Geospatial View of
Generic Statistical Business Process Model
GeoGSBPM

ModernStats World Workshop 2022
27-29 June 2022
Belgrade, Serbia

Prepared by: Juan Muñoz
INEGI/SSG
• Version 1.0

• Published on May 2021

• Geospatial Task Team members from: Australia, Canada, Germany, Finland, France, Hungary, Ireland, Mexico, Poland, Republic of Korea, United Kingdom, DDI Initiative, Eurostat, ILO, UNSD and UNECE

• Can be downloaded from: https://statswiki.unece.org/display/GSBPM/GeoGSBPM
Context

• Primarily, authoritative geospatial information is produced by the National Geospatial Information Agencies (NGIAs) or mapping agencies.

• All data with a geographic location is a constituent component of the data ecosystem, which the national statistical organisations often finds itself as the custodian.

• The data ecosystem in which statistical organisations operate is more diverse than ever, there are various actors, from government agencies, private companies to citizens, producing data with different tools and in different formats.

• With digitalisation and advance of technologies, data are also being generated by non-human agents at an explosive rate (e.g. sensor data, data from web-crawler, mobility data from cell phones).

• The value of geospatially enabled statistics is not limited to the public sector.

• Wide use of map services through the web has lowered the access barrier to location information and changed the way it is used for decision making for all spheres of the society.

• Geospatially relevant activities and considerations should be integrated into the regular production processes of statistical organisations, so that the design and production of geospatially enabled statistics can be conducted in a systematic and consistent way.
Importance of Geographic Information for the Statistical Process

• Tool in the planning and operation of the statistical production process
• Element to group the statistical data and to determine the granularity of its disaggregation
• A key to relating information from different datasets when seeking interoperability
• Provider of geographical context of the phenomena described by the statistical data
• Visualisation instrument that increases the potential of understanding and analyzing information
• An essential component for delivering the value of statistical information
Georeferencing Statistics
Georeferencing Statistics

- Number of dwellings
- Number of inhabitants
- Number and location of Business units
Georeferencing Statistics

Urban infrastructure in a city

Selection of a specific geographical area of interest

Number of inhabitants, health related business, and transit accidents

Statistics of the selected area
Scope of the document

• Expect to find:
  • Descriptions of geospatial-related activities needed to produce geospatially enabled statistics, using the framework of the GSBPM.
  • Recommendations related to the actions needed to collect, process and analyse geospatial information mapped to the phases and subprocesses of the production process described by GSBPM.

• Not expect to find:
  • A standard notation or a file format specification to describe geospatial information.
  • Recommendations of visualisation methods of georeferenced statistics.
Underlying Frameworks

- Generic Statistical Business Process Model (GSBPM), developed by the HLG-MOS. Coordinated by UNECE

