Annex 2. Results from national exercises on the GSBPM

Final report from the GEOSTAT 2 project

This annex is a part of the final report from the GEOSTAT 2 project. The results presented in this annex are linked to Chapter 7 of the main report.

Content

1 Implementing the GSBPM at Statistics Finland

1.1 The value and position of the GSBPM at Statistics Finland

At Statistics Finland the GSBPM was officially adopted as a framework for statistical production in 2015. The interest in the model grew over several years. However, Statistics Finland did not belong to the groups developing the standard but has participated in some later discussions. The Standards and Methods unit is responsible for the GSBPM at Statistics Finland. Other units use it, for instance, in documentation and in development projects. The organisational structure and the Intranet follow the GSBPM to a certain extent.

At the Standards and Methods unit it is seen that the GSBPM gives valuable benefits for the production of statistics. It gives a way of thinking about statistical production and a common vocabulary enabling cooperation, also internationally. Certain challenges are also seen. For example, common tools or templates to describe a certain statistical process are needed. For statistics based on administrative data sets, some clarification is also needed. The GSBPM has been a natural starting point for implementing statistical standards. It has paved the way for implementing other standards, such as the CSPA and the GSIM. The GSBPM is also seen as an important basis for building coherent service based IT architecture.

Already, the GSBPM has been in use in various production renewals but not in any relation with geospatial data.

1.2 The scope of the assessment and performing the tests

Statistics Finland's responsibility in the project was to formulate a common approach how to test the GSBPM from the geospatial point of view in order to have comparable results.

An IT architect took the role of a facilitator and expert of the GSBPM. The project group was also supplemented with an additional geospatial specialist. To the chosen approach ten-step instructions were gathered (Annex 3) for all project countries.

The content and results of testing at Statistics Finland are presented in Figure 3

1. Getting familiar with the GSBPM (Identification of the dimensions of Geospatial data)

    - The work from the geospatial point of view started with the GSBPM expert responsible for the model's adaptation and implementation at Statistics Finland.
    - The first objective was to understand the GSBPM in relation to geospatial data. By examining the production processes of different data sets (on a very coarse level), three different dimensions started to emerge: 1) the phases related to plain statistics (blue level), 2) the phases related to plain geospatial data (red level), and 3) the phases in between these two, the integration level (green level) (Figure 12 in Chapter 7). The integration of statistics and geospatial data requires an established code of practice, adequate technologies, as well as standards in order to function through the process.
    - The first finding of testing was that geospatially related phases may occur in all steps of the process.

2. Process descriptions of Geospatial data production processes (of population grid data, of any geospatial related statistical data), -> **a geospatial production process**: a *statistical geospatial* production process, a plain *geospatial* production process.

    - The work was carried out through a series of workshops, where storytelling workshop techniques were used. The identification of the phases and arranging the phases were done only with *Post-it notes.* The process flow was established and accepted by all members of the group before graphical software was used. The target was to have *a common understanding* of the process *before matching* the phases to the GSBPM.
    - The first process flow made described the production of statistical grid data. However, the next data in the target (statistics by postal code areas) showed an unexpected similarity. The finding was that the first geospatial related process described was also able to cover the other processes– only by adding a couple of new steps to the process flow. *A geospatial statistical production process* was described.
    - When the graphical presentation was created it was decided to divide the process into two separate process flows. The first process flow describes *the statistical geospatial process as a whole*, the other describes *the process relating directly to geospatial data*. These two process flows cross at particular points, because the geospatially related phases are often necessary in a specific stage of the main statistical geospatial process. (Figure 1)
    - The terms that are used should be generally understood. Therefore, the language that is in use in the original production situation was used.

3. Identification of correspondent GSBPM phases (Redefining of processes, handling duplicates and missing process phases)

    - The aim was to test how well the GSBPM could cover the original process phases.
    - The search of the corresponding GSBPM phases was quite challenging. There are no references to geospatial data. Tests were performed by raising the abstraction level of the documentation. The idea and the mission of a GSBPM phase was evaluated at a more general level. For example, the integration of geospatial data and statistics was seen as equivalent with the phase where two ordinary statistical data sets are joined.

4. Identification of dimension (Geospatial (red), Integrated (green), Statistics (blue))

    - The identification of the dimension of a particular phase may offer a tool to show and discuss where the geospatially related phases occur, as well as what the role of such phases is from the point of view of the whole process. When the renewals of these targets are the case, they are easily pointed out from the process whether the renewing concerns statistics, integration or geospatial dimension of the phase.

5. A synthesis of the geospatial statistical process by the GSBPM phases (Geospatially relevant GSBPM phases and sub-processes placed to the GSBPM framework)

    - *The most relevant process phases* were reorganised under the GSBPM's phases (Figure 2) by using the identified GSBPM phase numbers.

1.3 Statistics Finland's remarks

1) The production of geospatial statistics seems to fit into the GSBPM. However, in order to recognise the GSBPM phases, plenty of interpretation is needed first. The level of abstraction that is essential for understanding the process must be chosen. Raising the abstraction level of the GSBPM's documentation and adding references to geospatial issues would also help to locate process phases for each phase of the GSBPM.
2) All geospatial related production processes could fit into the same process flow (regardless of the focus and content of data). However, for clarity reasons, the production process may require separation of the main statistical production process from the process phases that focus entirely on geospatial data. In the case of Statistics Finland, a clear division was made and was seen as clarifying the complexity of the process. The main statistical production process may be supported by one or several sub-processes focusing entirely on geospatial data processing.
3) The GSBPM phases also cover the purely geospatial process phases (if interpreted by the logical meaning of the particular phase). A facilitator is needed to examine the process phases and the corresponding GSBPM phases.
   A facilitator is a person who guides the exercise and keeps the work on the right track. He/she understands the aims but does not have to know the content. He/she knows the methods for team working and he/she is an inspiring guide. If the team gets stuck on details, the facilitator helps move on and keep focus on the overall process to be defined.
4) The same GSBPM phase may occur many times, e.g. in process phases 5 to 5.8: Actions to produce statistical data and actions for integration of statistics and geospatial boundary data.
5) When creating the process flow, the genuine process phases are revised when the GSBPM is included. After finding the fitting GSBPM phases some missing phases may be detected. It is necessary to allow new phases to appear or irrelevant ones to be removed.
6) It is extremely difficult to settle the way to use the GSBPM model other than at the organisation level. The use of the model must be agreed on prior to common utilisation because the GSBPM documentation does not include geospatial related instructions.

