Statistics Finland

Geostat 2: Some experiences and conclusions – applying the GSBPM to geospatial statistical data production

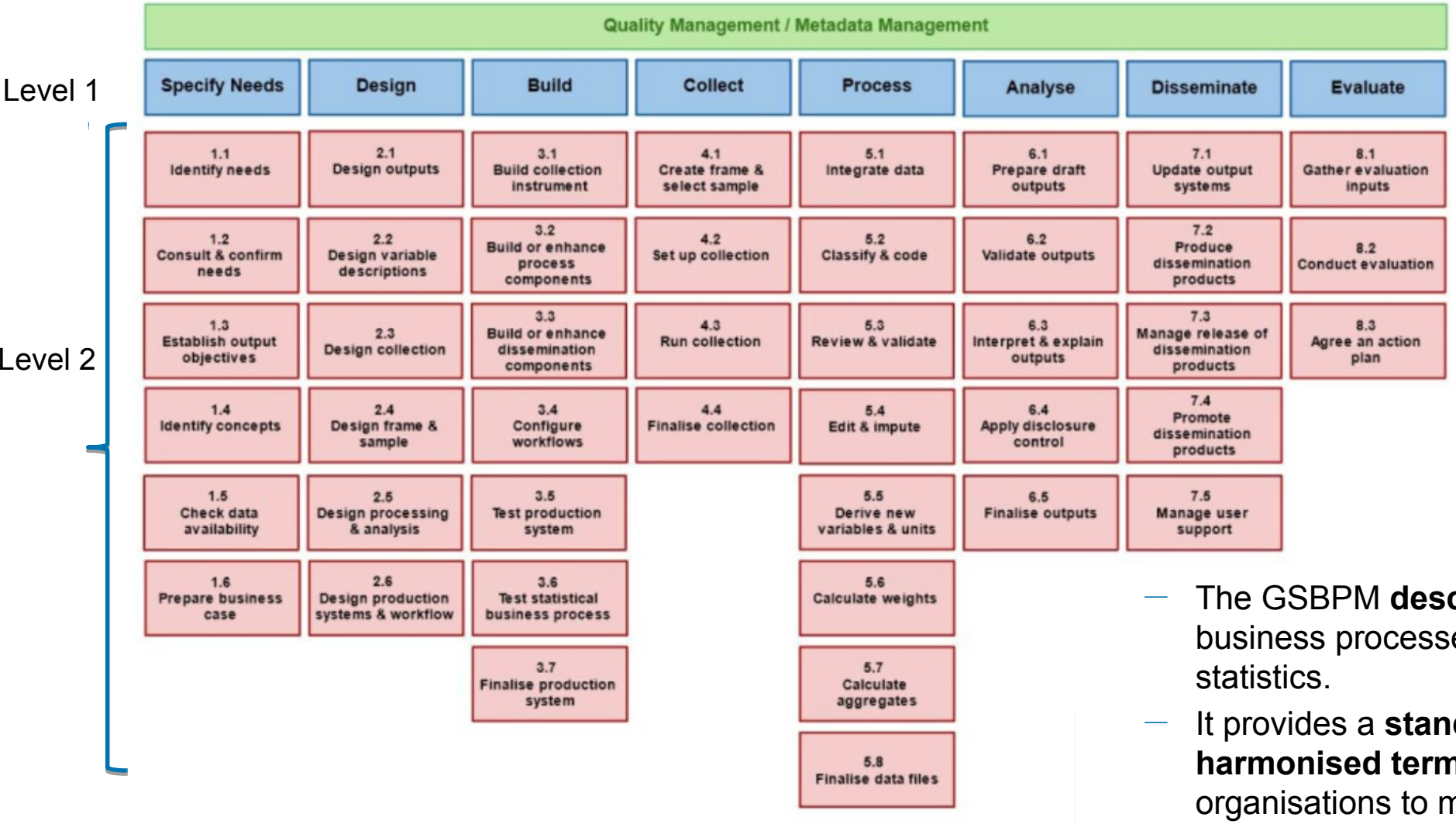
Rina Tammisto, Marja Tammilehto-Luode, Tuuli Pihlajamaa

