Geospatial view of
Generic Statistical Business Process Model

GeoGSBPM

InKyung Choi (UNECE)

@GFGS-UNECE-EFTA Webinar Coffee Talk (18 May 2021)
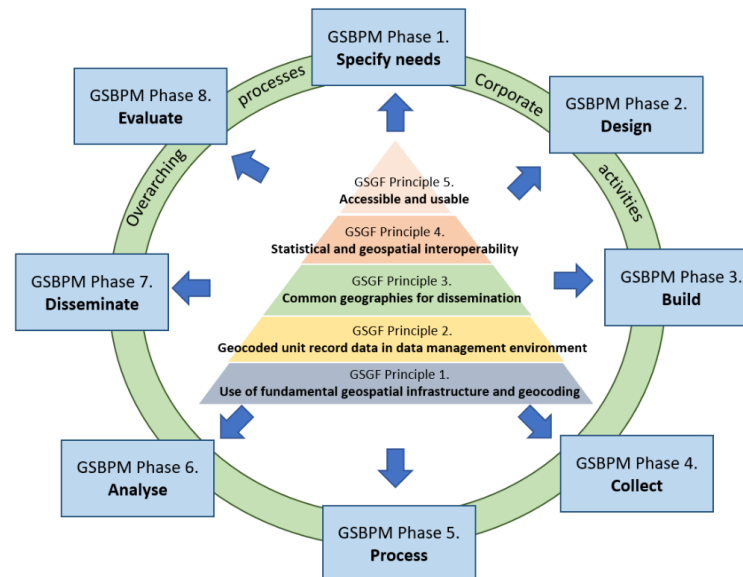
Content

1. Generic Statistical Business Process Model (**GSBPM**)

Content

1. Generic Statistical Business Process Model (**GSBPM**)

2. Geospatial view of GSBPM (**GeoGSBPM**)
   - 8 Phases
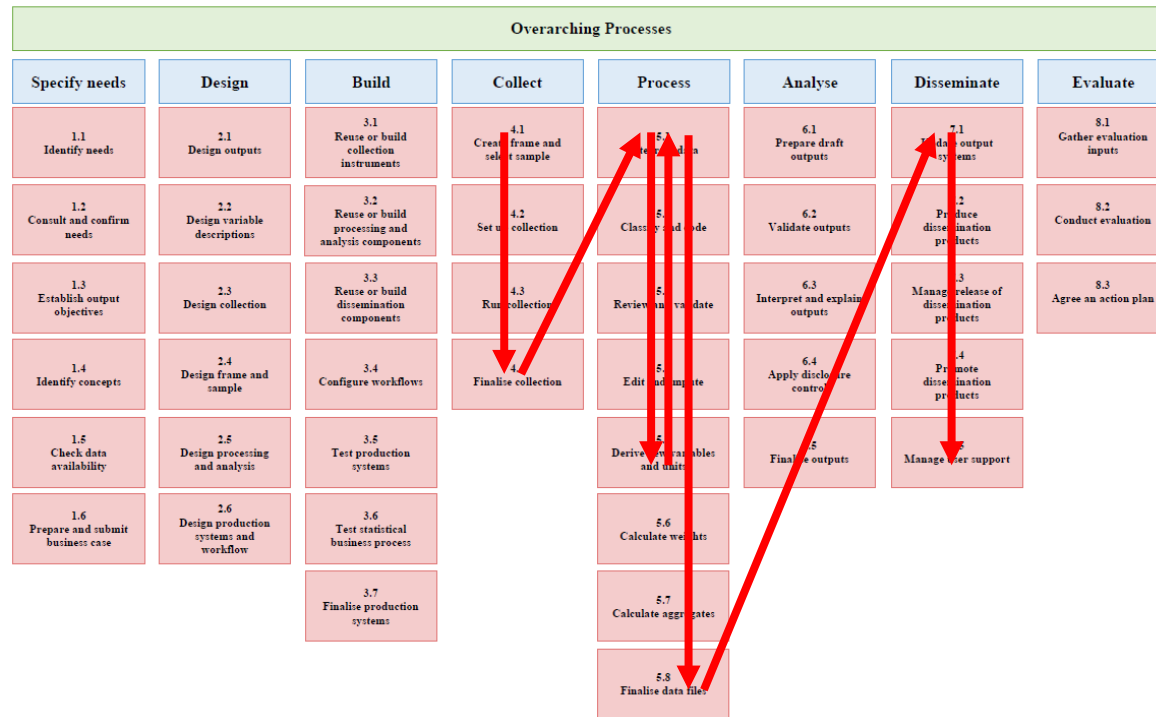   - Overarching process and corporate level activities

