

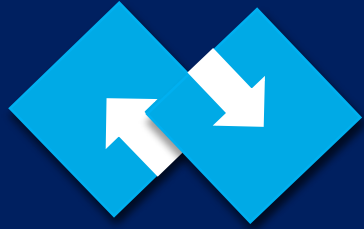
# Connecting SDMX and GSBPM

ModernStats World Workshop

27-30 October 2020

Prepared by Juan Muñoz López





# SDMX (ISO IS 17369)



SDMX stands for Statistical Data and Metadata eXchange.



Is an ISO standard designed to:



Describe statistical data and metadata



Normalize and improve the exchange of statistics across organizations.



Enable interoperable implementations within and between systems.



Improve exchange, reporting and dissemination of statistical data and their related meta-information.

# SDMX Sponsors

- The sponsors of the SDMX initiative are:
  - BIS, Bank of International Settlements
  - ECB, European Central Bank
  - Eurostat, The Statistical office of the European Union
  - IMF, International Monetary Fund
  - OECD, Organisation for the Economic Cooperation and Development
  - UNSD, United Nation Statistical Division
  - World Bank



# Statistical Metadata

SDMX deals with:

- **Structural metadata:** Defines the structure of statistical data sets and metadata sets (e.g. names of variables or dimensions of a statistical cube). Data must be linked to structural metadata; without these data identifiers and descriptors, they cannot be identified, retrieved or browsed.
- **Reference metadata:** set of concepts that describe and qualify statistical data sets and processing more generally, and which are often associated not with specific observations or series of data, but with entire collections of data or even the institutions which provide that data

## Source data for tables and graphs

- [Gender statistics - excel file](#) 

## Reference Metadata

### Data sources

Eurostat produces and disseminates a number of datasets that show how men and women compare in areas such as education, labour market, earnings, social inclusion and health in the EU. The most relevant and most frequently used datasets are listed in the ['Equality'](#) domain. For more information on data sources and availability, see the metadata files linked to the multidimensional tables or in other relevant articles.

**Gender overall earnings gap** 'Gender overall earnings gap', is calculated as follows:

$$GOEG = \frac{(E_m \times H_m \times ER_m) - (E_w \times H_w \times ER_w)}{(E_m \times H_m \times ER_m)} \times 100$$

where GOEG means Gender overall earnings gap, Em — Mean hourly earnings of men, Hm — Mean monthly hours paid to men, ERm — Employment rate of men (aged 15-64), Ew — Mean hourly earnings of women, Hw — Mean monthly hours paid to women and ERw — Employment rate of women (aged 15-64).

### Context

Gender statistics are indispensable for identifying inequalities between women and men, and needed for the purposes of gender policy development and implementation at global, European and national levels. [Four world conferences on women convened by the United Nations](#) between 1975 and 1995 have been crucial in putting the cause of gender equality at the very centre of the global agenda. In 1995, the Fourth World Conference on Women held in Beijing adopted the [Declaration and Platform for Action](#).

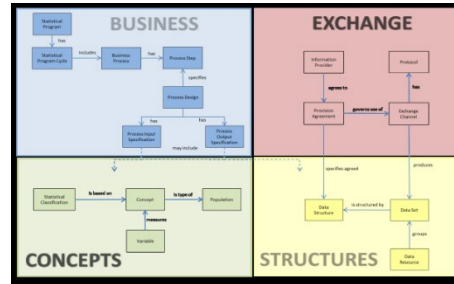
This specified critical areas of concern considered to represent the main obstacles to women's advancement, requiring concrete action by governments and civil society. These areas are as follows: women and poverty, education and training of women, women and health, violence against women, women and armed conflict, women and the economy, women in power and decision-making, institutional mechanisms for the advancement of women, human rights of women, women and the media, women and the environment and the girl-child.