European Forum for Geography and Statistics Conference 2016, Paris
15 November 2016

Geostat 2, Work Package 1: Geospatial data and the GSBPM

- Evaluate ***The Generic Statistical Business Process Model (GSBPM)*** from the point of view of *geospatial data* is involved in the statistical production process
  - Is the GSBPM usable?
  - What using requires?
  - Common view of the issues that should be developed
    - In the GSBPM model itself or
    - In its documentation (concrete suggestions if possible)
- In parallel with the project the *Global Statistical Geospatial Framework (GSGF)* was prepared – a wider perspective where to locate our work

GSBPM, The Generic Statistical Business Process Model



- The GSBPM **describes and defines** the set of business processes needed to produce official statistics.
- It provides a **standard framework** and **harmonised terminology** to help statistical organisations to modernise their statistical production processes, as well as to **share** methods and components.

The GSBPM documentation

http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model

2.1. Design outputs

49. This sub-process contains the detailed design of the statistical outputs, pro
to be produced, including the related development work and preparation of the sy
used in the "Disseminate" phase. Disclosure control methods, as well as processes
to any confidential outputs are also designed here. Outputs should be designed to
standards wherever possible, so inputs to this process may include metadata from
collections, international standards, and information about practices in other statis
from sub-process 1.1 (Identify needs).

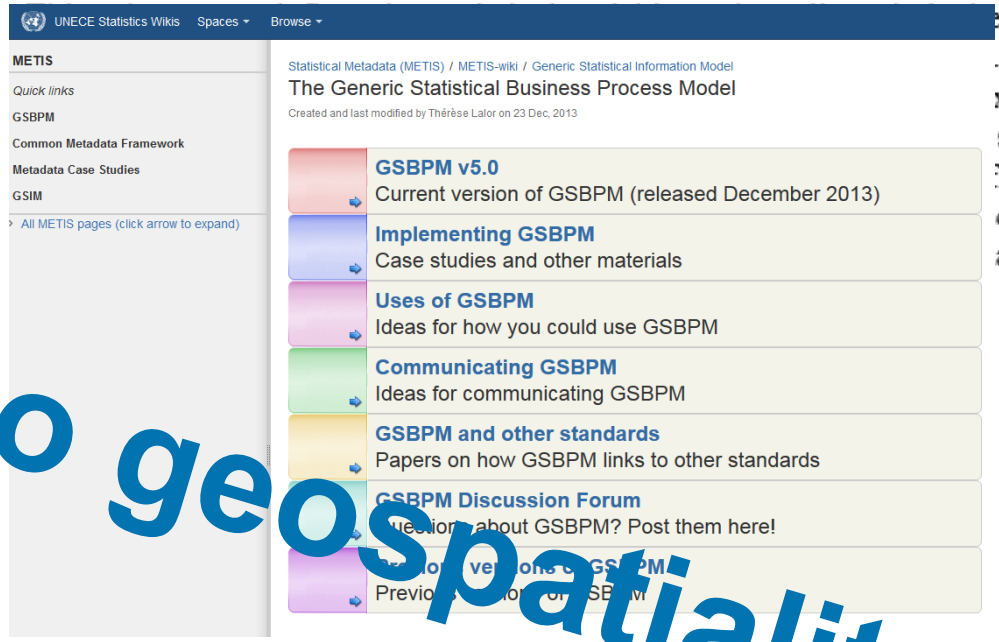
2.2. Design variable descriptions

5.1. Integrat data

75. This sub rces integrates data from one or more sources. It is where the results of sub-processes in the "Collect" phase are combined. The input data can be from a mixture of external or internal data sources, and a variety of collection modes, including extracts of administrative data. The result is a set of linked data. Data integration can include:

- combining data from multiple sources, as part of the creation of integrated statistics such as national accounts
- matching / record linkage routines, with the aim of linking micro or macro data from different sources
- prioritising, when two or more sources contain data for the same variable, with potentially different values