GSBPM

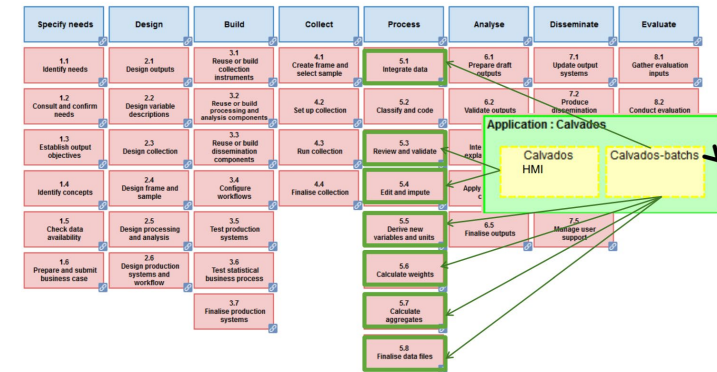## Generic Statistical **Business Process Model (GSBPM)**

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			

GSBPM

Generic Statistical Business Process Model (GSBPM)

Application modules can be linked to the different sub-phases of the GSBPM

GSBPM

Generic Statistical Business Process Model (GSBPM)

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			

Stovepipe vs. process oriented organization
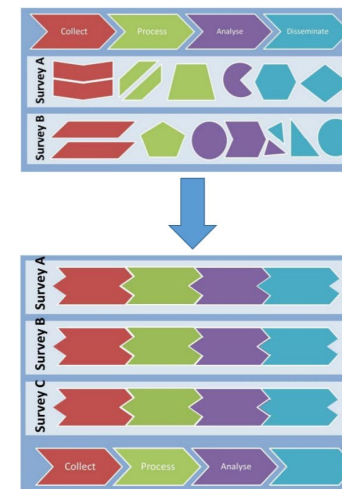
www.stat.gov.rs / stat@stat.gov.rs

Republic of Serbia presentation (UNECE ModernStats World Workshop 2018)

