Supporting Standards Group

Zoltán Vereczkei (Chair)

16 November 2021

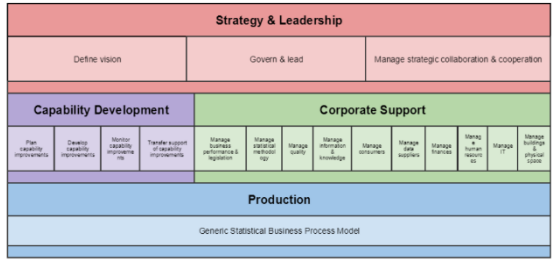
HLG-MOS Modernisation Workshop 2021 (Virtual)

Supporting Standards Group

- The goal of the group is to find ways how to develop, enhance, integrate, promote, support and facilitate implementation of the range of standards needed for statistical modernisation.
- Operational responsibility for the maintenance and development of the ModernStats models:

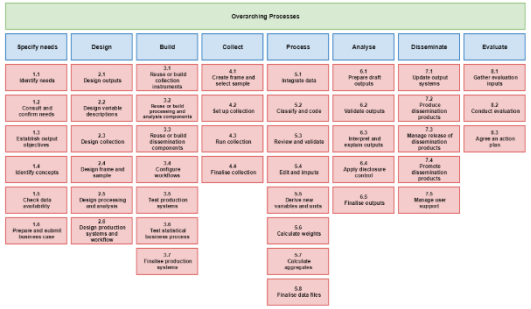
GAMSO

Generic Activity Model for Statistical Organisation



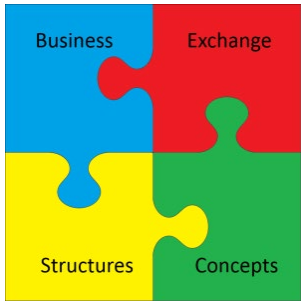
GSBPM

Generic Statistical Business Process Model



GSIM

Generic Statistical Information Model



CSPA

Common Statistical Production Architecture

Members

- 20 experts from 12 NSOs and 4 international statistical organizations

Members	Organizations
Zoltán Vereczkei - Chair	Hungary
Anna Dlugosz	Poland
Carlo Vaccari	Italy
Cory Chobanik	Canada
Csaba Ábry	Hungary
Dan Gillman	United States
David Barraclough	OECD
Edgardo Greising	ILO
Emanuele Baldacci	Eurostat
Essi Kaukonen	Finland

Members	Organizations
Flavio Rizzolo	Canada / DDI
Florian Vucko	France
Franck Cotton	France
Juan Muñoz	Mexico
Kevin McCormack	Ireland
Martina Hahn	Eurostat
Matjaz Jug	Netherlands
Mauro Bruno	Italy
Omurbek Ibraev	Kyrgyzstan
Waleed Mohamed	Egypt

Network of experts

5 task teams in 2021

1. Linking GSBPM and GSIM task team – 18 members
2. Geospatial task team – 20 members
3. GSIM task team – 21 members
4. Core Ontology for Official Statistics task team – 17 members
5. GSBPM “task” task team – 9 members

57 unique members from (17 national and 3 international organizations)
35 virtual meetings in 2021

Overview of 2021 activities

Task teams

Geospatial	**Completed, available here**
Linking GSBPM-GSIM	**To be completed by end of 2021**
Core Ontology (COOS)	**To be completed by end of 2021**
GSIM	**Started as „soft update”…**
GSBPM “task”	**On-going (started in Fall 2021)**

Group

ModernStats Governance	**Completed, available here**
ModernStats Usage Survey	**Completed, available here**