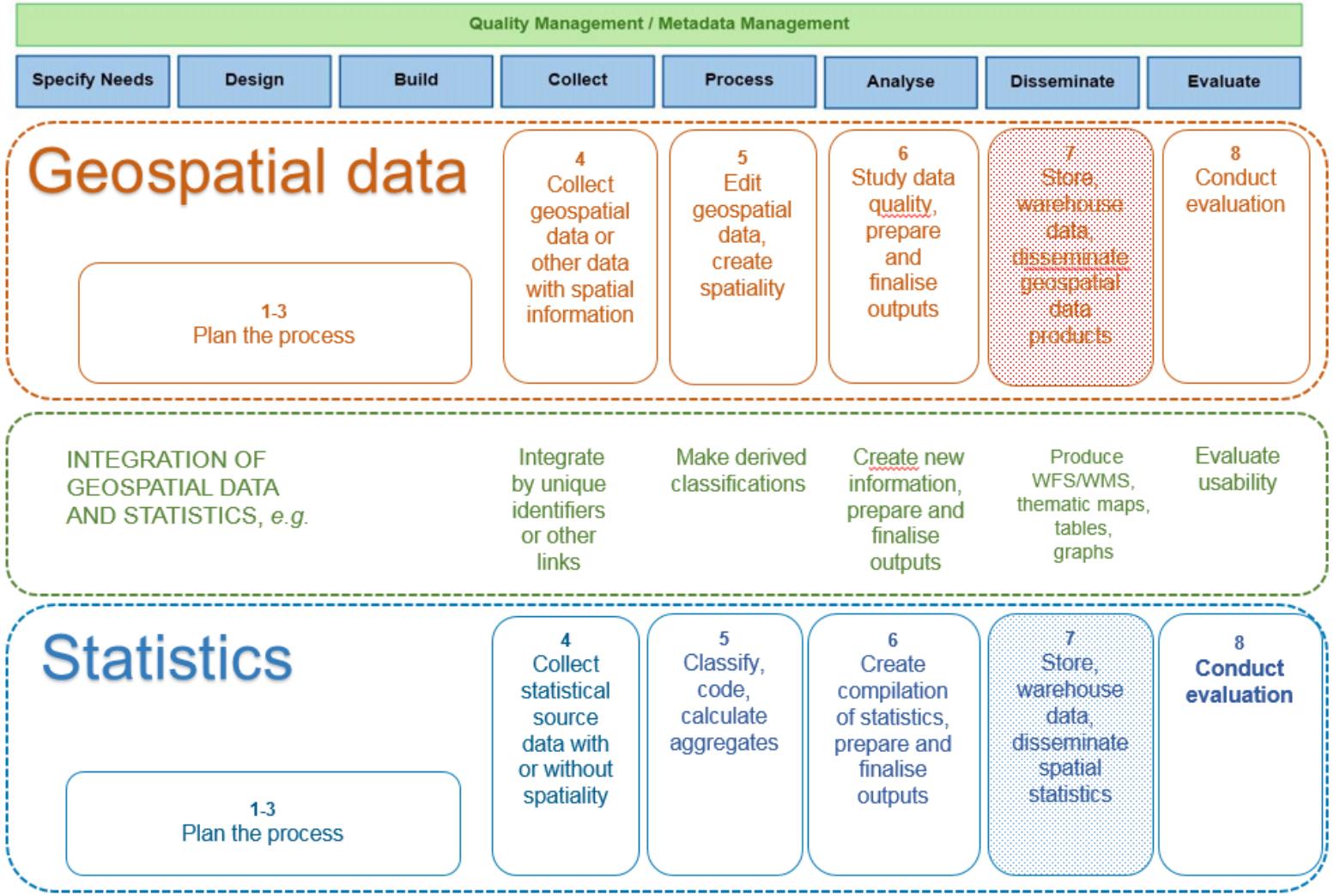
76. Data integration may take place at any point in this phase, before or after any of the other sub-processes. There may also be several instances of data integration in any statistical business