GSBPM

Usage of GSBPM is widespread or mature
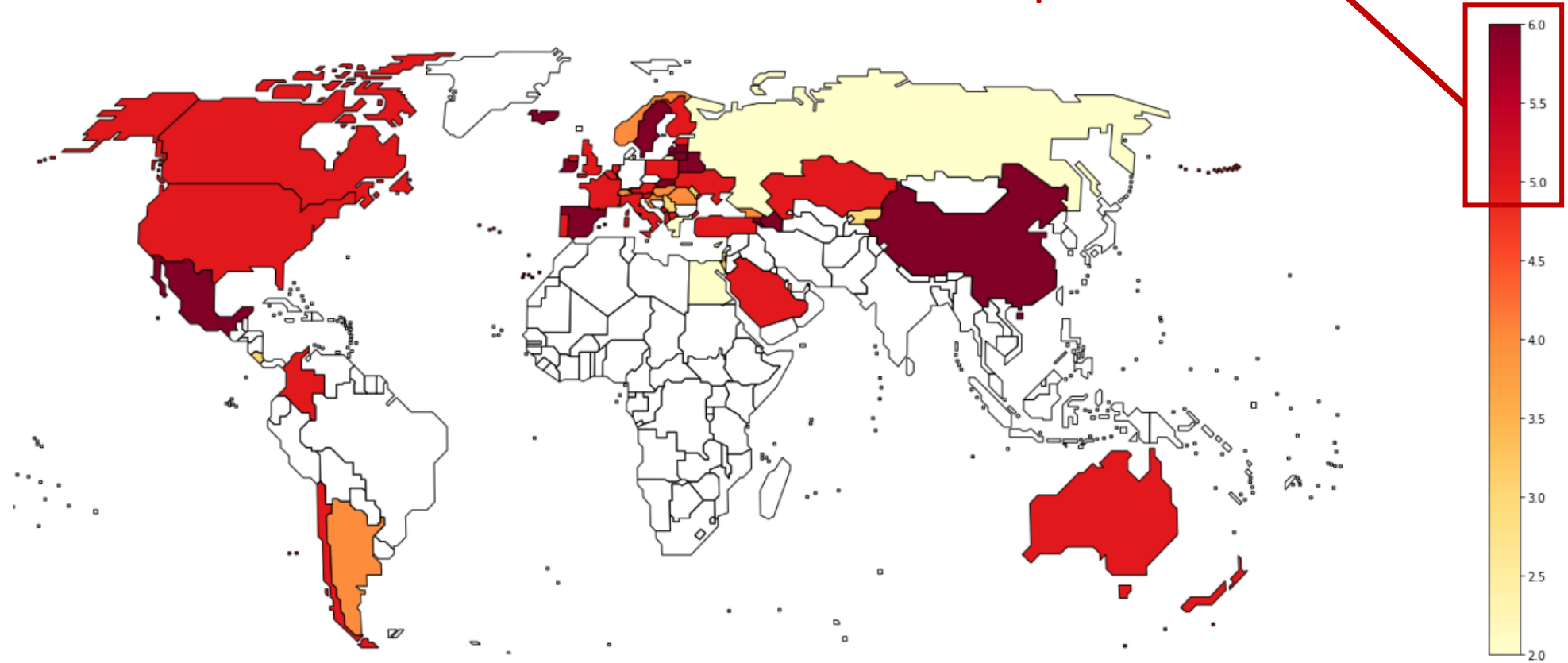
GSBPM and GSGF

GSBPM as tool to ensure GSGF principles to be followed

Immediate connection

GSBPM

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			

GSGF

5. Accessible and usable

4. Statistical and geospatial interoperability

3. Common geographies for dissemination of statistics

2. Geocoded unit record data in a data management environment

1. Use of fundamental geospatial infrastructure and geocoding

SOURCE: Australian Bureau of Statistics (ABS) / UN-GGIM, illustration by Statistics Sweden

Figure 2: The Global Statistical Geospatial Framework (GSGF)

Geospatial view of GSBPM (GeoGSBPM)

- Developed by Geospatial task team of HLG-MOS Supporting Standards Group
- GeoGSBPM describes geospatial-related activities and considerations using the framework of the GSBPM

Geospatial view of GSBPM (GeoGSBPM)

- Developed by Geospatial task team of HLG-MOS Supporting Standards Group
- 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. 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.

Geospatial view of GSBPM (GeoGSBPM)

- Developed by Geospatial task team of HLG-MOS Supporting Standards Group
- 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.

GSBPM original text

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.

Geospatial-related consideration and activities

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.

Geospatial view of GSBPM (GeoGSBPM)

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

GeoGSBPM – potential benefits

- Help production of geospatially enabled statistics to be conducted **in a systematic and consistent way**

- Provide a common framework to **manage quality and metadata** of statistical and geospatial information and services.

