

# Supporting Standards Group

on behalf of our Modernisation Group:

**Flavio Rizzolo** (chair)

Workshop on the Modernisation of Official Statistics

21-22 November 2023, Geneva

# The changing world of Official Statistics

- New challenges
  - The **cross-domain** nature of our work:
    - Multitude of factors, from a variety of domains, impacting **social** and **economic** statistics;
    - Ex: **industrial production and migrations** are affected/impacted/determined by climate, wars, social change ...
  - The **sources of data** – many more, and less reliable (unknown methodology and quality).
  - The **types of data** – larger, more complex, heterogenous.
- New requirements
  - More **complete**, **machine-actionable**, and **interoperable** (smart?) metadata;
  - Improved “**context**” for data (i.e., provenance, semantics);
  - Ability to **describe**, and **integrate**, data in new **data formats/structures**, e.g. streams, key-value;
  - Broader range of **technologies** (cloud, sensors, AI, etc.).
- **Standardization** (methods, processes, data, metadata, tools) enable statistical agencies to deal with new challenges and requirements more **effectively** and **efficiently**.

**ModernStat models** enable the implementation of **interoperable data and metadata systems** based on **open standards**, e.g. DDI, SDMX, VTL, S/XKOS, etc.

# The Supporting Standards Group can help



## Supporting standards

Modernisation of Official Statistics

Find ways to

*develop,*

*enhance,*

*integrate,*

*promote,*

*support, and*

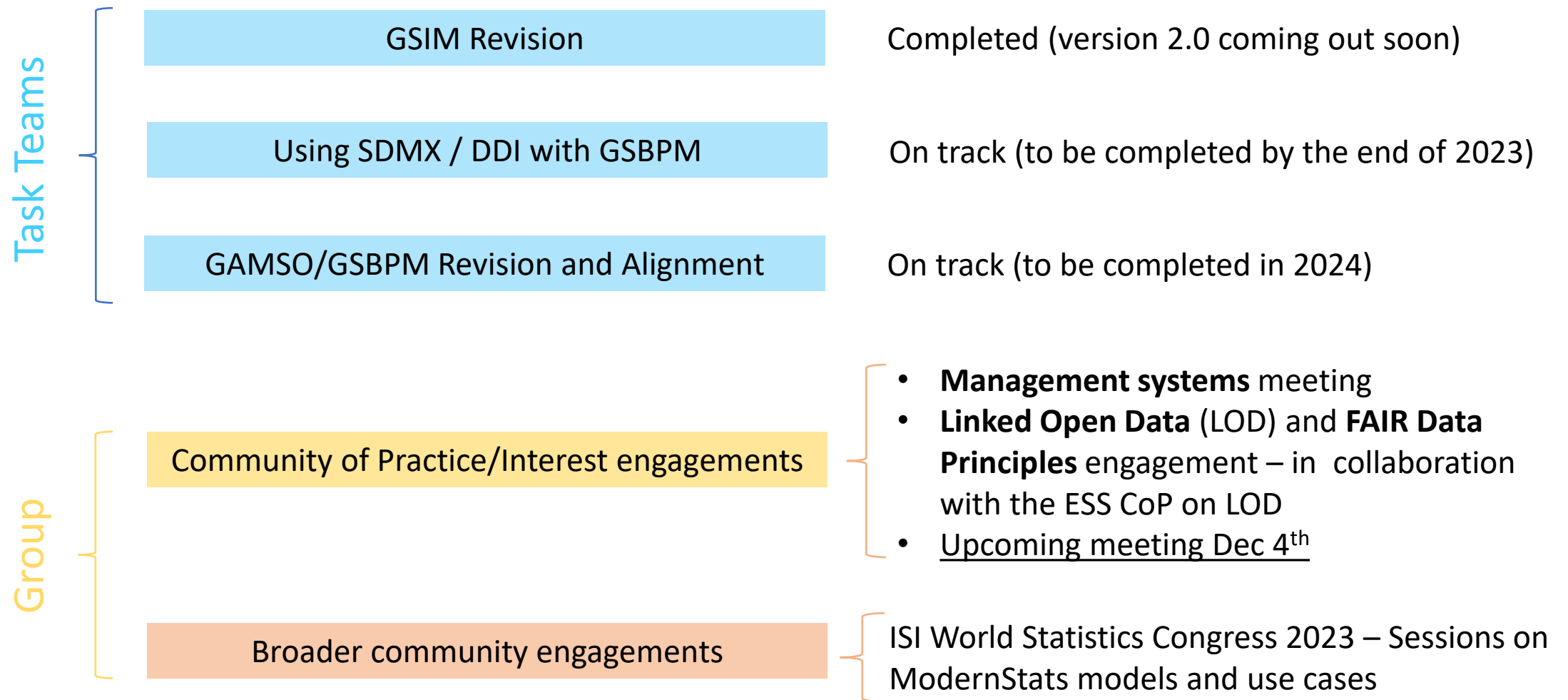
*facilitate the implementation of*

a range of standards needed for statistical modernisation

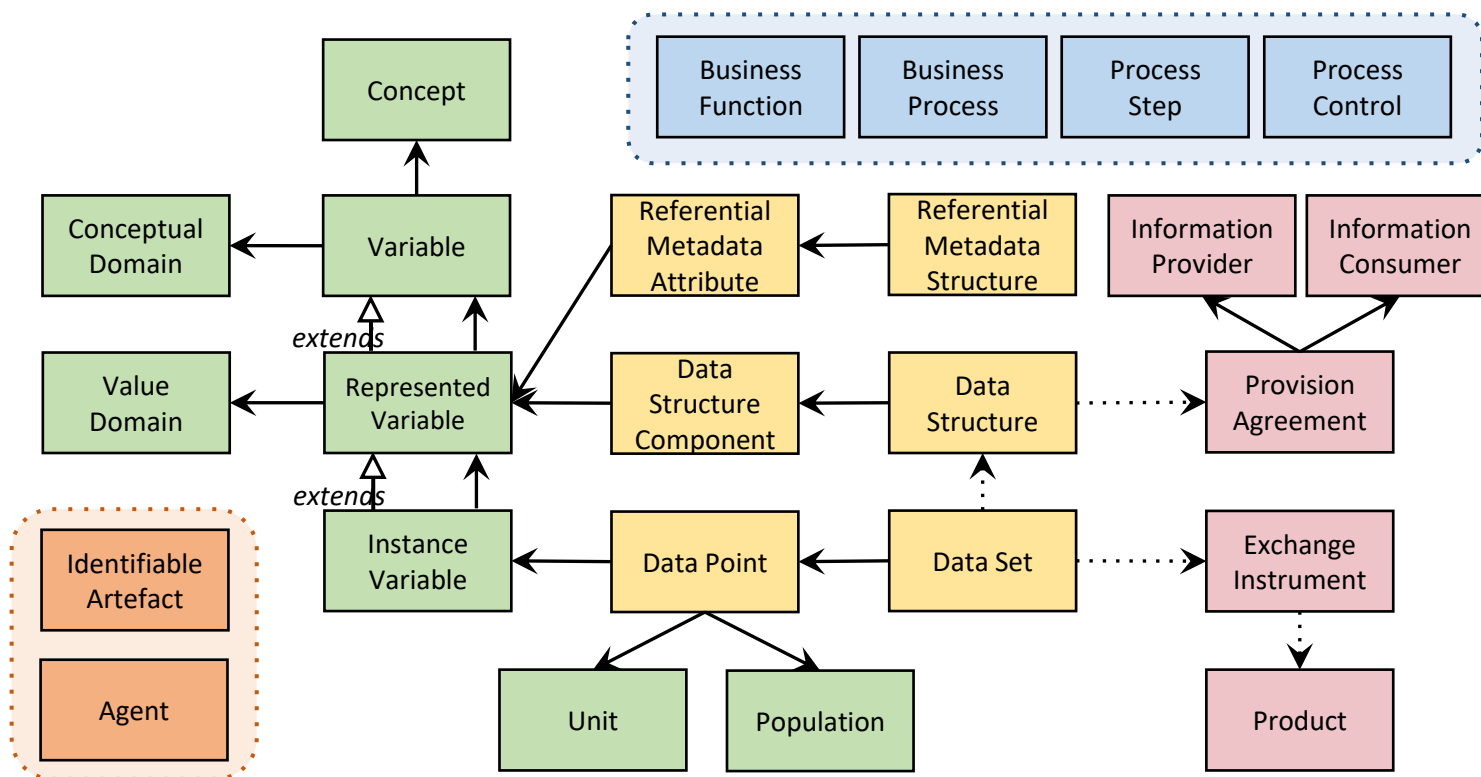
- **Operational responsibility** for the maintenance and development of the ModernStats standards.
- **Network of experts** with a extensive experience in key areas (standards, methodology, data/metadata management, architecture, software development, project management, etc.)