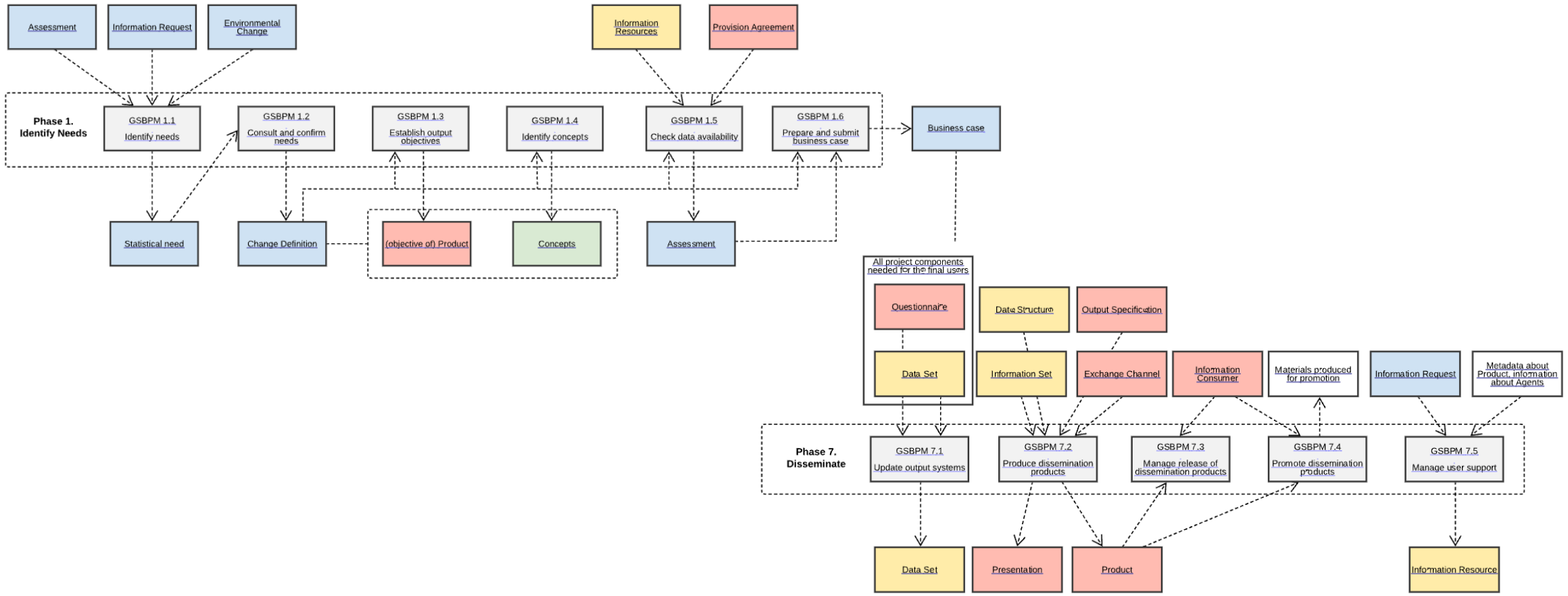
Linking GSBPM and GSIM task team

Sub-process 5.3 Review and validate

Process Input Specification	***Process Design***	***Process Output Specification***
***Core Input* type** *Data Sets* (*Unit Data Sets*) to be reviewed and validated *Data Structures* associated with *Data Sets* to understand *Data Sets* *Represented Variables* to be reviewed and validated *Process Methods* that specifies methodology for review and validation (e.g. calculating plausibility or validity) which can be represented as *Rules,* as designed in Phase 2 ***Parameter Input* type** Parameter values to be used for review and validation methodologies as specified in *Process Method* such as: • Limit value for edit *Rule* (interval for valid values) • Threshold for checking outlier ***Process Support Input* type** Auxiliary *Data Sets* or any *Information Resource* to be used for review and validation, e.g. historic comparison, macro-level comparison Technical / methodological handbooks, policies or guidelines to be followed regarding data validation as well as quality management	***Process Method*** Review *Data Sets* and *Process Methods* Apply *Process Methods* and *Rules* to review *Data Sets* Apply *Process Methods* and *Rules* to validate *Data Sets* Calculate quality measures specified by *Process Methods* Update *Data Sets* and associated element in *Data Structure* with results from review and validation	***Core Output* type** *Data Sets (Unit Data Sets):* updated *Data Set* *Data Structure* associated with *Data Set* Referential metadata: descriptions of the *Process Methods* used, quality information summarising *Process Metrics* or any other relevant information to be passed along with *Data Sets* ***Process Metric* type** Quality measures related to review and validation such as: • Number of validations conducted • Number of outliers detected Quality measures of *Process Step* such as: • Time spent to complete the *Process Step* (derived from *Process Execution Log)* • Cost spent to complete the *Process Step* ***Process Execution Log* type** Execution log such as • Time that *Process Step* started • Time that *Process Step* ended • Any message or event log generated from software used for review and validation
Process Control Design Have all *Represented Variables* needs to be reviewed? • If YES, go to next *Process Step* in *Business Process* • If NO, go to the beginning of current *Process Step*		