- Global Statistical Geospatial Framework (GSGF), developed by the United Nations Expert Group on the Integration of Statistical and Geospatial Information (UN EG-ISGI), endorsed by the Statistical Commission. Coordinated by UNSD
<p>| GSGF 1. Use of fundamental geospatial infrastructure and geocoding | GSBPM Phase | Specify Needs | GSBPM Phase | Design | GSBPM Phase | Collect | GSBPM Phase | Process | GSBPM Phase | Analyse | GSBPM Phase | Disseminate | Overarching processes / Corporate activities |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| When assessing data availability, the existence and availability of suitable geospatial information should be first identified from authoritative sources within the National Spatial Data Infrastructure (NSDI) | GSBPM Phase Specify Needs | Specify Needs | GSBPM Phase Design | Design | GSBPM Phase Collect | Collect | GSBPM Phase Process | Process | GSBPM Phase Analyse | Analyse | GSBPM Phase Disseminate | Disseminate | Overarching processes / Corporate activities |
| Geospatial variables (geographies) should be designed for the statistical unit level. Using point-based location as the base geospatial variable will provide considerable adaptability to changes over time and flexibility to aggregate up to various dissemination level geographies | GSBPM Phase Design | Specify Needs | Geospatial variables (geographies) should be designed for the statistical unit level. Using point-based location as the base geospatial variable will provide considerable adaptability to changes over time and flexibility to aggregate up to various dissemination level geographies | Design | Collect | Collect | Process | Process | Analyse | Analyse | Disseminate | Disseminate | Quality management include: identify the authoritative (external or internal) sources of reference data and establish quality profile of reference data |
| Geocoding should be conducted for each statistical unit that is collected and at the most detailed level (e.g. point-based geocoding as opposed to area-based geocoding) | GSBPM Phase Collect | Collect | Geocoding should be conducted for each statistical unit that is collected and at the most detailed level (e.g. point-based geocoding as opposed to area-based geocoding) | Collect | Process | Process | Process | Process | Analyse | Analyse | Disseminate | Disseminate | Quality management include: identify the authoritative (external or internal) sources of reference data and establish quality profile of reference data |
| Standardisation should take place before the integration of datasets. It can be done through, for example, matching location information in the datasets with centralised standard systems (e.g. address matching, geocoding) which should be based on the national geospatial information context | GSBPM Phase Process | Process | Standardisation should take place before the integration of datasets. It can be done through, for example, matching location information in the datasets with centralised standard systems (e.g. address matching, geocoding) which should be based on the national geospatial information context | Process | Analyse | Analyse | Analyse | Analyse | Disseminate | Disseminate | Disseminate | Disseminate | Quality management include: identify the authoritative (external or internal) sources of reference data and establish quality profile of reference data |
| GSGF 2. Geocode unit record data in a data management environment | GSBPM Phase Disseminate | Disseminate | GSBPM Phase Design | Design | GSBPM Phase Collect | Collect | GSBPM Phase Process | Process | GSBPM Phase Analyse | Analyse | GSBPM Phase Disseminate | Disseminate | Overarching processes / Corporate activities |
| The design of components includes: point-of-entry validation for geographical information; matching strategy; and, spatial analysis | GSBPM Phase Disseminate | Disseminate | The design of components includes: point-of-entry validation for geographical information; matching strategy; and, spatial analysis | Design | Collect | Collect | Process | Process | Analyse | Analyse | Disseminate | Disseminate | Quality management include: develop quality dimensions and metrics to be used at different stages, and a consistent matching strategy |
| The mechanism of matching or geocoding the statistical unit-record established in Design phase should be consistently applied | GSBPM Phase Analyse | Analyse | The mechanism of matching or geocoding the statistical unit-record established in Design phase should be consistently applied | Analyse | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Quality management include: develop quality dimensions and metrics to be used at different stages, and a consistent matching strategy |
| GSGF 3. Common geographies for production and dissemination of statistics | GSBPM Phase Analyse | Analyse | GSBPM Phase Disseminate | Disseminate | GSBPM Phase Collect | Collect | GSBPM Phase Process | Process | GSBPM Phase Analyse | Analyse | GSBPM Phase Disseminate | Disseminate | Overarching processes / Corporate activities |
| Needs of users in terms of geographies (e.g. size of unit, type) is discussed. Implications (e.g. cost, reliability, quality) should be communicated and consulted with users | GSBPM Phase Collect | Collect | Needs of users in terms of geographies (e.g. size of unit, type) is discussed. Implications (e.g. cost, reliability, quality) should be communicated and consulted with users | Collect | Process | Process | Process | Process | Analyse | Analyse | Disseminate | Disseminate | Overarching processes / Corporate activities |
| When grid geographies are used, the choice of grid system should take existing regional and global systems into consideration | GSBPM Phase Process | Process | When grid geographies are used, the choice of grid system should take existing regional and global systems into consideration | Process | Analyse | Analyse | Analyse | Analyse | Disseminate | Disseminate | Disseminate | Disseminate | Overarching processes / Corporate activities |
| Inaccuracies in geospatial information detected during field collection should be documented and transferred to the central geospatial information system for maintenance and update if necessary (if permitted under statistical confidentiality rules) | GSBPM Phase Analyse | Analyse | Inaccuracies in geospatial information detected during field collection should be documented and transferred to the central geospatial information system for maintenance and update if necessary (if permitted under statistical confidentiality rules) | Analyse | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Disseminate | Overarching processes / Corporate activities |</p>
<table>
<thead>
<tr>
<th>GSBPM Phase</th>
<th>Overarching processes / Corporate activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specify Needs</td>
<td>Alignement and harmonisation of geospatial metadata concepts with those of statistical metadata is critical</td>
</tr>
<tr>
<td>Design</td>
<td>Design of all production components should take into account standards used in the geospatial community</td>
</tr>
<tr>
<td>Build</td>
<td>Geospatial services have a broad stakeholder group, statistical organisations should check and consult with service inventories of stakeholders before building components on their own</td>
</tr>
<tr>
<td>Collect</td>
<td>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</td>
</tr>
<tr>
<td>Process</td>
<td>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 the public and private sectors</td>
</tr>
<tr>
<td>Analyse</td>
<td>Cataloguing and tagging the content using relevant metadata standards can greatly increase the usability of the analysis outputs. Geospatial product components should be cross-checked with other components (e.g. tabular aggregates, before release so that they do not breach privacy on their own as well as in combination with other outputs</td>
</tr>
<tr>
<td>Disseminate</td>
<td>Statistical organisations are encouraged to explore the semantic web standards as a long-term strategic objective with successive milestones to achieve dissemination of data and metadata within the framework of Linked Open Data (LOD)</td>
</tr>
</tbody>
</table>