# Overview of 2023 activities



# GSIM Revision → v2.0



Group	Changes
CONCEPTS	<u>Overall</u> (Minor) changes
STRUCTURES	<u>Referential Metadata</u> Improved clarity and usability (Use cases: ESS Standard for Quality Reports Structure and IM Sweden)
BUSINESS	<u>Process and Steps</u> Better separation between design and execution More clarity on types of inputs/outputs
EXCHANGE	<u>Exchange Channel</u> Better separation between specification, implementation and information exchanged
BASE	<u>Change Event</u> Improvement of how information about changes is maintained

## Feedback from:

- **Metadata Glossary, Linking GSBPM and GSIM, GeoGSBPM and COOS** activities
- **Statistical agencies** during GSIM specification review period

More information: <https://github.com/UNECE/GSIMRevision/>

# Using SDMX / DDI with GSBPM



## Benefits

- International implementation standards with large adoption base
- Free, open and collaborative development (for the most part)
- Growing ecosystem of tools readily available
- Better interoperability, reusability and quality



## Practical considerations

- Which standard to use for which use case?
- Which artefact to use for which stage of production process?
- How to create pipelines/workflows that use different standards at different steps?
- How to help people with the more technical aspects?

# Using SDMX / DDI by sub-process

## Objectives

- Provide an introduction to each standard in the context of GSBPM
- List relevant SDMX/DDI artefacts under each GSBPM sub-process
- Use the “Linking GSBPM and GSIM report” to inform and guide the exercise

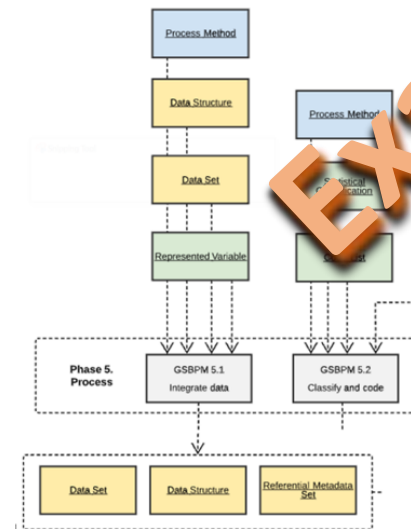
## Results and findings

- Identified where SDMX and DDI work better together (their strengths and “quirks”) using GSBPM sub-processes as context
- Improved interoperability between the two standards
- Good basis for further work on guidelines for implementing GSBPM (and GSIM) with DDI, SDMX, VTL, SDTL and other open standards

## Sub-process 5.1 Integrate data

### SDMX

- SDMX provides uniform data formats, cross-domain structural metadata and **content-oriented guidelines** regardless of the data source (including geospatial), ensuring enabling data harmonization “by design”.
- IT infrastructure based on **Web Services**, the **SDMX Registry** and the **mapping mechanisms** used in SDMX can be very useful to support the integration.
- SDMX Dataflow is helpful in this sub-process as a way to integrate data from different sources
- Other relevant SDMX artefacts include **DSD**, **Code Lists**, **Glossary**, **Code lists**.



For data integration, first data needs to be organized using DDI in some way, usually in **Data Sets**, and described in detail, usually with **Data Structures**.