Equality between women and men is a founding value of the EU (Article 2 of the [Treaty on European Union](#)) as well as a fundamental right (Article [19 of the Charter of Fundamental Rights](#) of the European Union). Following the 1995 conference in Beijing, the European Council requested an

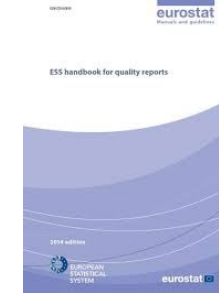
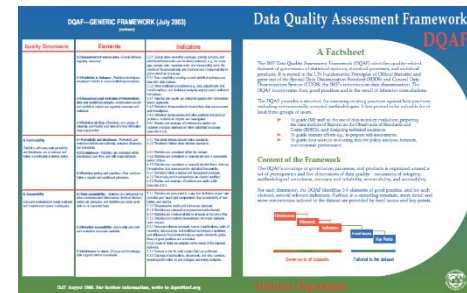
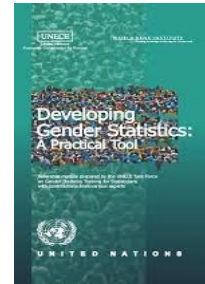


# Types of Reference Metadata

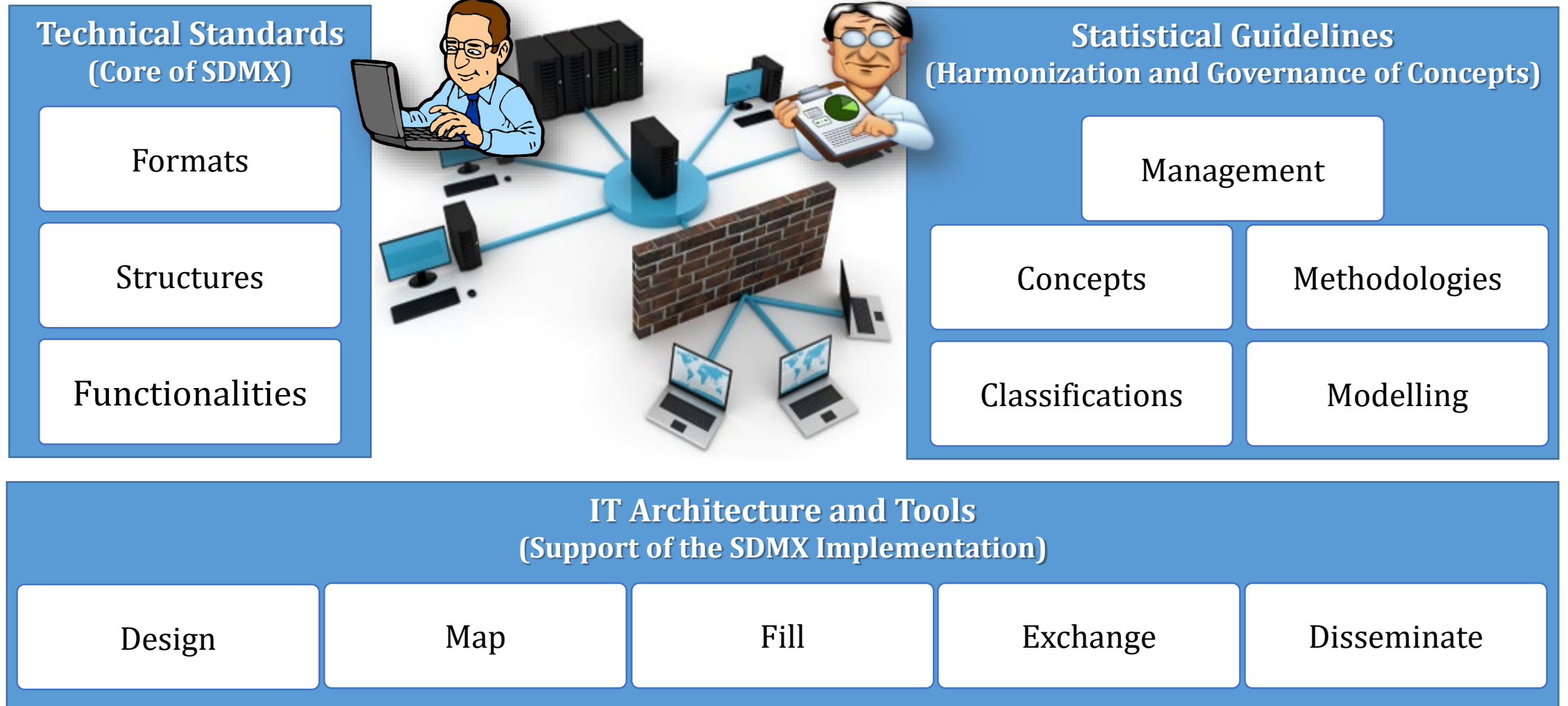
- Describes contents and quality of statistical data
- Includes:
  - **Conceptual Metadata.**
    - Concepts and practical implementation.
  - **Methodological Metadata.**
    - Methods used to generate the data.
  - **Quality Metadata.**
    - Quality dimensions of the resulting statistics.