**GSGF 4. Statistical and geospatial interoperability - data, standards and processes**

- **Design of all production components should take into account standards used in the geospatial community**
- **Geospatial services have a broad stakeholder group, statistical organisations should check and consult with service inventories of stakeholders before building components on their own**
- **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**
- **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 the public and private sectors**

**GSGF 5. Accessible and usable geospatially enabled statistics**

- **Discussion on the output format is useful as users for high spatial resolution data (e.g. city, municipal authority) might require data to be provided in certain formats that are digestible within their GIS system. Implications of the size of geographic units in terms of confidentiality risk should be discussed with users**
- **Design of these outputs should also take potential downstream uses into consideration. Accessibility and usability of geospatially enabled statistics and services can greatly increase by use of standards and open data formats**
- **Metadata elements are put together during development of dissemination components so that they can be disseminated along with the data products and services. To make it more findable and accessible for both internal and external users, metadata should be documented using standard taxonomy and vocabulary**
- **Cataloguing and tagging the content using relevant metadata standards can greatly increase the usability of the analysis outputs. Geospatial product components should be cross-checked with other components (e.g. tabular aggregates, before release so that they do not breach privacy on their own as well as in combination with other outputs**
Structure of the GeoGSBPM

General Description of Each GBPM Phase with annotations about relevant aspects about geographical information

2.5 Design processing and analysis

46. This sub-process designs the statistical processing methodology to be applied during the "Process" and "Analyse" phases. This can include among others, specification of routines and rules for coding, editing and imputation which may vary based on the mode of data collection and source of data.

47. Similar to statistical data, geospatial information obtained during the collection phase will undergo a range of processing (e.g., editing, imputation, validation) in particular, those from the third party as different sources might store the geospatial information in different ways. Geospatial information is often used as a key variable for integrating data from various sources, hence, the standardization of geospatial information to enable its use across the production process and the understanding of its quality are critical. Matching and non-matching strategy for integration and record linkage is developed in this sub-process. This sub-process includes design of processing methodologies specifically needed for geospatial information (e.g., point-in-polygon processing) as well as other geospatial services.

48. This sub-process also includes design of specifications for data integration from multiple data sources, validation of data and estimation Statistical disclosure control methods are also designed here if they are specific to this business process.

49. Geospatially enabled statistics, in particular at a high geospatial spatial resolution, can allow statistical organisations to produce analytical outputs at more disaggregated levels and conduct a wide range of spatial analysis (e.g., map visualization, spatial-temporal regression). Design of such analysis and analytical output can be conducted in this sub-process.

Reference to the Phases, Sub-Processes and Overarching Processes contained in GBPM
Discover the value of GeoGSBPM

- Satisfying expectations of the users by providing granular and flexible statistical information services.
- Ensuring interoperability of statistical and geospatial information.
- Building better tools and services to improve the quality of geo-referenced statistics.
- Geo-referencing Statistics to improve their analysis.
- Using geographical information to help in the collection of information.
- Enhancing the dissemination by providing geo-referenced views of the Statistics.
- Improving the integration and processing of data from different sources.
- Applying an enriched framework to evaluate the production of geo-referenced statistics.
Thank You

Comments, Questions, Any Feedback

Juan.Munoz@inegi.org.mx