Linking GSBPM and GSIM Task Team Presentation (UNECE ModernStats World Workshop 2020)
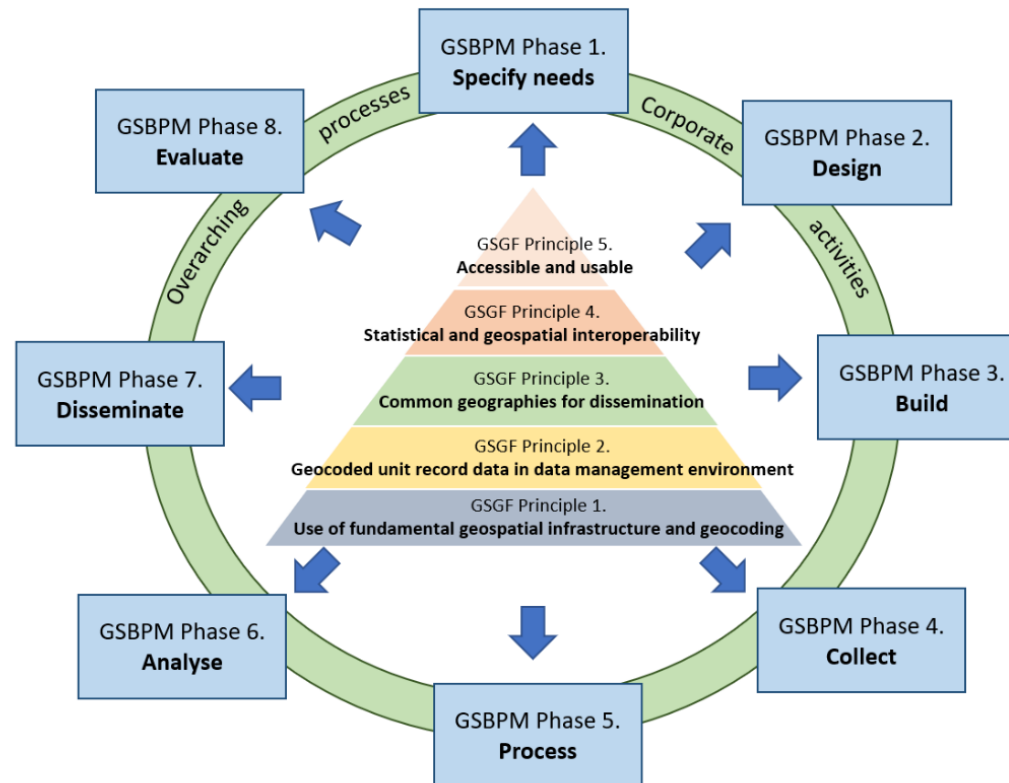
GeoGSBPM – potential benefits

- Help production of geospatially enabled statistics to be conducted **in a systematic and consistent way**

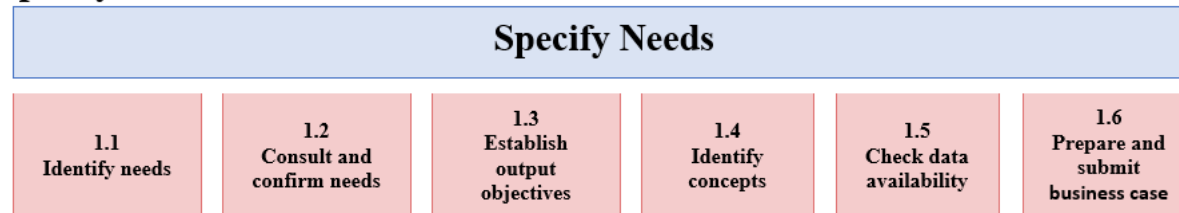- Provide a common framework to **manage quality and metadata** of statistical and geospatial information and services.