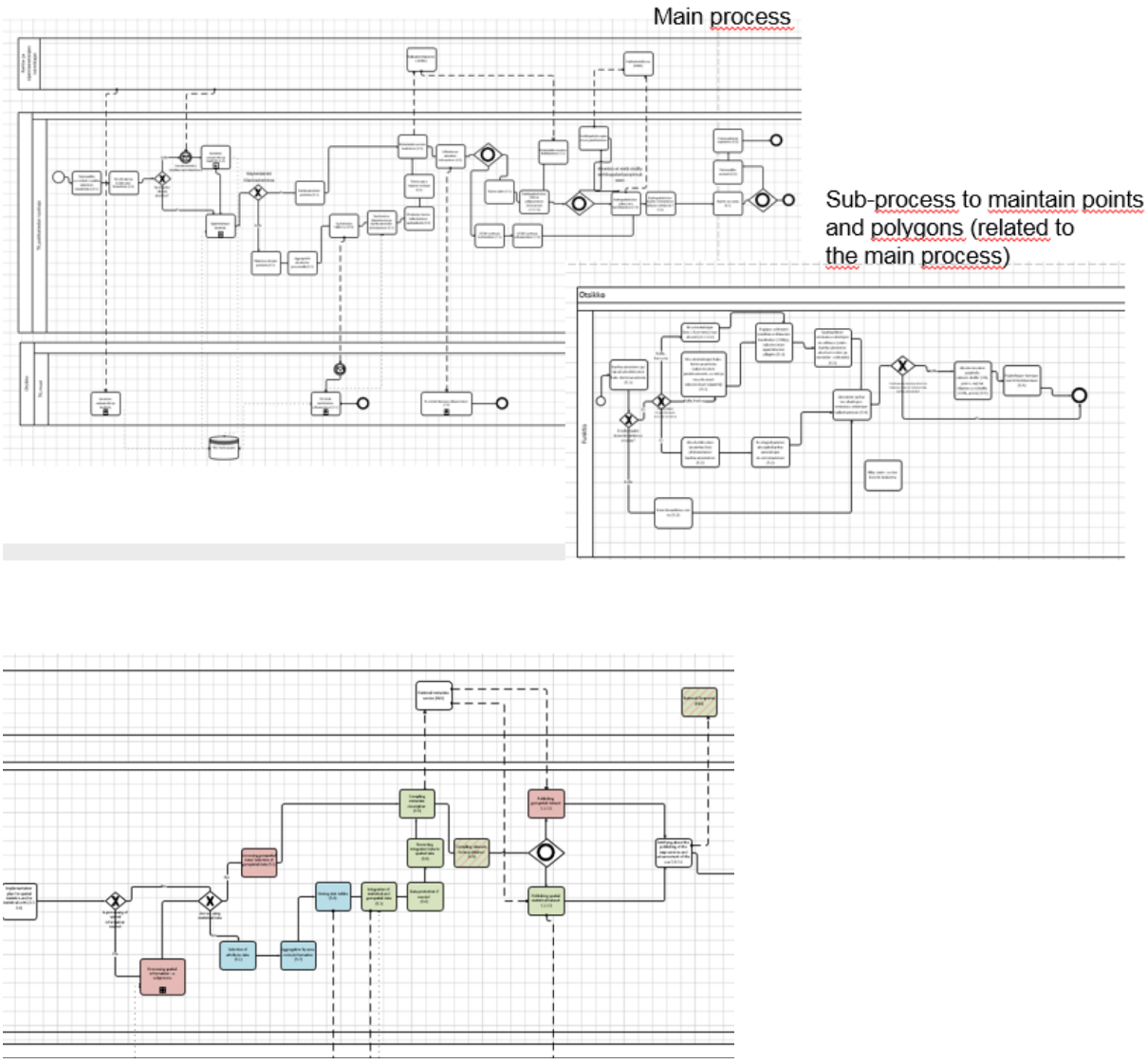

Figure 1: An example of a statistical production process where geospatial data are involved. The main statistical production process was separated from the elaborated process of maintaining purely geospatial data. The main process is also illustrated with colours which indicate data dimension in a particular process phase (Geospatial (red), Integrated (green) and Statistics (blue)).

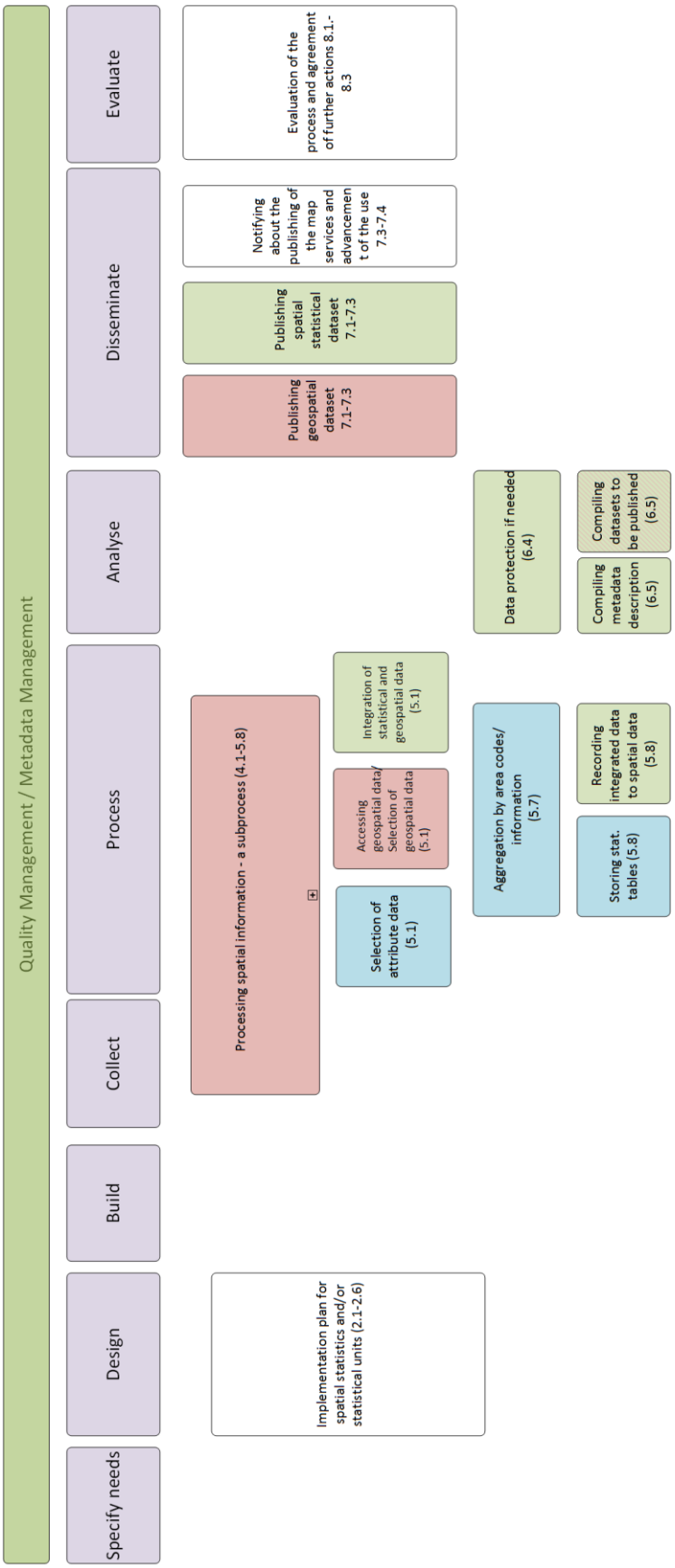

Figure 2: A synthesis of geospatially related production process of statistics set in corresponding GSBPM phases. Results are preliminary and made for testing the GSBPM in the Geostat 2 project.