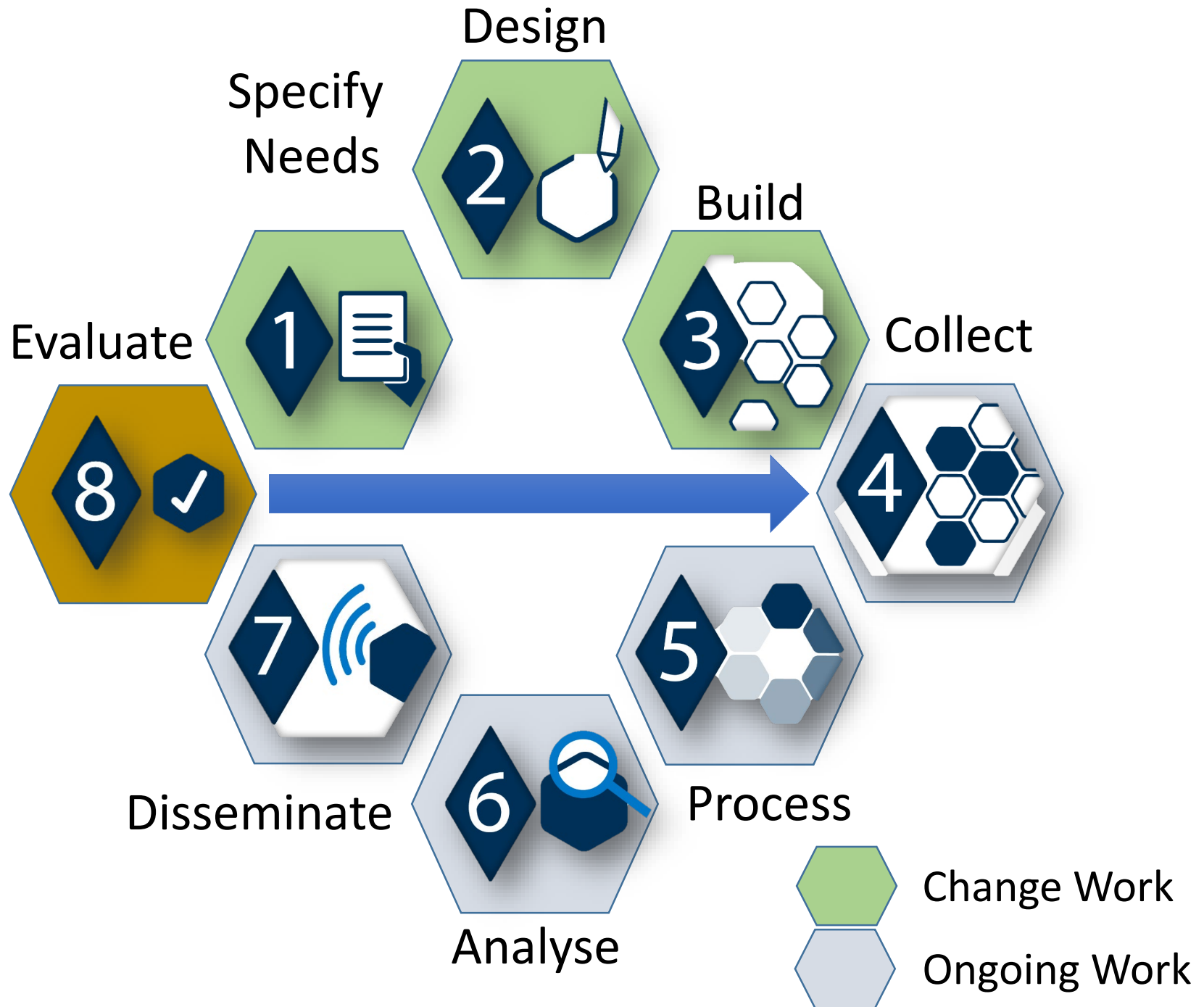
Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	1.1 Design inputs	1.1 Build release instrument	1.1 Create data and select sample	1.1 Integrate data	1.1 Prepare data output	1.1 Update output systems	1.1 Gather feedback
1.2 Collect & review needs	1.2 Design variable acceptances	1.2 Build or enhance data components	1.2 Install infrastructure	1.2 Control & validate	1.2 Validate outputs	1.2 Produce dissemination products	1.2 Conduct evaluation
1.3 Establish output objectives	1.3 Design selection	1.3 Build or enhance dissemination components	1.3 Real installation	1.3 Review & update	1.3 Integrate & system outputs	1.3 Manage release of dissemination products	1.3 Agree & action plan
1.4 Identify outputs	1.4 Design forms & design	1.4 Configure hardware	1.4 Finalize installation	1.4 QA & release	1.4 Apply disseminator format	1.4 Produce dissemination products	1.4 Manage user support
1.5 Check data availability	1.5 Design processing software	1.5 Real production system	1.5 Real production system	1.5 Deliver data, variables & data	1.5 Produce outputs	1.5 Manage user support	
1.6 Prepare business case	1.6 Design production systems & workflow	1.6 Test operational business process	1.6 Finalize production system	1.6 Calculate weights	1.6 Calculate aggregates	1.6 Produce data files	