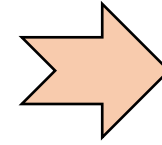
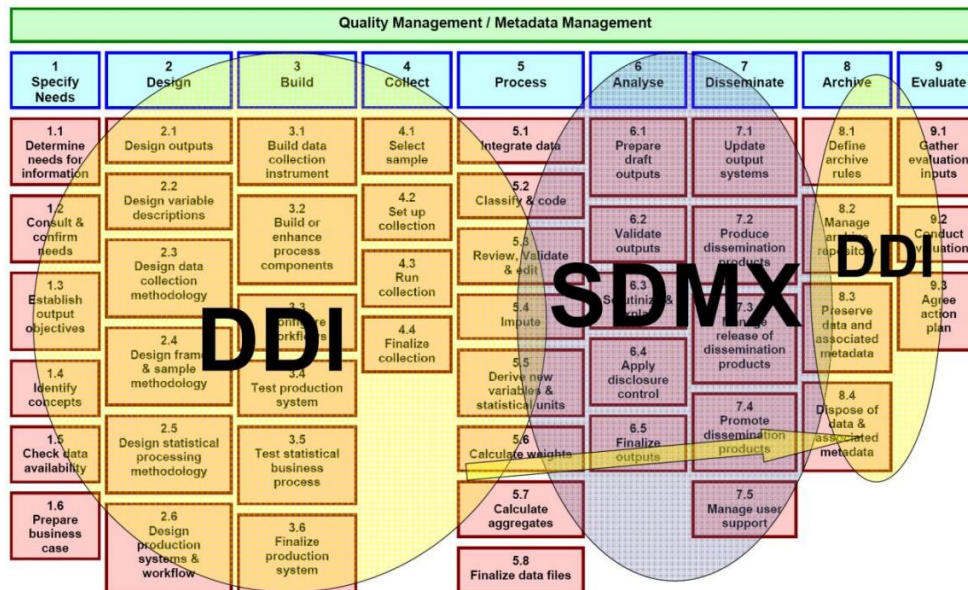
Second, data needs to be described consistently at three levels of detail, captured in DDI by the variable cascade: **Conceptual Variable**, **Represented Variable** and **Instance Variable** (or simply **Variable** in DDI-LC).

Third, mappings may have to be established between the variables used to integrate the **Data Sets/Structures**, often via an **Instance Variable Map** (or **Comparison** in DDI-LC).



# Relevance of SDMX / DDI by sub-process

Circa 2010



Identify Needs Phase	Design Phase	build phase	collect phase	process phase	analyse phase	disseminate pha	evaluationphase
phase		anything in build can be reused in build phase. DDI as an input					
1,1 identify needs	design output	reuse or build collection instrun	create frame and select samp	integrate data classify	prepare draft outputs	update output systems	gather evaluation input
1,2 consult and confir	design variable descriptions	reuse or build pr	reuse or build pr	reuse or build pr	validate outputs	produce dissemi	conduct evaluation
1,3 establish output o	design collection	reuse or build dissemination crun	reuse or build dissemination crun	reuse or build dissemination crun	interpret and explain outputs	manage release dissemination p	agree action plan
1,4 identify concepts	design frame and sample	configure workfl	finalise collection	edit and impute	apply disclosure	promote dissemination	production
1,5 check data availa	design processin and analysis	test production system		derive new variables	finalise outputs	manage user support	
1,6 prepare and subn	design productio systems and wo	test statistical business process		calculate weight			
		finalise production system		calculate aggregates			
				finalise data files			

New (2023)

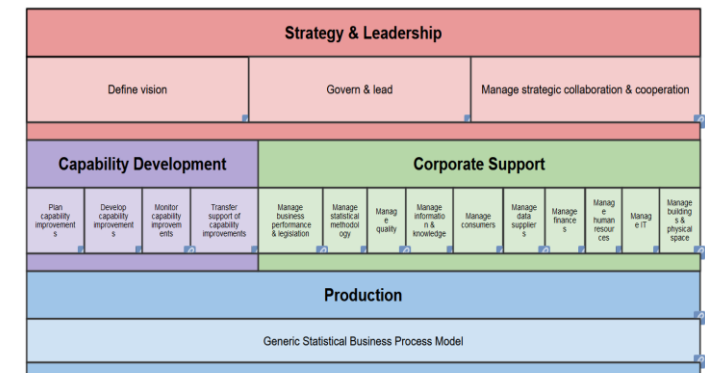
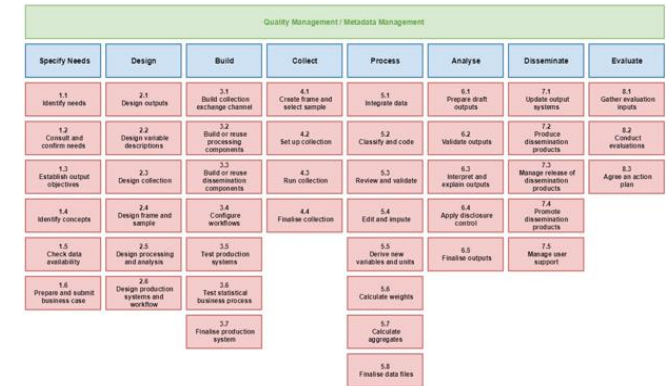
- The standard is relevant
- The standard is marginally relevant
- The standard is not quite relevant