- Facilitate **sharing of geospatial services, methods and tools** that can be applied regardless of data types, domains and output formats
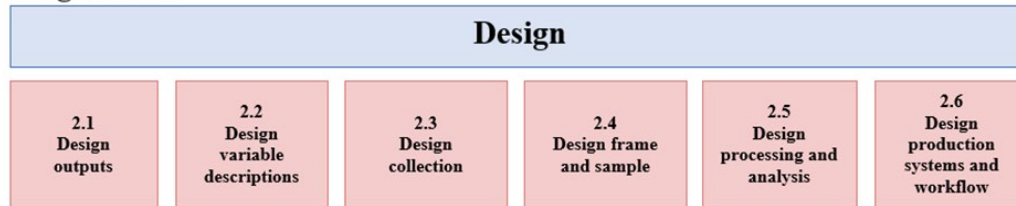
GeoGSBPM – Phase 1. Specify Needs

Specify Needs Phase

Specify Needs					
1.1 Identify needs	1.2 Consult and confirm needs	1.3 Establish output objectives	1.4 Identify concepts	1.5 Check data availability	1.6 Prepare and submit business case

- **Type of geography (e.g. grid, administrative boundary) and size of geography unit can have a significant implication on the cost of production, reliability and the risk of privacy breaches**, they have to be carefully examined from the very beginning of the production process

- **When assessing data availability, the existence and availability of suitable geospatial information should be first identified from authoritative sources with the National Spatial Data Infrastructure (NSDI)**

- Geospatial information may exist at the point-based level (e.g. x-y coordinates) or at the coarse area-based level (e.g. administrative boundary) and this should be compared with geospatial requirements that users specified (e.g. geography type, size, date of reference of the data, availability of time series).

GSGF 1.

1. Use of fundamental geospatial infrastructure and geocoding

GeoGSBPM – Phase 2. Design

Design Phase

Design

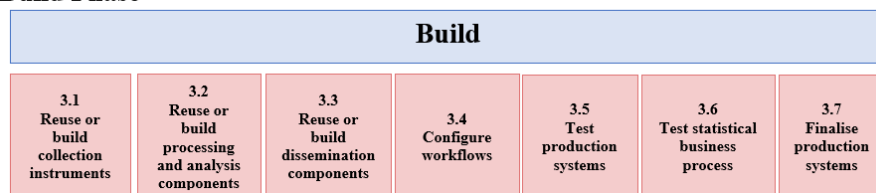| 2.1 Design outputs	2.2 Design variable descriptions	2.3 Design collection	2.4 Design frame and sample	2.5 Design processing and analysis	2.6 Design production systems and workflow

- Geographic viewing tools are a powerful way to help users to understand the geographic context of the issues that they are trying to solve with the data. Therefore, **it is recommended to include spatial visualisation and GIS services components in the output** ... **Accessibility and usability of geospatially enabled statistics and services can greatly increase by the use of standards and open data formats (e.g. XML, GeoJSON)**.

- Geospatial variables should be designed at the statistical unit level using **point-based location 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. DGGS) as it will greatly increase usability of the output.

- When geospatial information is collected along with data, the efficient and sustainable way to ensure quality is to make sure the information is accurate from the source. Therefore, design of the collection instrument should include a **point-of-entry validation tool**.
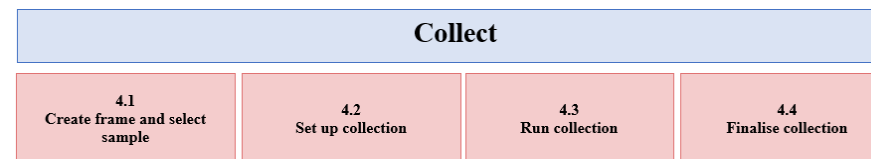
GSGF 5.
GSGF 2.

