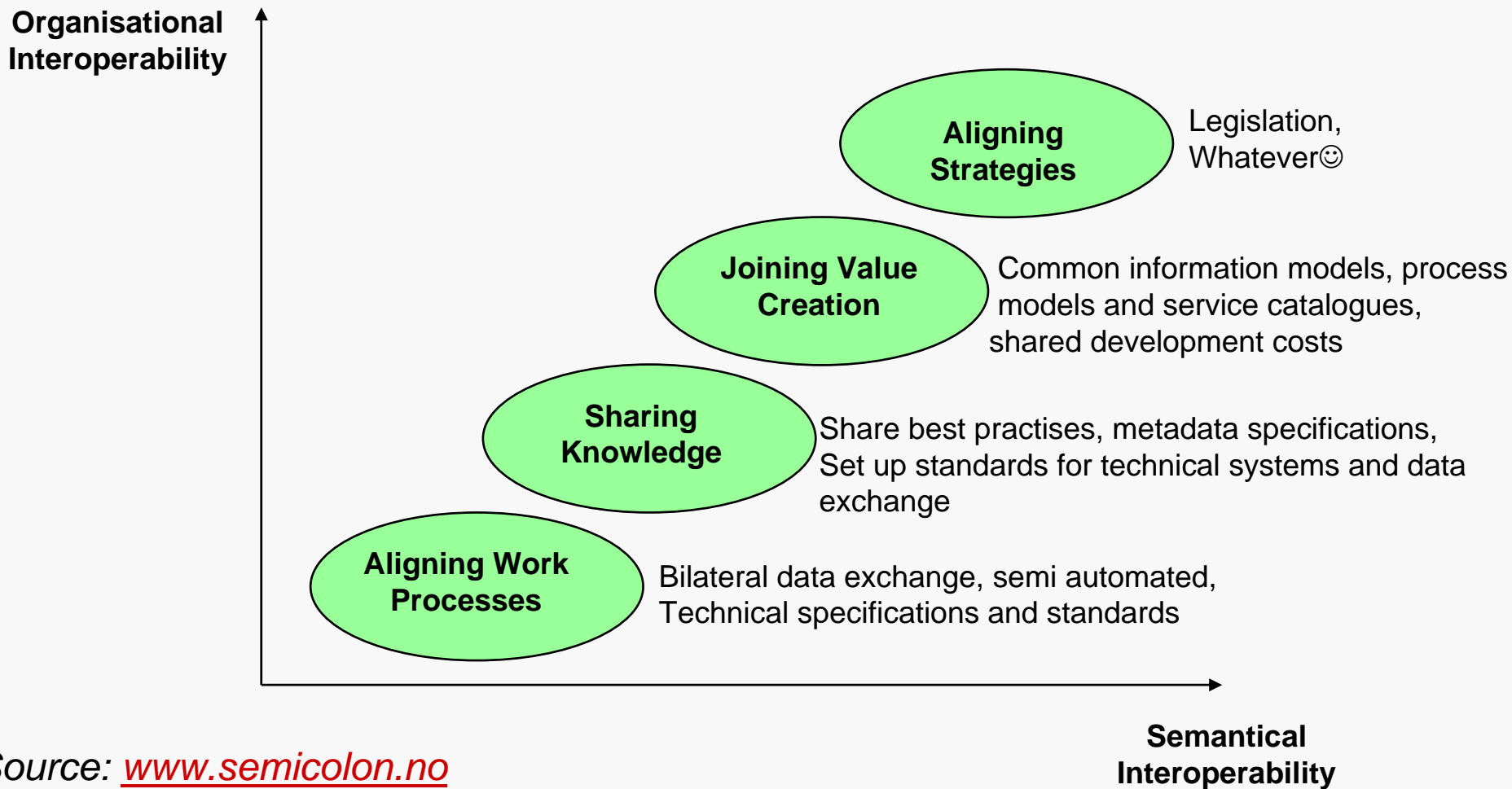
Generic Statistical
Business Process Model

Best Practice
Statistical Methods

Generic Statistical
Information Model

ICT-
Architecture
(Principles)

Maturity growth in e-Government

Organisational Interoperability

Aligning Strategies

Legislation, Whatever☺

Joining Value Creation

Common information models, process models and service catalogues, shared development costs

Sharing Knowledge

Share best practises, metadata specifications, Set up standards for technical systems and data exchange