# Main Components of SDMX



# GSBPM as a Cycle



# GSBPM Levels

Overarching Processes							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and 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 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

**Level 0: Process**

**Level 1: Phases, 8**

**Level 2: Sub-processes, 44**



Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs <b>Review of existent data and metadata flows</b>	2.1 Design outputs <b>Design data and metadata flows</b>	3.1 Reuse or build collection instruments <b>Select/Build DSDs, MSDs and VTL artefacts, and decide on SDMX supported patterns</b>	4.1 Create frame and select sample <b>Review and make use of available SDMX data and metadata flows</b>	5.1 Integrate data <b>Populate databases from SDMX dataflows, update the registry</b>	6.1 Prepare draft outputs <b>Internal use of SDMX datasets</b>	7.1 Update output systems <b>Automated by SDMX, just ensure that it is connected to dissemination systems</b>	8.1 Gather evaluation inputs <b>Use logs from ICT infrastructure (like queries to web services) if available</b>
1.2 Consult and confirm needs <b>Stablish provision agreements</b>	2.2 Design variable descriptions <b>Design Concept Schemes, DSDs and MSDs</b>	3.2 Reuse or build processing and analysis components <b>Check for available tools in the SDMX.ORG website</b>	4.2 Set up collection <b>Run pilots to check SDMX dataflows</b>	5.2 Classify and code <b>Update codelists, and version DSDs and MSDs as needed</b>	6.2 Validate outputs <b>Check gathered information, review metadata reports</b>	7.2 Produce dissemination products <b>Provide flexible data and metadata services using the SDMX Registry and Web Services</b>	8.2 Conduct evaluation <b>Assess the benefits from actual SDMX implementation and plan for improvements</b>
1.3 Establish output objectives <b>Decide dataflows</b>	2.3 Design collection <b>Design the ICT architecture for collection, integration, exchange and dissemination</b>	3.3 Reuse or build dissemination components <b>Check for available tools in the SDMX.ORG website</b>	4.3 Run collection <b>Collect/Exchange the information</b>	5.3 Review and validate <b>Run VTL artefacts to help in the process</b>	6.3 Interpret and explain outputs <b>Detect discrepancies using VTL artefacts, use the SDMX datasets and metadata reports to build stories</b>	7.3 Manage release of dissemination products <b>Use embargo capabilities to automate this task</b>	8.3 Agree an action plan <b>Make the agreements to improve the actual SDMX implementation to get more benefits</b>
1.4 Identify concepts <b>Review cross-domain and global concepts, build a concept scheme</b>	2.4 Design frame and sample <b>Design constraints</b>	3.4 Configure workflows <b>Deploy the ICT infrastructure, build the mappings</b>	4.4 Finalise collection <b>Automated by SDMX, check it with VTL tools</b>	5.4 Edit and impute <b>Run VTL artefacts to help in the process</b>	6.4 Apply disclosure control <b>Check confidentiality using VTL tools, generate new dissemination DSDs/Web services</b>	7.4 Promote dissemination products <b>Promote the use of new enabled SDMX capabilities, like direct connection to data and metadata</b>	
1.5 Check data availability <b>Check data from available data and metadata flows</b>	2.5 Design processing and analysis <b>Design constraints, validation and transformation artefacts</b>	3.5 Test production systems <b>Check data and metadata flows support (ICT infrastructure and SDMX artefacts)</b>		5.5 Derive new variables and units <b>Run VTL artefacts to help in the process</b>	6.5 Finalise outputs <b>Connect SDMX datasets to dissemination systems, use of internationalization features intrinsic to SDMX</b>	7.5 Manage user support <b>Check the use of SDMX enabled services to be ensure alignment with user's needs</b>	
1.6 Prepare and submit business case <b>Design the strategy to make use of SDMX</b>	2.6 Design production systems and workflow <b>Finish the design of the collection and exchange architecture</b>	3.6 Test statistical business process <b>Run pilots to check resulting SDMX messages and reports</b>		5.6 Calculate weights <b>Run VTL artefacts to help in the process</b>			
		3.7 Finalise production systems <b>Finalize mappings and ICT infrastructure</b>		5.7 Calculate aggregates <b>Run VTL artefacts to help in the process</b>			
				5.8 Finalise data files <b>Automated by SDMX, use ICT infrastructure to move to dissemination</b>			