5. Accessible and usable

2. Geocoded unit record data in a data management environment

GeoGSBPM – Phase 3. Build / 4. Collect

Build Phase

Build						
3.1 Reuse or build collection instruments	3.2 Reuse or build processing and analysis components	3.3 Reuse or build dissemination components	3.4 Configure workflows	3.5 Test production systems	3.6 Test statistical business process	3.7 Finalise production systems

Collect Phase

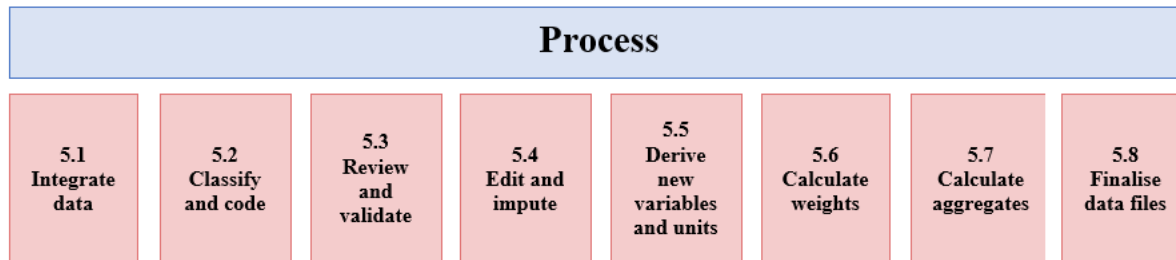| Collect			
4.1 Create frame and select sample	4.2 Set up collection	4.3 Run collection	4.4 Finalise collection

- While statistical services are often used mainly within statistical organisations, geospatial information services often have a much broader stakeholder group … **geospatial information and their related services may already exist in the inventory of these stakeholders and should be checked to avoid duplication of efforts** before building the components… If and when a new geospatial service is needed, it should be built in collaboration with NGIA and other stakeholders.

- **When inaccuracies of geospatial information are detected during the field collection (e.g. new settlement or district), this information should be documented and used to update the geospatial information systems of the statistical organisation, as it can affect downstream tasks.** Subject to maintaining statistical confidentiality, some corrections may also be fed back to mapping and cadastral agencies where appropriate.

GSGF 4.

4. Statistical and geospatial interoperability

GeoGSBPM – Phase 5. Process

Process Phase

Process

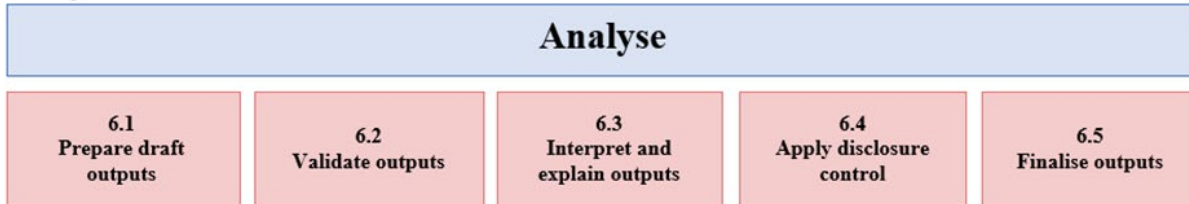| 5.1 Integrate data	5.2 Classify and code	5.3 Review and validate	5.4 Edit and impute	5.5 Derive new variables and units	5.6 Calculate weights	5.7 Calculate aggregates	5.8 Finalise data files

- Geospatial information (e.g. address, x- and y-coordinate, or a geographical name) can play an important role in bringing together information from various domains by enabling integration of datasets from different sources using the location information as a matching key variable (e.g. integrating administrative data with survey data using address or postal code that exists in both datasets). **To ensure the quality of the integration, standardising the geospatial information in the different datasets is critical. This standardisation would normally take place before the integration of datasets and can be done through, for example, matching location information in the datasets with a centralised standard system (e.g. address matching, geocoding) which should be part of the national spatial data infrastructure.** This linkage, ideally done through consistent, unambiguous and persistent identifiers (PIDs), can also allow the dataset to use various additional geospatial information within the address registry or geocode database.

GSGF 1.