Linking GSBPM and GSIM task team

(Examples of “clickable” version of information flow: phase 1 and phase 7)

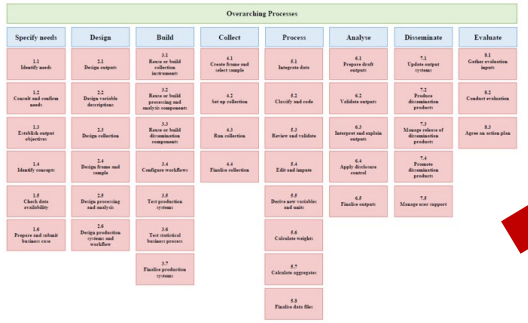


Linking GSBPM and GSIM task team

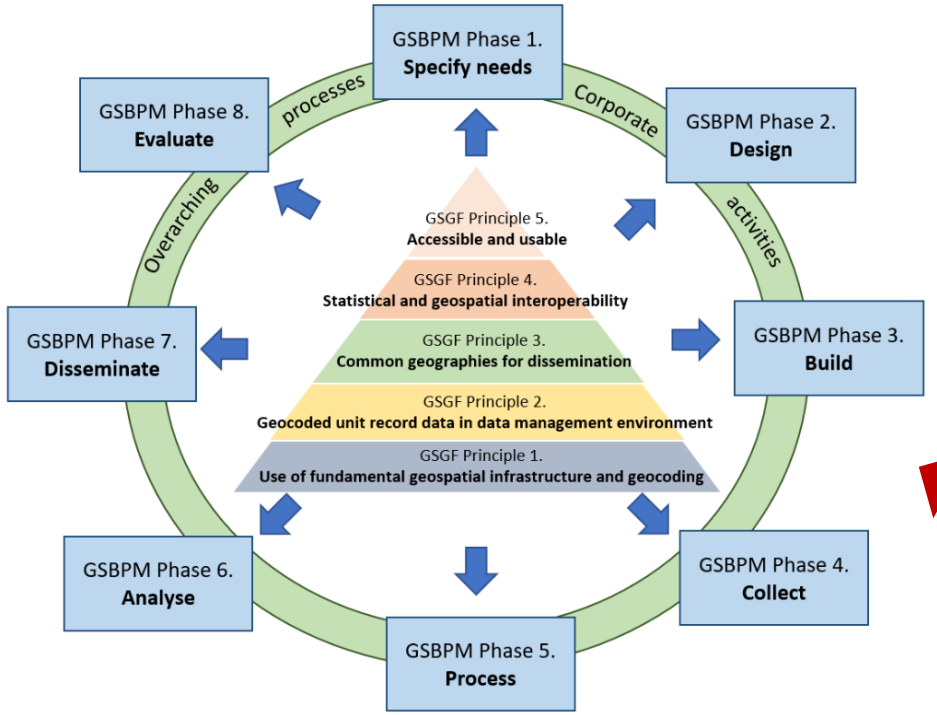
- Integrated view for GSBPM and GSIM.
- Help us (you) to better understand the relationship between the GSBPM and GSIM.
- Valuable to help you to move forward with implementation of ModernStats models (especially GSIM after GSBPM or the two models together).
- Easier interpretation of GSIM objects and definitions
- Available for GSBPM sub-processes.

Geospatial task team: GeoGSBPM

- Global Statistical Geospatial Framework (GSGF) provides principles that supports the production of harmonized and standardized geospatially enabled statistical data.
- GeoGSBPM operationalizes the GSGF principles throughout the production process.