Identify Needs Phase	Design Phase	build phase	collect phase	process phase	analyse phase	disseminate pha	evaluationphase
phase		anything in build can be reused in build phase. DDI as an input					
1,1 identify needs	design output	reuse or build collection instrun	create frame and select samp	integrate data classify	prepare draft outputs	update output systems	gather evaluation input
1,2 consult and confir	design variable descriptions	reuse or build pr	reuse or build pr	reuse or build pr	validate outputs	produce dissemi	conduct evaluation
1,3 establish output o	design collection	reuse or build dissemination crun	reuse or build dissemination crun	reuse or build dissemination crun	interpret and explain outputs	manage release dissemination p	agree action plan
1,4 identify concepts	design frame and sample	configure workfl	finalise collection	edit and impute	apply disclosure	promote dissemination	production
1,5 check data availa	design processin and analysis	test production system		derive new variables	finalise outputs	manage user support	
1,6 prepare and subn	design productio systems and wo	test statistical business process		calculate weight			
		finalise production system		calculate aggregates			
				finalise data files			

Source  
Steven Vale: Exploring the relationship between DDI, SDMX and GSBPM

# GAMSO-GSBPM Revision and Alignment

- GAMSO and GSBPM can be realized in practice:
  - By using implementation standards, e.g. SDMX, DDI, BPMN, BPEL, ProvOne, etc.
  - By using existing standards-based tools and services, e.g. .Stat Suite, SDMX-RI, SDMX Converter, Colectica, Aria, RDS, etc.
- Over the past few years feedback has been compiled on two ways:
  - Directly from **Statistical agencies** during the **GSBPM review** period
  - Indirectly from the **GSBPM, Core Ontology for Official Statistics (COOS), Linking GSBPM and GSIM** and **Ethics** Task Teams



## Recent updates (2023)

- Greater visibility of overarching processes
- Renamed “Collect” phase to “Acquire” and review sub-processes to include all types of data intakes