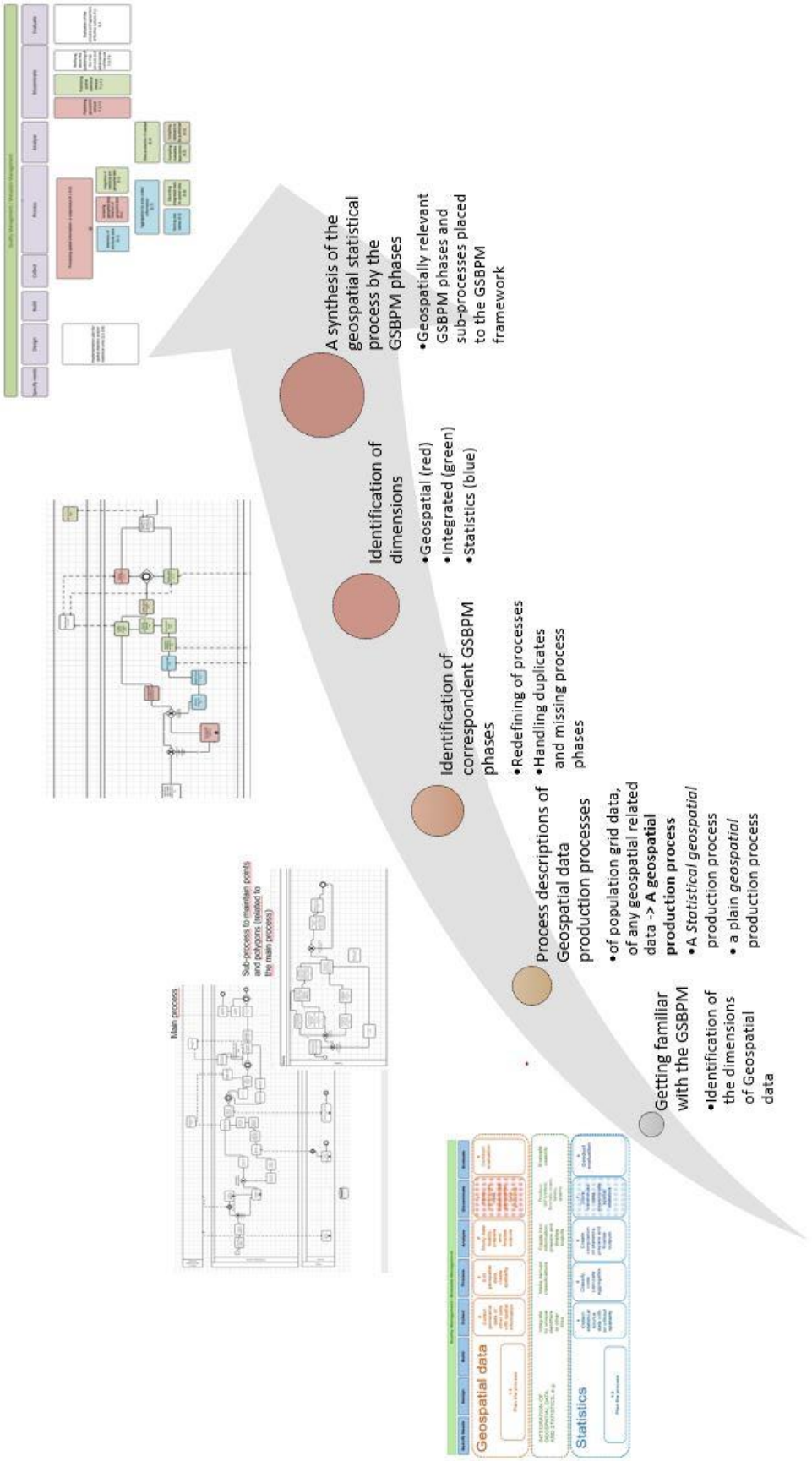

Figure 3: Stages of testing geospatial related statistical production processes at Statistics Finland.

2 Implementing the GSBPM at Statistics Sweden

2.1 Existing process model, process chart and process architecture

In 2008 Statistics Sweden adopted a process-based approach to its operations. The aim was primarily to reduce unnecessary variation in the statistical production and lowering the cost of IT support. Therefore Statistics Sweden has during the last couple of years focused on the statistics production process which is based on the GSBPM.

But a business does not only consist of one core process. Statistics Sweden has, just as other businesses and government agencies, several processes needed to support the statistical production, which can't be found in the GSBPM such as manage and develop as examples. Much work has also been done to describe these processes.

To bring control and develop the business processes one need to know how they look. Statistics Sweden´s process chart gives an overall picture of Statistics Sweden´s total operations and how different parts fit together and interact. The process chart is also a basis for designating responsibility and ownership and various process description initiatives can be placed into the chart.

The purpose of the process chart is not to conduct a detailed mapping of Statistics Sweden´s entire activities but instead focus on the processes that are involved in business development. The process chart gives an overall picture of operations and how different parts fit together and interact. It is divided at the top level in the four business areas which are:

- Lead and manage operations
- Develop products, services and abilities
- Deliver products and service
- Support operations