GSBPM



GSGF

Geospatial task team: GeoGSBPM

- GeoGSBPM describes geospatial-related activities and considerations using the framework of the GSBPM

(Example of GSBPM sub-process 2.2 Design variable description)

2.2 Design variable description

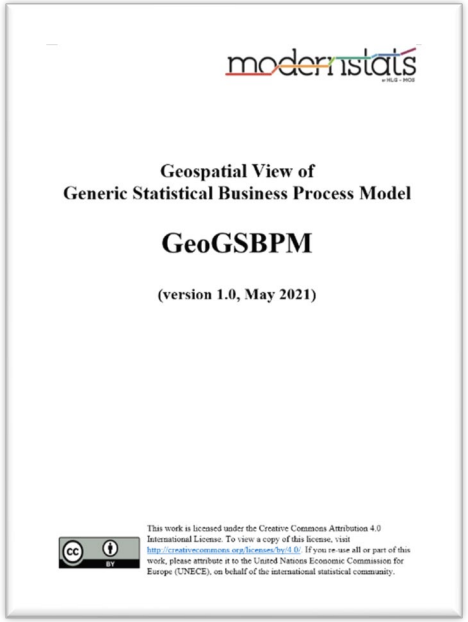
28. This sub-process defines the variables to be collected via the collection instrument, as well as any other variables that will be derived from them in sub-process 5.5 (Derive new variables and units), and any statistical or geospatial-classifications that will be used. It is expected that existing national and international standards will be followed wherever possible.

29. Geospatial variables (geographies) that are used while collecting data at a statistical unit level are not usually the same as those that are used for dissemination. Hence, they should be designed at the statistical unit level using point-based location[8] as the base geospatial variable, as it will provide a considerable adaptability to changes over time and flexibility to aggregate up to various dissemination-level geographies. For gridded geographies, it is important to use a grid system that is comparable with the existing regional or global grid system (e.g. Discrete Global Grid System (DGGS)[9]) as it will greatly increase usability of the output. Different types of grid (e.g. hexagon, rectangular) and their advantages and disadvantages can be assessed when designing gridded geographies.

30. This sub-process may need to run in parallel with sub-process 2.3 (Design collection), as the definition of the variables to be collected, and the choice of collection instruments may be inter-dependent to some degree. Preparation of metadata descriptions of collected and derived variables, statistical and geospatial classification is a necessary precondition for subsequent phases.