Aligning Work Processes

Bilateral data exchange, semi automated, Technical specifications and standards

Semantical Interoperability

Source: www.semicolon.no
Analytical Framework for e-Government Interoperability

Industrializing Statistics
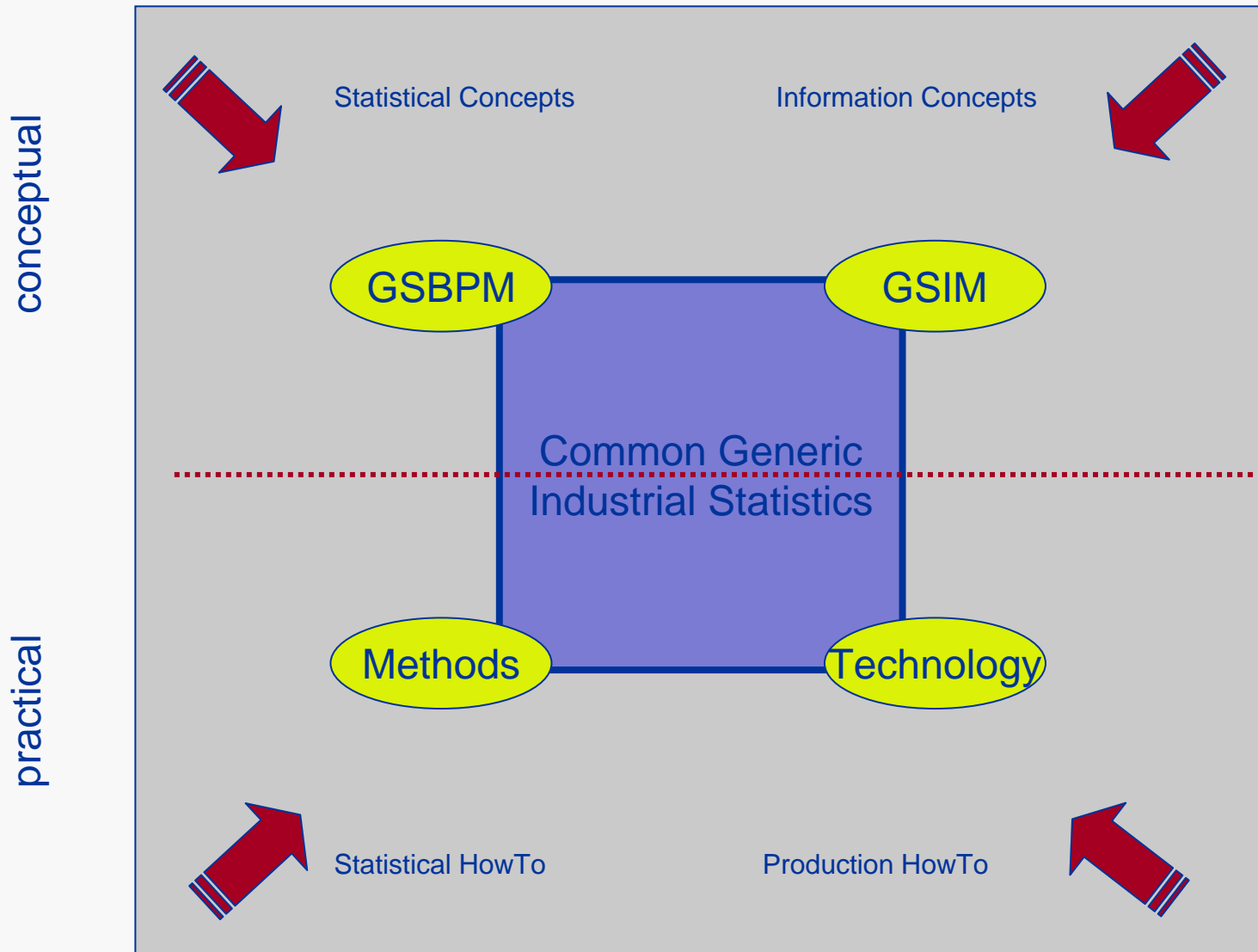
Some conclusions

- Statistics is a key component of the semantic web and linked data, and the developments of the semantic web are crucial to refining statistics and statistics production

- The semantic web developments will bridge data collection and data dissemination in statistics

- Innovation in data integration will be emphasized

- Statistical classifications will be exploited outside the statistics community

- More communities/sectors/branches will collaborate on the development of standards and compete heavily on the implementations

Thank you for your attention!