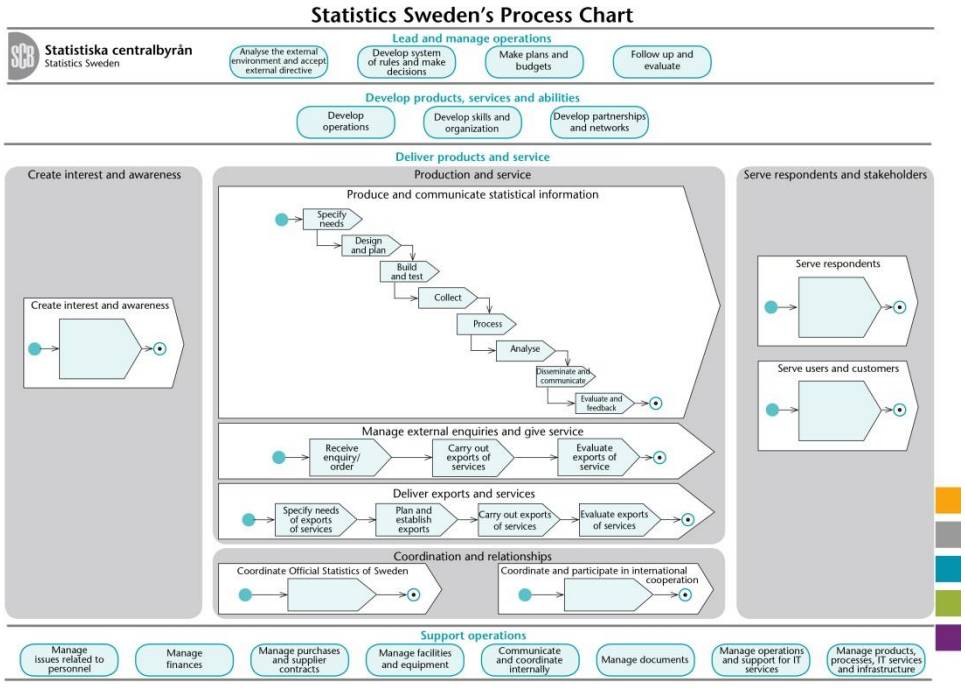

Figure 4: Statistics Sweden´s Process Chart

Processes that provide value to external customers are described in *Delivering Statistics Sweden's products and services*. Processes to govern and manage Statistics Sweden are described in *Lead and manage operations* and the results of these processes are used in all other processes. Processes that are designed to develop the business are described in *Develop Statistics Sweden's products, services and abilities*. Processes needed for support to the other processes are described in *Support operations*.

The process chart with its underlying uniform process descriptions are an important component to streamline development by providing an overall picture where

- Process managers can be appointed
- IT support can be linked to processes
- You get an overview of where the development work is carried out to prevent the development being carried out in several different places
- It is possible to identify reusable components.

The process chart should also be used as a basis when measuring and analysing the processes.

The process *Production and service* within the process *Deliver products and services* is the same as in the GSBPM. Statistics Sweden has also eight core processes to describe the statistical production with the underlying processes being more detailed in the Swedish process chart.

Statistics Sweden's process chart is complemented by a process architecture that consists of six levels in order to easier identify and develop the processes.

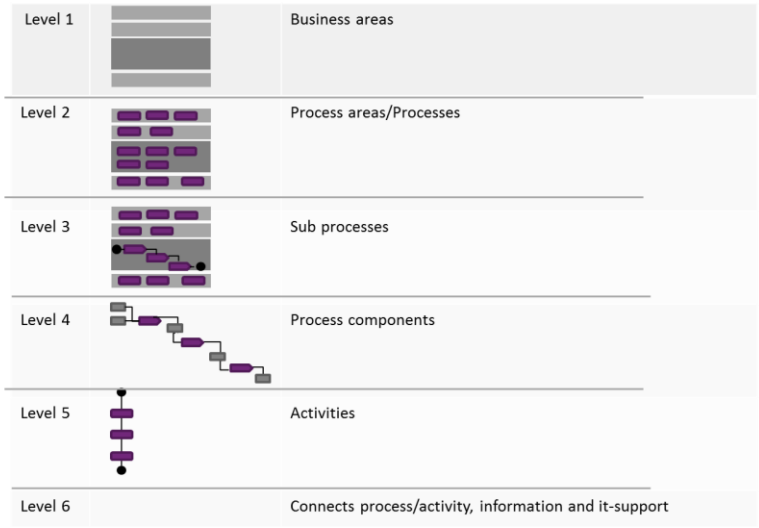

Figure 5: Statistics Sweden´s process architecture

2.2 The scope of the assessment

It was decided to undertake the assessment of the GSBPM in relation till geospatial information management in the form of a workshop with participation of geospatial experts and process architects. Taking into account the limited time at disposal, it was not possible to map a complete work flow for geospatial statistics production. The approach was rather to look at each process of the domestic process chart/GSBPM and to assess the relevance of the processes as regards to geospatial

information management and to identify possible gaps in the process model. Further it was decided to look at management of geospatial information (as is) through the production process model and without special attention to a particular statistical product. Considering the scope of the GEOSTAT 2 project the main priority was given to the first 5 phases of the model, though all phases were discussed.

2.3 Statistics Sweden's remarks

Below is a brief summary of the most relevant topics that were discussed during the workshop. The remarks are structured according to the phases of the GSBPM. Some of the topics discussed are generic in relation to the GSBPM. They are presented under *General remarks*.

Phase 1. Specify needs

The *Specify needs phase* does not pose any particular challenges in terms of geospatial information management. The sub-processes are generic and reflect considerations to be made regardless of the information to be handled or output to be created. However, while users of sectorial statistics may more easily be identified for consultation on their needs and expectations, the stakeholders in geospatial statistics represent a very heterogeneous community. Geospatial statistics products cut cross various sectors (economy, demography, spatial planning, environment etc), framed by no common policy context. Hence, a structured consultation with the user community may be a bigger challenge for producers of geospatial statistics.

Phase 2. Design

On sub-process level a few issues came up for discussions during the assessment. One conclusion was that *Design output* does not fully recognise production of geospatial statistics. The GSBPM is generic in this sense, whereas Statistics Sweden's process chart is more detailed, and the output products mentioned does not address geospatial products. Integration of geography and statistics entails a much richer flora of potential outputs than those traditionally mentioned (tables, charts etc).

It was also concluded that there is a general need to acknowledge the power of geography in the design phase. In Statistics Sweden sample surveys entailing geospatial data is not a common approach, partly because vast access to geocoded register data overrides this need, but also partly because the knowledge about the potential of geographical sampling is low. In many countries sample survey based collection of core statistical data is a common approach, thus geographical sampling deserves to be promoted in order to raise awareness.

Phase 3. Build

As for the *Specify needs phase*, the *Build phase* does not call for any particular changes in regards to geospatial information management, at least not on an overarching level. However on sub-process level there were a few issues discussed. Typically for geospatial statistics applications, *Build collection instrument* is not a relevant sub-process, as direct collection of geospatial data is not conducted. Processing of remote sensing data was discussed as something in between directly collected and obtained data. The raw data (satellite imagery itself) is acquired but in order to transform the spectral data to meaningful information, image classification algorithms are required. This type of transformation of raw data is not recognised in a proper way.