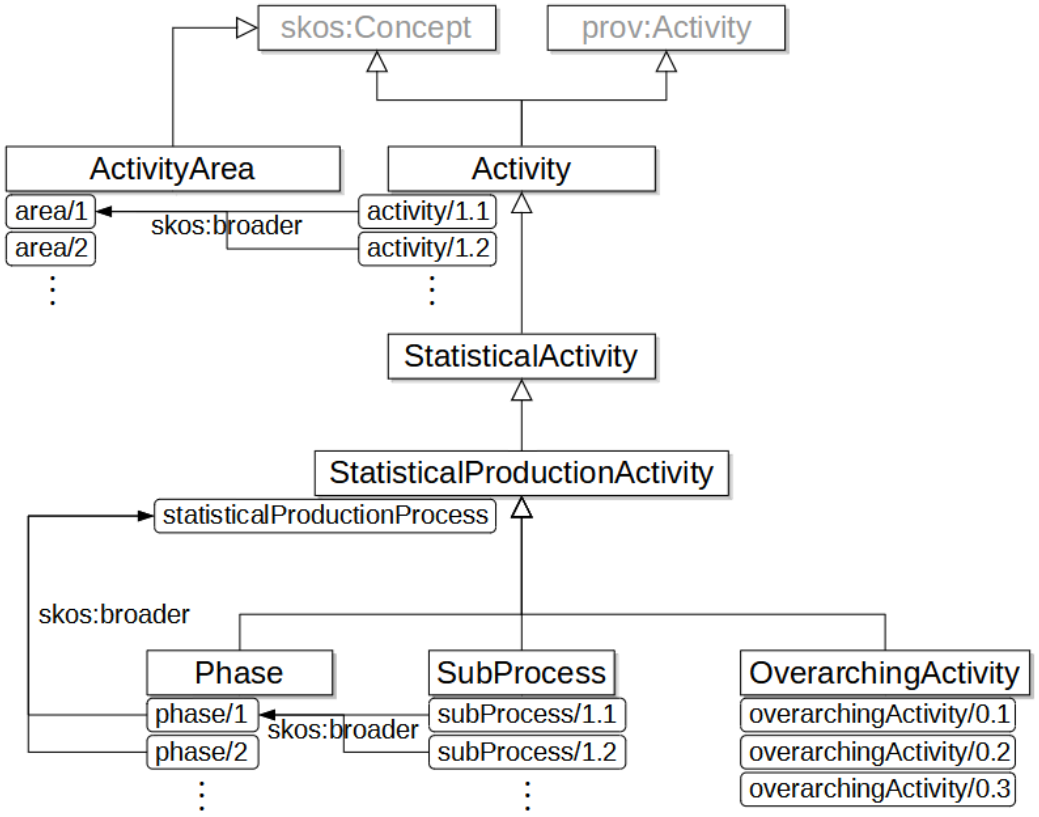
GSBPM original text

Geospatial-related consideration and activities

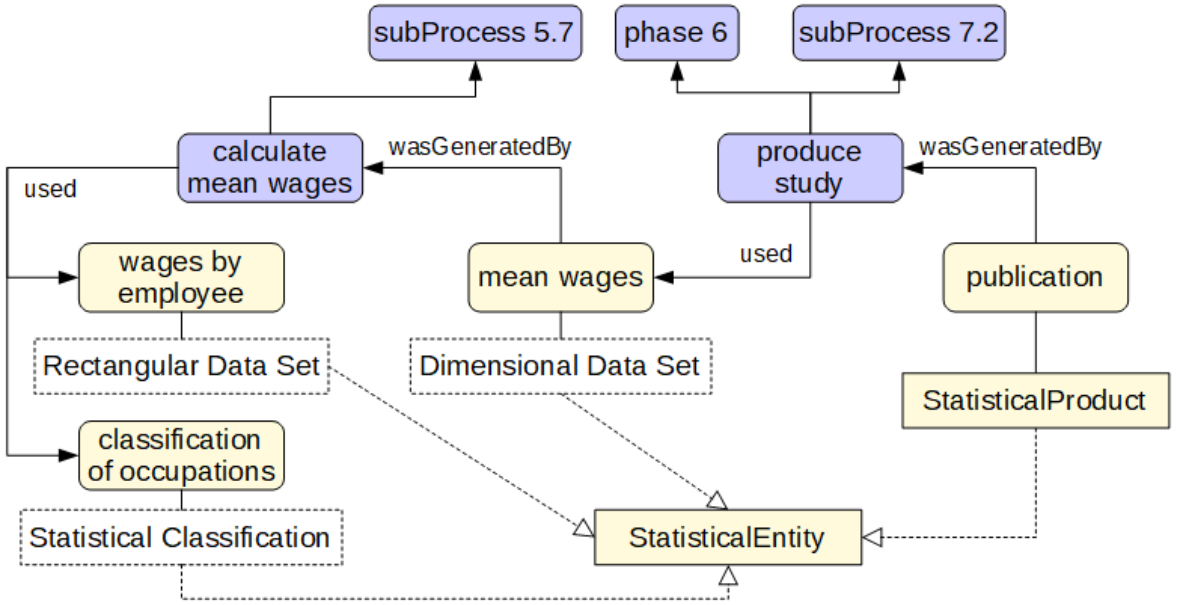


Available on [GeoGSBPM wiki]

Core Ontology task team



COOS integrating ModernStats models (GSBPM, GAMSO)



Example: COOS and external vocabularies it links to can allow to represent and share concrete statistical processes

Core Ontology task team

- Trends:
  - Collaboration in order to define business models aiming at building common representations of processes and information in the domain of official statistics: ModernStats models: GSBPM, GSIM, GAMSO, CSDA.
  - More and more statistical offices turn to semantic standards in order to formalize their data and metadata.
  - Core ontology suggests formal representations for the core concepts used in our domain.
- What is available? Ontology specification, a governance document, a URI policy and OWL ontology (under expert review).
- Details on Github: https://linked-statistics.github.io/COOS/coos.html