SDMX related recommended tasks mapped to GSBPM

Overarching Processes **Registry, DSDs, MSDs**

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs Global and international <b>DSDs, MSDs.</b> <b>SDMX, Registry</b>	2.1 Design outputs <b>DSDs, MSDs.</b> <b>Registry</b>	3.1 Reuse or build collection instruments <b>Sdmx.org, Registry</b>	4.1 Create frame and select sample <b>Registry</b>	5.1 Integrate data <b>Dataflows, Registry</b>	6.1 Prepare draft outputs <b>VTL artefacts, APIs, Registry</b>	7.1 Update output systems <b>VTL artefacts, APIs, Registry,</b> <b>Dataflows, Metadataflows,</b> <b>APIs</b>	8.1 Gather evaluation inputs <b>Technical Standards, COG,</b> <b>Statistical Guidelines</b>
1.2 Consult and confirm needs <b>Provision agreements, Registry</b>	2.2 Design variable descriptions <b>COG, Statistical</b> <b>Guidelines, DSDs, MSDs.</b> <b>SDMX, Registry</b>	3.2 Reuse or build processing and analysis components <b>Sdmx.org, Registry</b>	4.2 Set up collection <b>Mappings, Constraints</b>	5.2 Classify and code <b>Codelists, DSDs, MSDs</b>	6.2 Validate outputs <b>VTL artefacts, APIs, Registry</b>	7.2 Produce dissemination products <b>VTL artefacts, Registry,</b> <b>Constraints, Dataflows,</b> <b>Metadataflows, APIs</b>	8.2 Conduct evaluation <b>Technical Standards, COG,</b> <b>Statistical Guidelines</b>
1.3 Establish output objectives <b>DSDs, MSDs.</b> <b>SDMX, Registry</b>	2.3 Design collection <b>COG, Statistical</b> <b>Guidelines, DSDs, MSDs.</b> <b>SDMX, Registry</b>	3.3 Reuse or build dissemination components <b>Sdmx.org</b>	4.3 Run collection <b>Mappings, Constraints</b>	5.3 Review and validate <b>VTL artefacts, APIs</b>	6.3 Interpret and explain outputs <b>VTL artefacts, APIs, Registry</b>	7.3 Manage release of dissemination products <b>VTL artefacts, Registry,</b> <b>Constraints, Dataflows,</b> <b>Metadataflows, APIs</b>	8.3 Agree an action plan <b>Technical Standards, COG,</b> <b>Statistical Guidelines</b>
1.4 Identify concepts <b>Glossary,</b> <b>Registry,</b>	2.4 Design frame and sample <b>Constraints</b>	3.4 Configure workflows <b>Mappings, constraints, APIs,</b> <b>VTL artefacts</b>	4.4 Finalise collection <b>Mappings</b>	5.4 Edit and impute <b>VTL artefacts, APIs</b>	6.4 Apply disclosure control <b>VTL artefacts, APIs, Registry,</b> <b>Constraints</b>	7.4 Promote dissemination products <b>VTL artefacts, Registry,</b> <b>Constraints, Dataflows,</b> <b>Metadataflows, APIs</b>	
1.5 Check data availability <b>DSDs, MSDs.</b> <b>SDMX, Registry</b>	2.5 Design processing and analysis <b>Registry, COG, Statistical</b> <b>Guidelines</b>	3.5 Test production systems <b>Mappings, constraints, APIs</b>		5.5 Derive new variables and units <b>VTL artefacts, APIs</b>	6.5 Finalise outputs <b>VTL artefacts, APIs, Registry,</b> <b>Mappings, Dataflows</b>	7.5 Manage user support <b>Technical Standards, COG,</b> <b>Statistical Guidelines</b>	
1.6 Prepare and submit business case <b>COG, Statistical Guidelines,</b> <b>DSDs, MSDs.</b> <b>SDMX, Registry, Sdmx.org</b>	2.6 Design production systems and workflow <b>MSDs, Registry, Statistical</b> <b>Guidelines</b>	3.6 Test statistical business process <b>Mappings, constraints, APIs</b>		5.6 Calculate weights <b>VTL artefacts, APIs</b>			
		3.7 Finalise production systems <b>Mappings, constraints, APIs</b>		5.7 Calculate aggregates <b>VTL artefacts, APIs</b>			
				5.8 Finalise data files <b>Dataflows, Registry</b>			