1. Use of fundamental geospatial infrastructure and geocoding

GeoGSBPM – Phase 6. Analyse

Analyse Phase

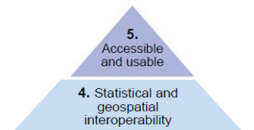| Analyse				
6.1 Prepare draft outputs	**6.2** Validate outputs	**6.3** Interpret and explain outputs	**6.4** Apply disclosure control	**6.5** Finalise outputs

- When preparing the analysis output, it is important to pay **attention to semantic interoperability, so that the output can be understood and used without ambiguities by users from different domains**. Cataloguing and tagging the content using relevant metadata standards greatly increases the usability of the analysis outputs.

- Geospatially enabled statistics, in particular for high spatial resolution, may carry a greater risk of privacy breach and require in-depth disclosure control. When an interactive mapping application is used, the tool should be configured in a way that users are not allowed to drill down to the spatial resolution level that statistical unit might be disclosed. Geospatial information adds a new dimension to data with which an individual statistical unit can be more easily identified in combination with other information. Therefore, a **geospatial product component should be cross-checked with other components (e.g. tabular aggregates, anonymised micro datasets) before release so that it does not breach any privacy on its own as well as in combination with other outputs.**
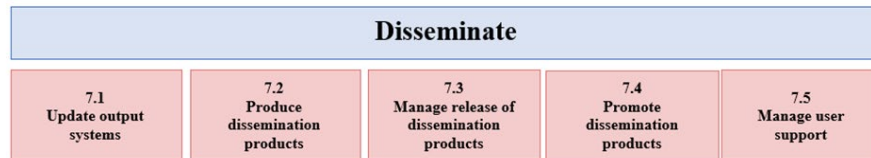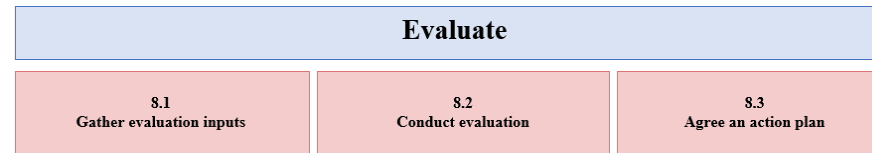
GSGF 5.

GSGF 4.

5. Accessible and usable

4. Statistical and geospatial interoperability

GeoGSBPM – Phase 7. Disseminate / 8. Evaluate

Disseminate Phase

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

Evaluate Phase

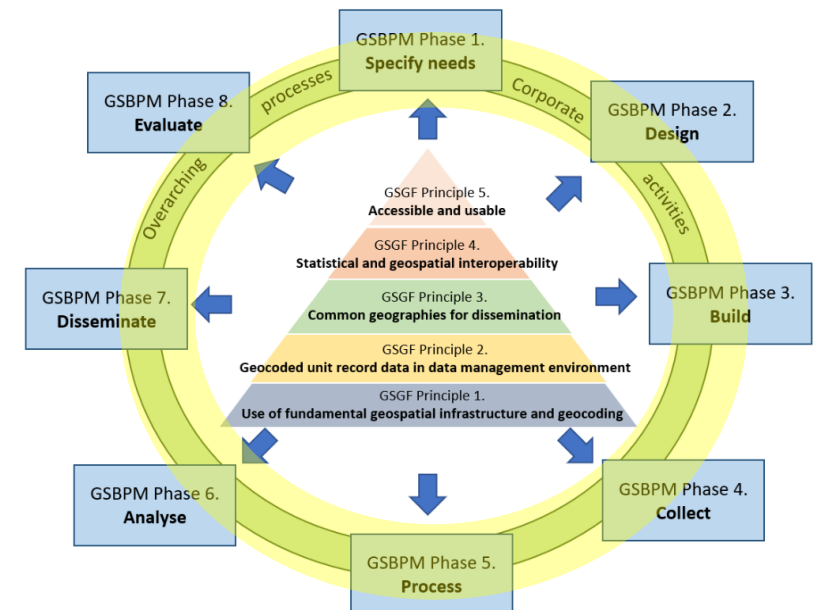| Evaluate		
8.1 Gather evaluation inputs	8.2 Conduct evaluation	8.3 Agree an action plan

- **Cataloguing and tagging the products using relevant metadata standards can greatly increase the discoverability and accessibility of the products both internally and externally. International standards should be used as a norm to ensure that the products can be found and consumed easily across a range of various user groups from public and private sector.**

- Geospatial information products could require **additional support** (e.g. tutorial video) as users may not be familiar with the data concept / type / structure, file format and associated GIS tools.