Phase 4. Collect

The *Collect phase* of the GSBPM (as well as the corresponding collect phase in Statistics Sweden's process chart) is mainly focused on direct collection of survey data. From a geospatial data management perspective, data collection is rather a matter of acquisition of data from NMCAs or other data producing institutions. Depending on the objectives of the output, geospatial information can be combined with data from direct collection. However this is not the typical case in Statistics Sweden.

The weak support for non-survey data sources in Statistics Sweden's process chart has been recognised also from the point of view of register based statistics. In essence, people involved in register based statistics and geospatial experts share some common concerns regarding the needs to incorporate guidance for non-survey sources under all phases. Recently an effort was made to extend Statistics Sweden's process chart with better guidance for register based statistics. Such effort would be advisable also for the GSBPM.

Phase 5. Process

Many of the typical operations associated with spatial analysis are found under the *Process phase*. Except from the problems emerging from semantic differences (se *Analyse* below) the *Process phase* seems to mirror quite well the work flow of geospatial statistics. Of particular interest is sub-process 5.1 *Integrate data*, as this is the hallmark of geospatial data processing. The description of this sub-process would benefit from a more explicit reference to integration of statistical and geospatial information.

Phase 6. Analyse

A generic problem has been identified emerging from semantic discrepancy between terminologies found in the statistical and geospatial community. The term *Analyse* according to the GSBPM is a good example. The *Analyse phase* is described as follows:

"In this phase, statistical outputs are produced, examined in detail and made ready for dissemination. It includes preparing statistical content (including commentary, technical notes, etc.), and ensuring outputs are "fit for purpose" prior to dissemination to customers. This phase also includes the sub-processes and activities that enable statistical analysts to understand the statistics produced."

From a spatial analysis perspective, *analyse* is rather recognised as the process where properties of spatial objects and their relations to other objects are described. *Analyse* entails a vast variety of operations from simple point-in-polygon operations to more advanced distance and density calculations. The process that a spatial analyst would refer to as *analyse* is rather occurring in the *Process phase* according to the GSBPM.

The description found under sub-process 5.5 *Derive new variables and units* is another interesting case as it captures the denotation of spatial analysis:

"This sub-process derives data for variables and units that are not explicitly provided in the collection, but are needed to deliver the required outputs. It derives new variables by applying arithmetic formulae to one or more of the variables that are already present in the dataset, or applying different model assumptions".

Spatial relations are typically derived variables. In a case where population proximity to green areas is to be calculated, dwelling location of population and green areas are collected or mapped data, whereas the distance between population and green areas must be derived through spatial analysis. In a geospatial sense, distances between objects (population and green areas) are *analysed*.

Phase 7. Disseminate

Although the *Disseminate phase* features the key element of dissemination from a geospatial statistics perspective, there is a need to recognise geospatial data output as something fundamentally different from regular statistics. In general, there is a vague understanding of the differences between "maps" and "geospatial information" within the statistical community. Maps (actually recognised as a product component by the process chart in Statistics Sweden but not by the GSBPM) should be understood as a representation of statistics, no different from graphs or tables, whereas geospatial data is a more complex form of output. In fact, some geospatial data outputs are not statistics at all. Statistics Sweden, like many other NSIs, produces geographical delimitations of localities and other geographical-statistical units. These are disseminated as products along with statistical information describing the properties of the urban areas, but the boundaries themselves are also used as spatial input data by Statistics Sweden and other users for further analysis.

Geospatial statistics output products can to some extent be considered as intermediates. They represent something in between micro data and a final statistical product. There are also resemblances between registers and geospatial statistics products.

Phase 8. Evaluate

One of the most important issues that were discussed with relevance for the *Evaluate phase*, is how to assess quality from a geographical point of view. Quality assessments are conducted on regular basis to ensure a high quality in the core registers. Typically issues related to consistency in geocodes etc are not part of the quality assessments. There are few or no recommendations on how to proper assess quality from this angle. The units responsible for maintaining the registers typically have a limited understanding of the impact of poor geocodes for geospatial statistics products. Quality control mechanisms based on geospatial information is not implemented as part of the maintenance schemes, hence preventing geocoding errors to be effectively assessed and detected.

General remarks

A general conclusion is that the structure of the GSBPM is sufficient to structure the work on geospatial data management. After all, in most processes, geospatial information is not fundamentally different from other information processed in NSIs. No urgent need to modify the model in terms of introducing new sub-processes was concluded.

Semantics is a crucial factor to take into consideration in order to use the GSBPM (or national implementation of process charts) or to streamline management of geospatial information into core statistical production. Unless people can relate to the vocabulary of the process model, it will have little or no impact as a tool to improve and streamline the geospatial statistics production.

The semantic discrepancy between the statistical and geospatial terminology occurring in the GSBPM may pose obstacles to a successful use of the model to structure management of geospatial information.

3 Implementing the GSBPM at Statistics Austria

3.1 Current situation

The GSBPM has certainly arrived at Statistics Austria, primarily in the "Quality Management and Methods"-section. They are considering and working on implementing the phases of the GSBPM into the processes of the time recording system and project planning. Furthermore some departments recently received the task to check how their working steps fit into the GSBPM processes.

Concerning the GSBPM and its focus on and inclusion of the geospatial aspect, after a short discussion with the "Quality Management and Methods"-section, their opinion was that it would be enough to include the geospatial aspect of the data in the GSBPM descriptions of the relevant phases. So they did not see the need to redefine the model itself. This may be due to the fact that the storage of the building ID with the unit record data meanwhile is taken for granted and through the geo-coordinate behind the building-Id there are no obstacles to the whole world of geospatial analysis.

3.2 The scope of the assessment

It was decided to undertake the assessment of the GSBPM in relation till geospatial information management in the form of a workshop with participation of geospatial experts and people working for the register-based census. The workshop was conducted as a simple story mapping exercise.

It turned out that the register census staff had a very structured overview and time schedule for each working step, describing when what is done and who is responsible. It also included the dependencies between the working steps (e.g. what has to happen before). Interestingly each of them knew a lot about his/her own working step, but neither of them had the full picture.

3.3 Statistics Austria's remarks

As for the exercise of assessing the GSBPM – with the knowledge of the main processes of GSBPM and their overview of working steps, the main outcomes of the workshop were:

1) Previous censuses focused more on the output areas. Therefore the understanding of the place for the geospatial aspect of data in the GSBPM was seen with the latter processes starting from *collect*, *process* and *analyse*.
2) The building-Id plays an essential role in the register based census. Some of the working steps only were possible because the building-Id was there, which originally was not planned but they were just lucky that it was there when they needed it. Examples for this are designing household and family units.
3) With this realisation it soon came clear that it is advantageous to include the building-Id right from the beginning, starting with the *Specify Needs phase*. Apart from the possibility to react more easily to administrative changes it contributes to the improvement of data quality and has numerous advantages in all phases such as the possibilities to build new variables (such as households and family units as mentioned above), creating new information (commuting distance from door to door) or the dissemination of new data products (such as small area statistics / grid statistics).