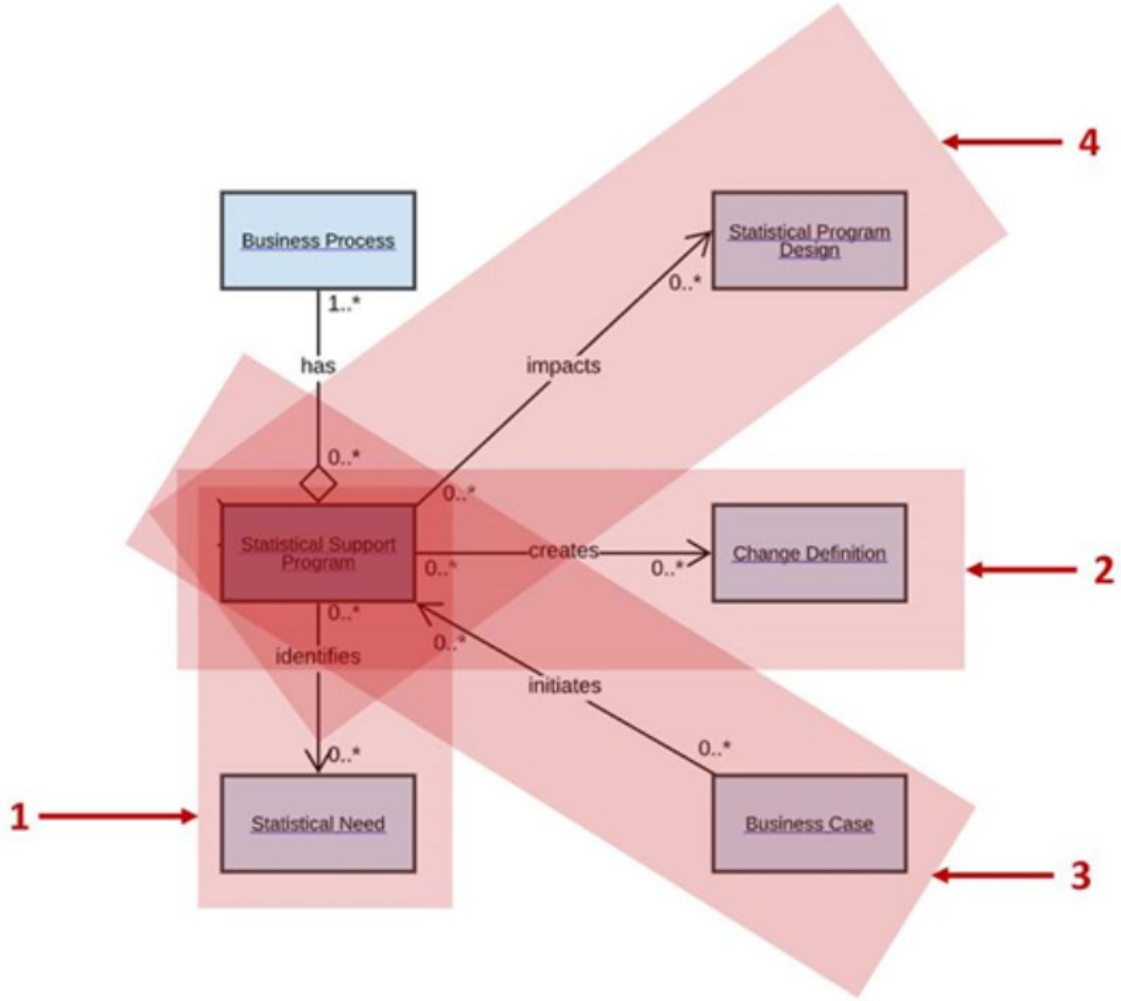
GSIM task team

- Produced GSIM Communication Paper in 2020.
- We started as a „soft update” but ended up to overcame challenges with a lot of added value for the future GSIM version. These are fundamental for the model (re-assessment of several key GSIM objects and modify them) → full revision.
- 2021: lot of focus on interoperability: Linking GSBPM and GSIM and Core Ontology task teams provided valuable inputs and renewed perspectives on how GSIM could be used in conjunction with GSBPM and GAMSO.

GSIM task team

- Two GSIM objects „Statistical Program” and „Statistical Support Program” were heavily discussed, with various scenarios (e.g., GSBPM Phase 1, Phase 8, GAMSO corporate support) and organizational set-up.

- As the level of process standardization increases in the statistical organization, the role of these two objects, and their relationships to GSBPM, become more important.



GSBPM „Tasks” task team

- Many countries try to indentify lower level to the GSBPM model and requested compiling a common list for the community.
- Country examples available – thank you!
- The work started during the fall of 2021 and will continue but already started compiling first proposals (based on common principles for phases and sub-processes).
- Mixture of country examples, GSBPM expertise and GSBPM model issues (potential lessons for a future GSBPM revision).
- Expect to conclude around June 2022.