- It is important to note that the extent and depth of evaluation depend on the knowledge and experience of staff about the process under the evaluation. **As geospatial production processes may be less well-established and understood, a systematic quality framework and capacity building of staff may be required for a thorough evaluation**.
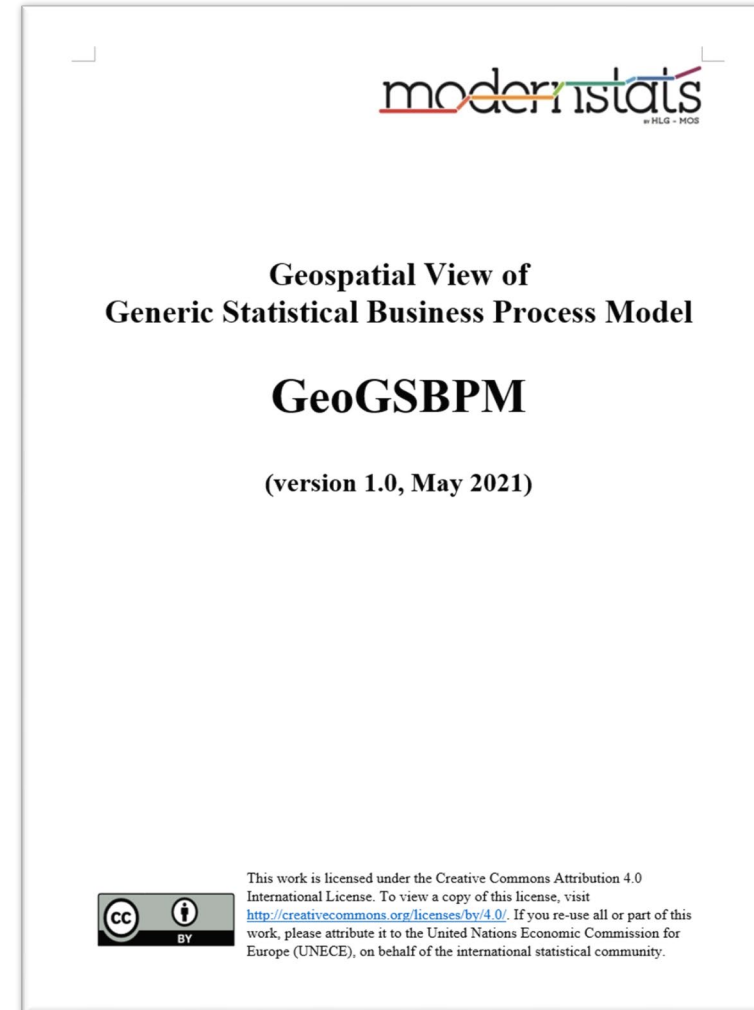
GSGF 4.

4. Statistical and geospatial interoperability

GeoGSBPM – Overarching processes and corporate-level activities

Activities that apply throughout the entire production process ("GSBPM overarching process") or that are conducted at a corporate level to support the production ("GAMSO corporate-level activities")

- **Strategic collaboration and cooperation**

- **Metadata management**

- **Quality management**

Thank you!

Geospatial View of
Generic Statistical Business Process Model

GeoGSBPM

(version 1.0, May 2021)

Available on GeoGSBPM wiki

(https://statswiki.unece.org/display/GSBPM/GeoGSBPM)