4 Implementing the GSBPM at the Statistical Office of France (INSEE)

4.1 Current situation

Insee has already launched initiatives regarding geospatial issues, such as setting up a strategic statistical committee for the management of geography, drawing up a map of the various geographical activities within the French statistical system, collaborating with the French NMCA to build an authoritative address register. In this a context, the GSBPM is seen as a powerful tool that could help to implement the vision.

4.2 The scope of the assessment

Just like in Sweden and Austria, it was decided to undertake the assessment of the GSBPM in relation till geospatial information management in the form of a small workshop with participation of geospatial experts and experts from the Quality Unit, Census Department, Methodology unit, IT department, Regional statistics unit and Business Register Unit.

4.3 Insee's remarks

The main conclusions were the following:

1) Insee stands at a turning point between an internal and scattered management of spatial issues and a coordinated outsourcing to IGN of these issues. Before implementing the GSBPM model, it may be fruitful to first implement the UN-GGIM Global Statistical-Geospatial Framework to the current and targeted French situations. As a consequence, Insee endorses the GEOSTAT 3 project that aims at building an European version of the global framework.
2) More precisely, regarding the GSBPM exercise:
    a. All the attendees agree on the future usefulness of the GSBPM, and on the need of several meetings to have a more precise view of its implementation at Insee.
    b. The adaptation of the GSBPM model to geospatial issues should include specific processes or sub-processes dedicated to the integration of geospatial data in statistics itself. In France, there is no unique identifier of the addresses. The integration is carried out with algorithm comparing the denominations in the address register and those in the administrative files. When the score is not high enough integration is done by a devoted team, and then by imputation.
    c. A process targeted at maintaining, updating, ensuring the consistency of authoritative geospatial data and statistics over time should be added to the GSBPM.
    d. Managing confidentiality is one of the big issues to manage geospatial statistics. It might need a specific process as well.
    e. Depending on the version of the GSBPM, the definition of some terms are not included. In the version used by the workshop, we could not get the exact meaning of geospatial data, geospatial statistics and spent maybe too much time talking about it. A specific attention should be paid on terminology.

5 Implementing the GSBPM at Statistics Portugal

5.1 Current situation

According to Statistics Portugal mission, quality commitment is a milestone for statistical production and dissemination processes, duly aligned with the European Statistics Code of Practice and the General Guidelines of Official Statistical Activity 2013-2017.

The Statistical Production Processes Handbook (last version dated 2010) identifies and systematically documents all the phases, processes and sub-processes that are encompassed in every statistical survey/statistical activity. It also enables identifying the connections and linkages between related activities, which allow a better and more effective functioning of the organization at process level.

Taking into account the core activity of Statistics Portugal this document constitutes a reference for Statistics Portugal staff for those in particular who are directly involved in performing specific tasks of the process of production and dissemination of a survey/statistical activity.

The process Model is organized into 4 phases deployed into processes and sub-processes and it is organized as follows:

- Identification and description of the phases, processes and sub-processes of a survey/statistical activity;
- Identification of the documentation to be produced within each process;
- Identification of reference documents to be used in each process;
- The production process matrix – a summary of the production process at the level of tasks, identifying the responsibilities of each unit involved.

The phases identifying the major stages of the statistical production and the process/sub-process are:

- Conception: includes three processes related with the first planning stage of the statistical operation, such as technical dialogue between the various stakeholders and the feasibility and the methodological studies.
- Development: Includes the processes related with the development of statistical methods and respective infrastructure, as well the data collection process and data treatment and analysis processes. The security of statistical information process is also included in this phase.
- Dissemination: Includes the tasks concerning the main objective of the statistical production process, i.e. to disseminate to society quality official statistical information according to international best practices. SP Portal is the main channel for the dissemination of statistical data. It releases at first-hand all official statistics, including those produced under ONA – i.e., those entities with whom SP signed a Protocol for delegation of statistical functions - and which are simultaneously released on their websites.
- Assessment: includes horizontal tasks related with quality assessment during the statistical production process. The result of the quality assessment aims at identifying strengths and weaknesses along the process and improvement actions for the quality of the statistical production, for further corrections.

The following scheme presents the business model at Phase, Process and Sub-Processes level. This version does not reflect in an explicit way the integration of geospatial component that exist at the moment. The current work and effort to update the statistical production process, will allow us to describe in an unequivocal terms this relationship.

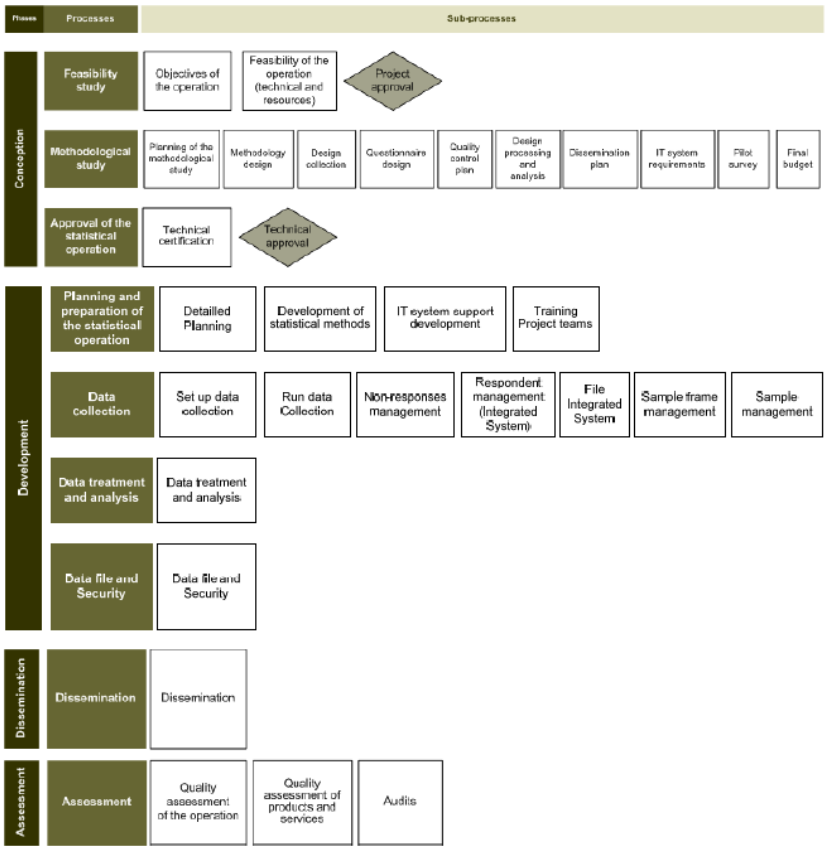

Figure 6: The Process Chart

5.2 Spatial Data Infrastructure

Nowadays, Statistics Portugal's integrated system is designed to allow for: (1) an integration of the statistical data warehouse systems and the spatial data, including both statistical units that support statistical projects and the microdata obtained from statistical projects; (2) the provision of the appropriate tools for statistical dissemination and to support statistical production: the tools (and spatial data) to run spatial analysis, to design samples and to collect data

5.3 SDI and the Statistics Portugal GSBPM chart

Following the Peer review 2015 recommendation 10, statistics Portugal is currently redefining its statistical production process model, following the guidelines and overall structure of the GSBPM.