## Upcoming work (2024)

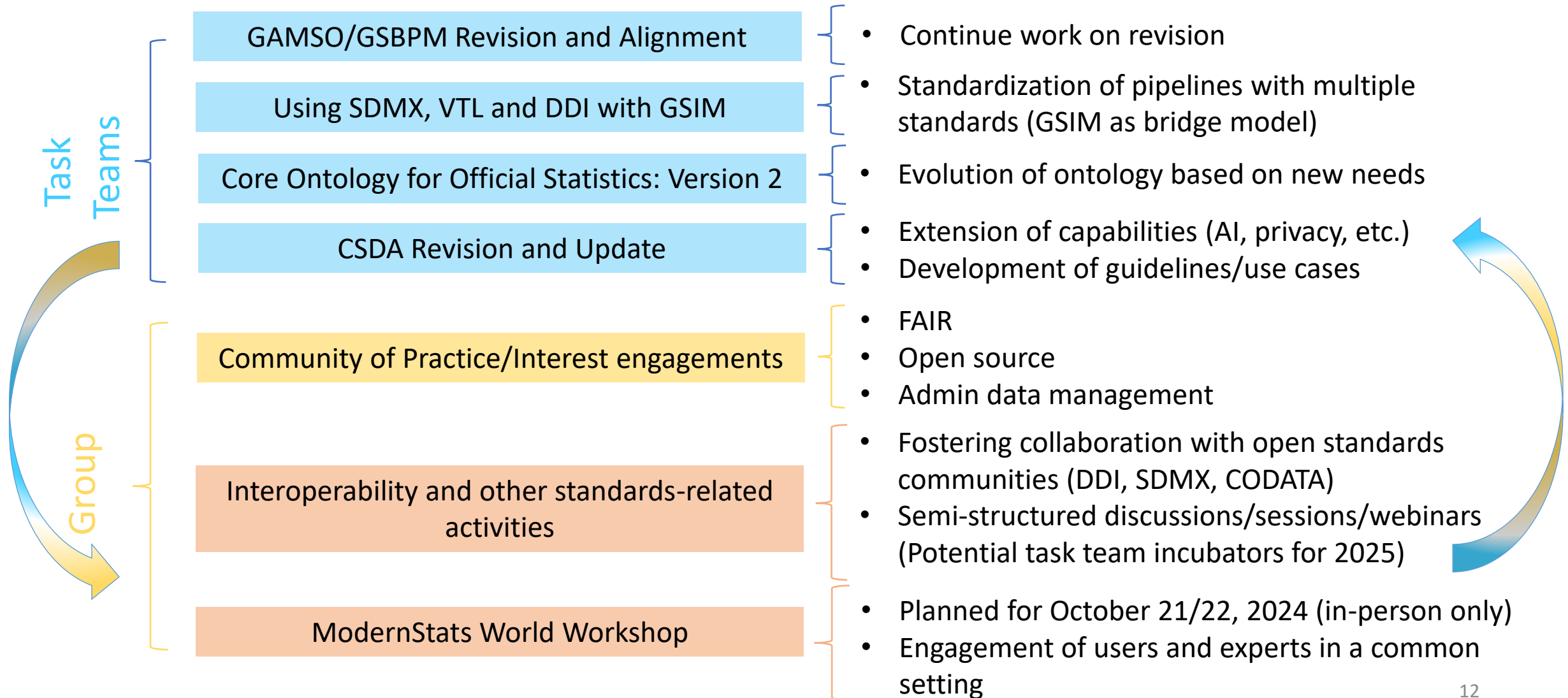
- Changes to the numbering/identification of GSBPM sub-processes to emphasize the non-linearity of the model
- Integration of ethics aspects in GAMSO

More information: [https://github.com/UNECE/GSBPM\\_GAMSO\\_Revision](https://github.com/UNECE/GSBPM_GAMSO_Revision)

# Future directions

- Evolution of ModernStats standards → be relevant and stay relevant
  - CSDA revision and update
- Integration of ModernStats standards → work better together
  - Between official statistics and the outside world (FAIR, CODATA, RDA)
  - How to use SDMX, VTL and DDI with GSIM
  - Core Ontology for Official Statistics (COOS) extension
- ModernStats Community of Practice/Interest → solve problems together
  - Bring together experts and users
  - Discuss topics that the SSG should explore and further develop
  - Discuss topics of interest parallel to the revision of the models
  - Increase the understanding and the use of ModernStats models within different user groups.