GSBPM „Tasks" task team

(Example of GSBPM „task" level activities from countries)

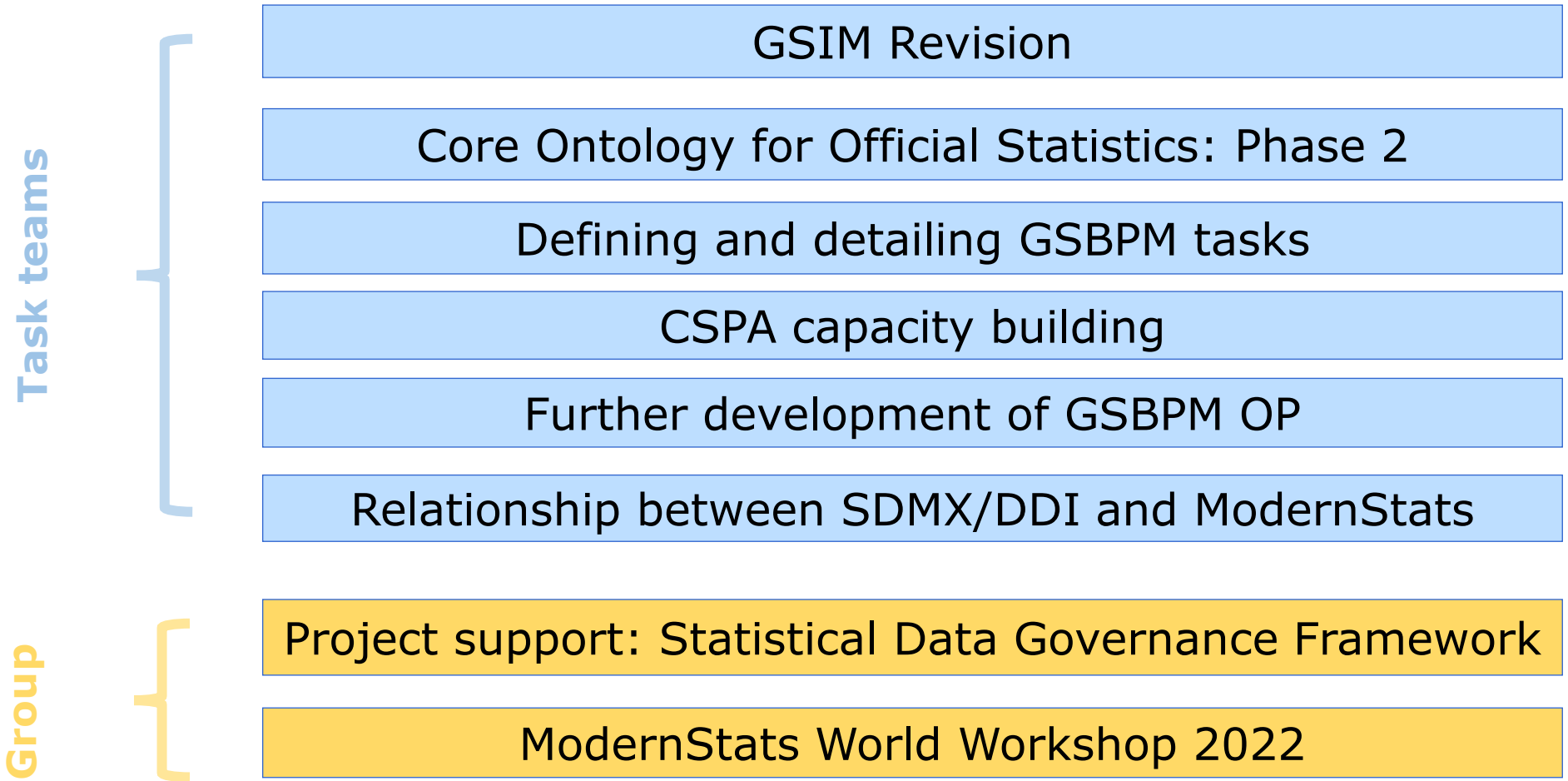
Sub-process	Tasks								Linking GSBPM and GSIM (Process Design)
	Australia	Costa Rica	Norway	Romania	Serbia	Spain	France	Azerbaijan	
1.1 Identify needs	1.1.1 Assess the problem	•identify the main user persons and entities; •identify the information needs within the entity; •identify the external information needs of the entity; •verify whether existing information needs are already investigated in otherEOs; •compare similar EOs; •identifylegal, financial, human and technological orother resourceconstraints, and •other activities.	1.1.1 Identify need for information 1.1.2 Develop concepts	D.01.0.01 Identifying user groups depending on the needs by key groups or large user profiles (internal or external) D.01.0.02 Identifying the indicators required by the European Regulations in the statistical field D.01.0.03 Identification and monitoring of new needs (new data requests) D.01.0.04 Analysis of specific national and international practices D.01.0.05 Drawing up of the list with the requested indicators, specifying which can be provided and which cannot D.01.0.06 Periodic analysis (biennial, five-year, etc., as the case may be) of the unmet needs of the users in order to analyse the possibility of improving the statistics produced (including new indicators, new data sources, new levels of disaggregation of existing indicators, etc.).	1.1.1 Identify needs for statistical outputs 1.1.2 Authorize requests for detailed establishment of needs	1.1.1 Identify external needs for information 1.1.2 Identify internal needs for information 1.1.3 Compare similar statistical operations 1.1.4 Identify restrictions 1.1.99 Other tasks	- Make an inventory of (new) needs (user's survey, quality survey, international regulation...) - Benchmark other NSI's practicings and methods	1.1.1 Identifying needs based on national priorities 1.1.2 Study International methodologies, standards and practices and identifying needs 1.1.3 Study normative base and internal needs 1.1.4 Research, analyse users' needs and organise meetings with users 1.1.5 Prepare identified needs for discussions	- Check all *Information Requests*, *Environment Changes* and *Assessments* received - Conduct an initial analysis of *Information Requests*, *Environment Changes* and *Assessments*, e.g. by comparing with practice amongst other statistical *Organisations* - Produce the *Statistical Needs*