This is an opportunity to reflect, in its processes and sub-processes, the progressive integration of the geographic information of statistics Portugal SDI at the different phases of the statistical production process.

This integration was boosted by the use, in 2012, of the 1Km^2 Grid as the reference geography, alongside with the building points of the BGE (that are the geographical expression of the FNA

households and business units), to support the sampling design and data collection processes. Most recently Statistics Portugal has developed the GEOINQ platform to support and manage, in an integrated system, the data collection for all surveys that Statistics Portugal develops. The GEOINQ platform incorporates several of the SDI components that play a key role in the overall system.

Despite the early stage of the Statistics Portugal GSBPM chart review, several tasks within different subprocesses and subprocesses within different phases have already identified that also suit geographical information. In fact, the geospatial data and the SDI is present in most of the phases of production process in the context of a survey.

Below is a first draft of the reviewed GSBPM chart and further discussion and clarification will be needed until we can fully assign the geographical information, of the several components of the SDI, in the processes and sub-processes of the different phases of Statistics Portugal GSBPM.

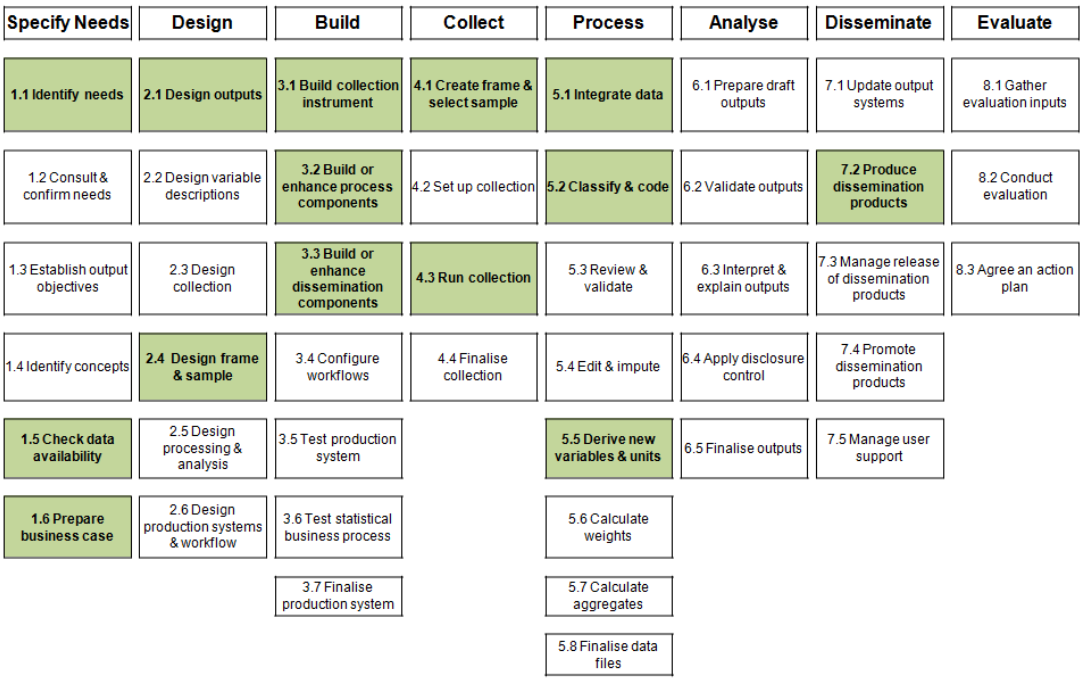

Figure 7: First draft of the reviewed GSBPM chart recognising geospatial information

6 Implementing the GSBPM at Statistics Poland

6.1 Methodology of comparative analysis of spatial data handling in the CSO with the phases of the GSBPM

In order to make a precise analysis of the processes implemented in the CSO in terms of spatial data compliance with the processes recommended by the generic model, the methodology of mapping the equivalent business processes was adopted, taking as a reference the GSBPM model. This model represents a generic reference standard of statistical business process.

The comparative analysis method was based on subsequent comparison of business functions in each GSBPM phase and sub-process with actual processes and business functions implemented in CSO statistical production with particular emphasis on spatial data.

Compliance analysis was made by mapping various activities carried out in CSO to phases, processes and sub-processes recommended in GSBPM. The results of analysis showed that all actions undertaken in CSO are fully compatible with GSBPM sub processes for phase 1 - Specify needs, phase 3 – Build, phase 5 - Process and phase 8 - Evaluate. In those phases there are no additional activities that are not covered by the GSBPM sub processes. In the other phases the situation is as follows:

Phase 2. Design

The CSO performs more activities than GSBPM sub-processes provides in this phase. This applies in particular to all spatial aspects of statistical design of the production process. This phase of GSBPM model shows shortcomings in the following aspects:

- designing spatial processes,
- designing geocoding methodology for population sampling frame,
- sampling,
- spatial analysis,
- geostatistical data processing.

It can be stated, that the spatial aspect is completely ignored in this GSBPM model phase and deserves to be supplemented with sub-process 2.5a “Design geocoding frame, sample & data collection", which should be located between sub processes 2.5 and 2.6 of Phase 2. Sub process 2.5a would ensure designing a proper relationship between spatial data and statistical data as well as necessary spatialisation of the sampling frames, data and analytical products through well thought and planned geocoding, so that GIS tools can be used at all stages of massive statistical surveys processing.

Phase 4. Collect

The CSO performs more activities than GSBPM sub-processes provides in this phase. This phase of GSBPM model shows shortcomings in the field of spatial geocoding of the population sampling frame, sample and the data collected through electronic channels (multimode data collection). It can be stated that the spatial aspect is ignored in this GSBPM model phase and deserves to be supplemented with sub-process "4.1a Geocode frame & sample", which should be located between 4.1 and 4.2 subprocesses of Phase 4.