# Overview of 2024 agenda

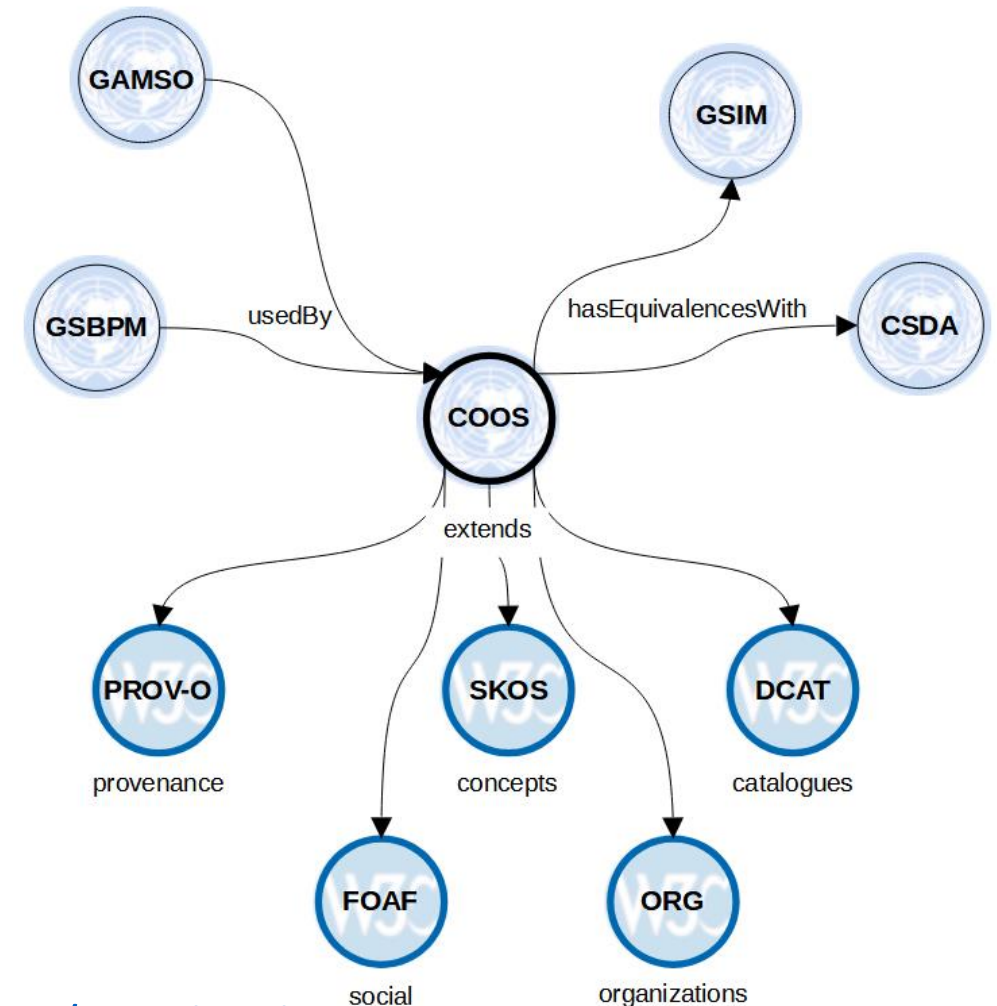


# Core Ontology for Official Statistics (COOS)

- COOS as an integration model for the core set of ModernStats standards backed by elements of well-known standard vocabularies.
- COOS defines a conceptual integration framework to provide semantic coherence across these models based on a common vocabulary of terms, definitions and a well-defined set of inter- and intra-model relationships formalized in RDF/OWL, using standards vocabularies, e.g. SKOS, PROV, DCAT, DC, ORG, etc.

## Proposed COOS v 2.0 activity (2024):

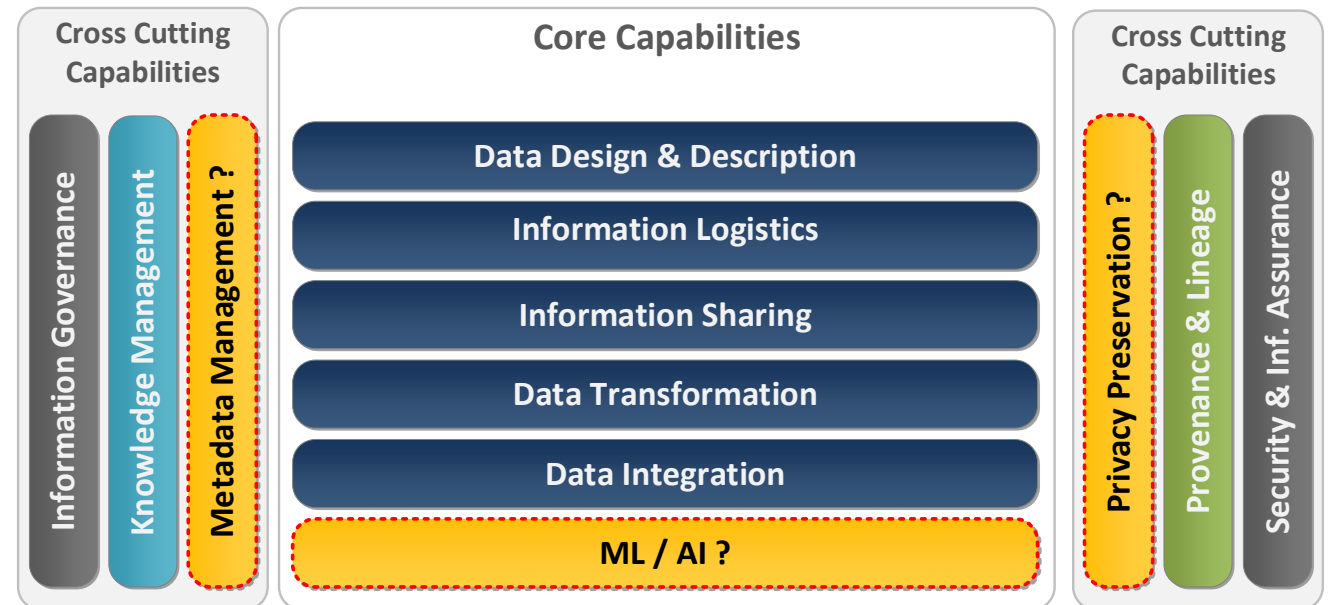
- Integration of feedback from review
- Alignment with evolving ModernStats models
- Review v 1.1 RDF representation
- Review usage of other standards vocabularies
- Review internal consistency