Mapping of  
SDMX artefacts  
to SDMX

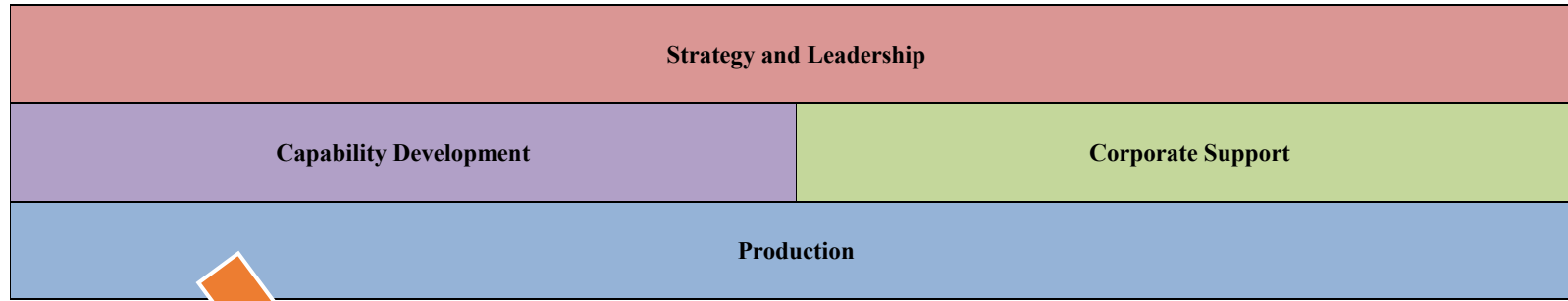
## Overarching Processes

Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and 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 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

Available SDMX tools are grouped under GSBPM activities at [the Awesome list](https://github.com/SNStatComp/awesome-official-statistics-software)

<https://github.com/SNStatComp/awesome-official-statistics-software>

# GAMSO



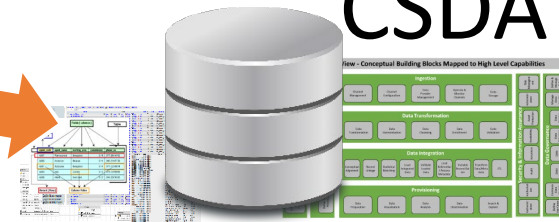
Relation With Other Standards and Models

## CSPA

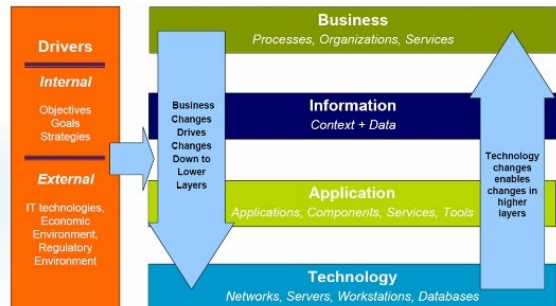


Overarching Processes							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Reuse or build collection instruments	4.1 Create frame and select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult and confirm needs	2.2 Design variable descriptions	3.2 Reuse or build processing and analysis components	4.2 Set up collection	5.2 Classify and 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 Reuse or build dissemination components	4.3 Run collection	5.3 Review and validate	6.3 Interpret and explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame and sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit and impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing and analysis	3.5 Test production systems		5.5 Derive new variables and units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare and submit business case	2.6 Design production systems and workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production systems		5.7 Calculate aggregates			
				5.8 Finalise data files			

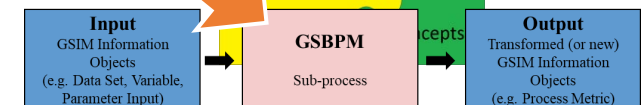
## CSDA



## TOGAF®



## GSIM



# Conclusions

- The relation SDMX-GSBPM keeps along the whole statistics production process
- There is an ecosystem of coexisting standards that have been created to improve different aspects related to the production of official statistics
- As this standards and models evolve, their tend to converge and cover existing gaps. In order to obtain the best benefits of them and to avoid overlaps and inconsistencies it is needed to visualize them as an integrated framework.

# Questions?

[Juan.Munoz@inegi.org.mx](mailto:Juan.Munoz@inegi.org.mx)

**modernstats**  
BY HLG - MOS