Aside from geocoding of the sampling frame and sample, geocoding collected variables at an early stage of their acquisition is an important issue requiring separate treatment. Data geocoding will ensure from the beginning appropriate relation between spatial data and statistical data, necessary data spatialisation, so that GIS tools can be used at all stages of massive statistical surveys processing. Spatial reference of statistical data, apart from its analytical advantages, can serve as a strong individual data connector in the data integration process, provided, among others, in the next phase - the sub process 5.1. Therefore, we propose the creation of new subsequent sub process 4.3a “Geocode collection” located between 4.3 and 4.4 sub-processes of Phase 4.

Phase 6. Analyse

The CSO performs more activities than GSBPM sub processes provides in this phase. This phase of GSBPM model shows shortcomings in the following aspects:

- spatial analysis implementation methods,
- creation of analytical structures, tables and spatial cubes as well as cartographic presentation methods and their validation.

Spatial data requires separate treatment because of its specificity, the need for new data models, methodologies and methods of geocoding and analysis. We propose to improve the model by modifying Phase 6 and adding sub-process 6.2a “Prepare spatial analyzes & maps" which should be located between sub processes 6.2 and 6.3.

Phase 7. Disseminate

The CSO performs more activities than GSBPM sub processes provides in this phase. This phase of GSBPM model shows shortcomings in the field of publication of specific spatial analysis results and maps in the form of choropleth maps and diagram maps. These aspects require extraordinary procedures associated with the use of GIS tools and data interpretation. Publication of spatial data requires separate treatment due to its specificity, the need for new data models usage, methodologies, and joining methods and data analysis. We propose to improve the model by modifying the Phase 7 and adding sub process 7.2a “Manage spatial analyzes & maps using GIS" which should be located between sub-processes 7.2 and 7.3.

6.2 Remarks by CSO

To summarize conducted comparative analysis of the spatially referenced statistical data production processes implemented in the CSO versus the GSBPM model in terms of spatial data, we can draw the following conclusions:

1) The CSO performs all activities provided in the GSBPM model.
2) Model implemented and tested in practice in the last census round appears to be more comprehensive than the theoretical generic GSBPM model.
3) Through the practical implementation of relevant processes relating to spatial data and mapping them with GSBPM model the important areas not included in the model were pointed out, although they are important and necessary in the actual production process.
4) By indicating these areas, the potential shortcomings and imperfections of GSBPM model were diagnosed.
5) Shortages of GSBPM model essentially concerned statistical data spatialisation aspects from the stage of designing the data collection, geocoding, analysis and providing spatial characteristics of statistical products.
6) Diagnosed shortcomings are so important for the ongoing work of UN-GGIM, UNECE and Eurostat, in the field of issues involved in combining spatial data with statistical data, that there should be a significant modification of the GSBPM model in terms of geospatial aspects.

Therefore, we suggest that the current GSBPM model v 5.0 should be modified by adding the following sub-processes:

- Phase 2: sub process "2.5a Design geocoding frame, sample & data collection";
- Phase 4: sub processes "4.1a Geocode frame & sample" and "4.3a Geocode collection";
- Phase 6: sub process "6.2a Prepare spatial analyzes & maps";
- Phase 7: sub process “7.2a Manage spatial analyzes & maps using GIS"

The introduction of above sub-processes would enrich the model with a spatial component of the statistical production process, which will allow a better understanding of spatial data as well as its role and place in the statistical production process and the standardization of methodologies merging statistical data with spatial data. This issue is still poorly perceived by the statistical community and is still in the early development phase. For obvious reasons spatial elements could not be fully taken into account when constructing the model. The introduction of the proposed changes in the GSBPM model would contribute to strengthening the position of geospatial statistics in the statistical production process and could be an important achievement of the GEOSTAT 2 project.

Finally, the updated version of the graphic GSBPM model could look as follows:

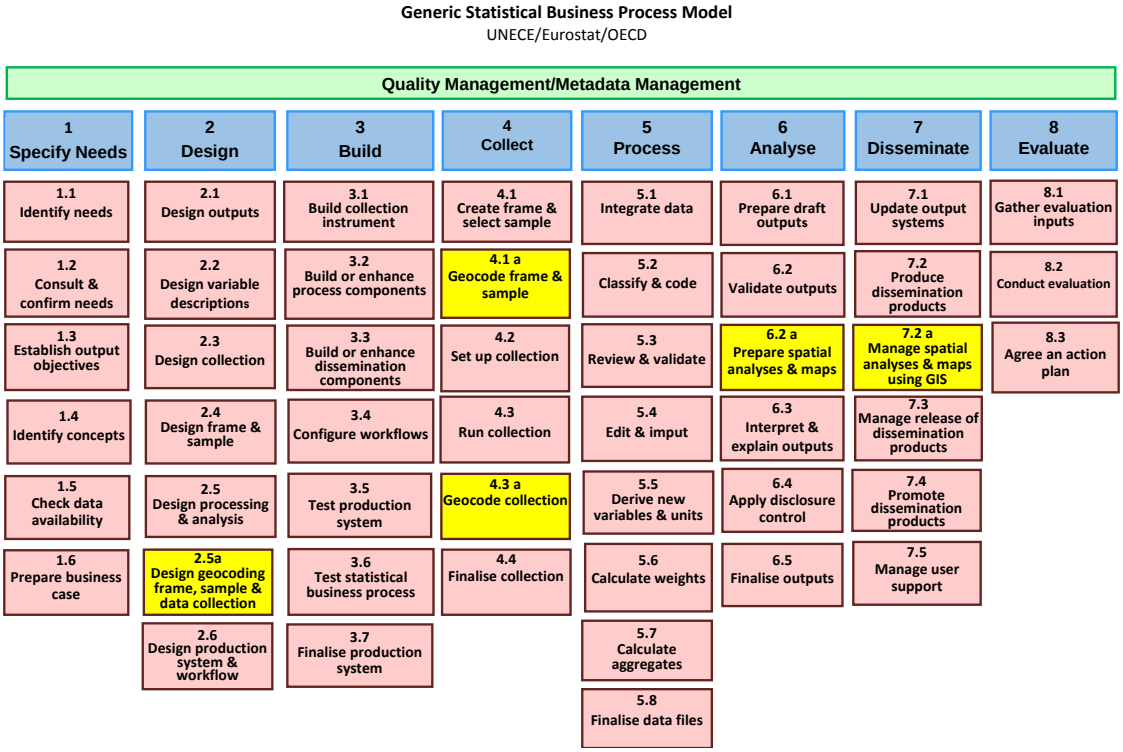

Figure 8: Graphical representation of the revised GSBPM model as proposed by CSO, including 4 new sub processes.