More information: <https://linked-statistics.github.io/COOS/coos.html>

# Common Statistical Data Architecture (CSDA)

- The **Common Statistical Data Architecture (CSDA)** includes a catalog of information capabilities.
- **Information Capabilities:** standardized, conceptual building blocks that systematically describe what an agency **does** with both **data** and **metadata**.
- **CSDA** complement GSIM, GSBPM and GAMS0
  - Integrates standard vocabularies, processes, software and people into a cohesive **data-centric framework**.
  - Expands coverage of what we do into new areas (e.g. knowledge management, provenance, security, search and discovery, persistence)
- Development frameworks and IT systems can be used to make CSDA capabilities **machine-actionable**.



## Proposed CSDA Revision and Update Activity (2024):

- Extension to new areas of interest that arose over the past five years (e.g., **generative AI, FAIR**) and areas that were not included in the 2018 specification (e.g., machine learning, privacy preservation)
- Development of **guidelines** and **use cases** to accelerate adoption

# Current members

Expert	Organization (Country)	Expert	Organization (Country)	Expert	Organization (Country)
Anna Dlugosz	GUS (Poland)	Edgardo Greising	ILO	José Luján	INEGI (Mexico)
Carlo Vaccari	Independent expert	Flavio Rizzolo	StatCan (Canada)	Juan Muñoz	INEGI (Mexico)
Christopher Jones	UNECE	Florian Vucko	Insee (France)	Juan Rioja	INEGI (Mexico)
Cory Chobanik	StatCan (Canada)	Franck Cotton	Insee (France)	Manuel Cuellar	INEGI (Mexico)
Dan Gillman	BLS (USA)	Giorgia Simeoni	Istat (Italy)	Martina Hahn	Eurostat
Daniel Blanc	INE (Uruguay)	Helda Mitre	Instat (Albania)	Matjaz Jug	CBS (Netherlands)
David Barraclough	OECD	InKyung Choi	UNECE	Mauro Bruno	Istat (Italy)
Denis Grofils	Pacific Community	Janusz Dygaszewicz	GUS (Poland)	Olivier Sirello	BIS

Many thanks to all the experts, both current members and reviewers, committed to the work of the **Supporting Standards Group** and its **Task Teams!!**

# We need you!

- Please consider joining the **Supporting Standards Group** and one of our **Task Teams** to help us realize our ambitious work programme!
- We need all types of expertise, especially around **open standards, methodology, data/metadata management** and **architecture**.
- Excellent opportunity to take part in the modernisation programme!
- Information on the Supporting Standards Group on the [HLG-MOS Modernisation Group page](#)
- Contact us!
  - Christopher Jones: [jonesc@un.org](mailto:jonesc@un.org)
  - Flavio Rizzolo: [flavio.rizzolo@statcan.gc.ca](mailto:flavio.rizzolo@statcan.gc.ca)



# Thank you!

<https://statswiki.unece.org/display/hlgbas/Modernisation+Groups>