1.2 Consult and confirm needs	1.1.2 Consult clients to understand the nature of the problem 1.1.3 Document stakeholder needs 1.1.4 Research & validate needs 1.1.5 Align to ABS capabilities & priorities 1.1.6 Confirm stakeholder needs	•define the strategy for consultation with user persons and entities; •contact and consult the persons and entities users to detail and delimit the information needs, that is, to establish thescope; •make a diagnosis on the importance of producing the new O or changes in the existing ones where it is described: involved, problems, deficiencies, background and justification or importance (importance for the entity, society and country, including stakeholders) and •other activities.	1.2.1 Identify key users 1.2.2 Specify need in consultation with key users 1.2.3 Confirm need with key users	D.02.0.01 Organizing meetings with users who have data needs, satisfying and confirming their needs D.02.0.02 Prioritizing needs according to the possibility of production, the response burden, costs and confidentiality criteria D.02.0.03 Drawing up protocols/conventions for institutional collaboration D.02.0.04 Analysis of the legal framework that has implications on statistical observation and measurement	1.2.1 Consulting with users and establishing in detail their needs	1.2.1 Contact users 1.2.99 Other tasks	- Consult stakeholders to understand (in detail) their needs for the statistics - Identify changes in the needs	1.2.1 Consultation of identified needs with stakeholders 1.2.2 Specification of needs 1.2.3 Discuss and confirm identified needs	- Consult the *Statistical Needs* with internal and external stakeholders to confirm in detail the needs for the statistics - Produce the *Change Definitions* specifying the needs in detail

Different level of granularity, local settings, etc.

Future directions

- Ensuring integrated view of our models in practice:
  - Revision of existing models that take input directly from lessons learned from the community, the use of our models and the task teams (example: GSIM).
  - Investing resources into developing common foundations that help us to understand how the models work together and provide further practical guidance (example: COOS).
  - Moving different ModernStats models closer to each other (example: linking GSBPM and GSIM).
- (Re)integration of CSPA into the „big picture”.
- Clear revision procedures and principles for our ModernStats models: ModernStats Governance Guidance; available here.

Overview of potential 2022 activities



Are you interested?

- Please consider joining our Task Teams and help us to realise our ambitious work programme!
- Excellent opportunity to add your expertise and take part in the modernisation programme!

Acknowledgments

- Experts committed to the work of the Supporting Standards Group
- Chairs of the task teams
- UNECE and InKyung Choi for her invaluable support
- You

Thank you!

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