Not any reference to geospatiality

Statistics Finland

Integration of Geospatial data and Statistics



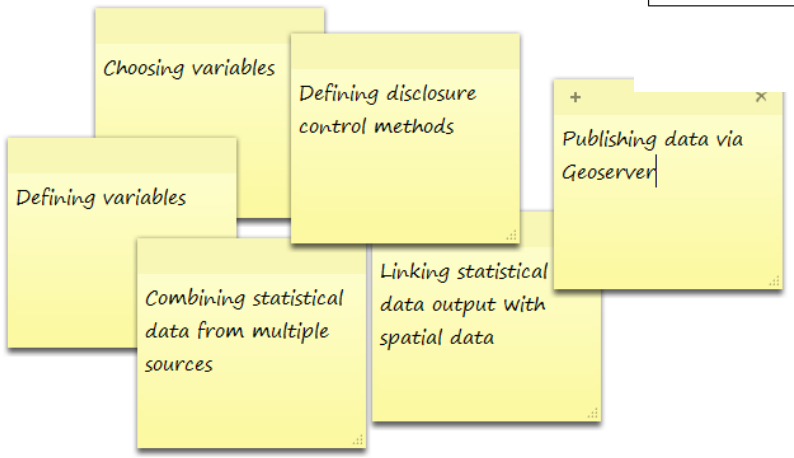
Modelling of processes

Story Mapping method:

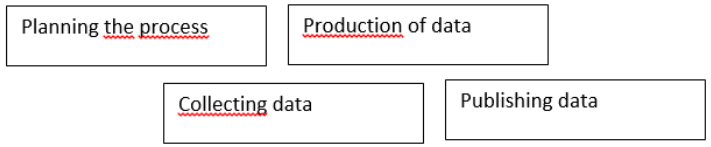
http://www.slideshare.net/SteveRogalsky/user-story-mapping-in-practice

1. Writing down free cognitive associations (tasks) of the process
2. Classification of notes and identification of phases + completion by phases still missing
3. Identification of GSBPM phases + completion by phases still missing
4. Creation of the model + recognising actors

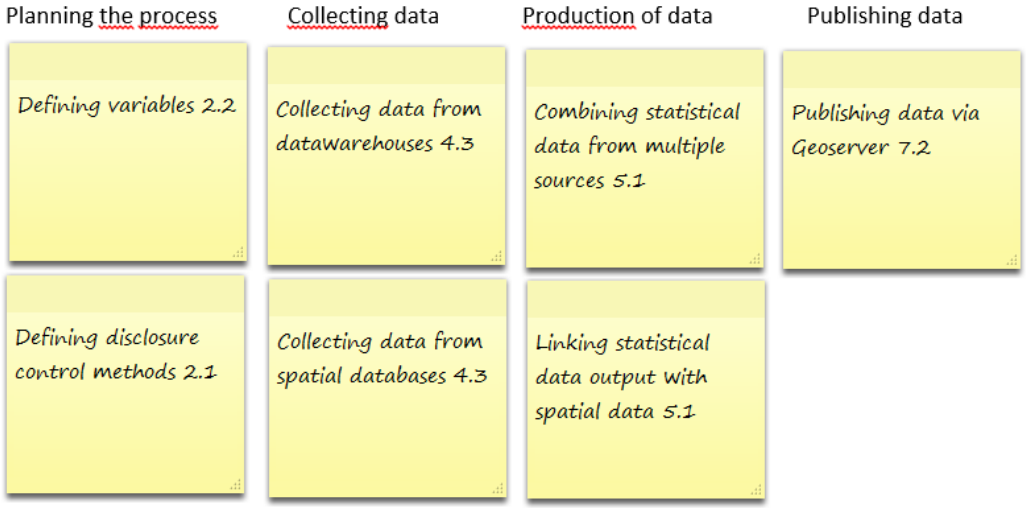
TASKS: CLASSIFICATION



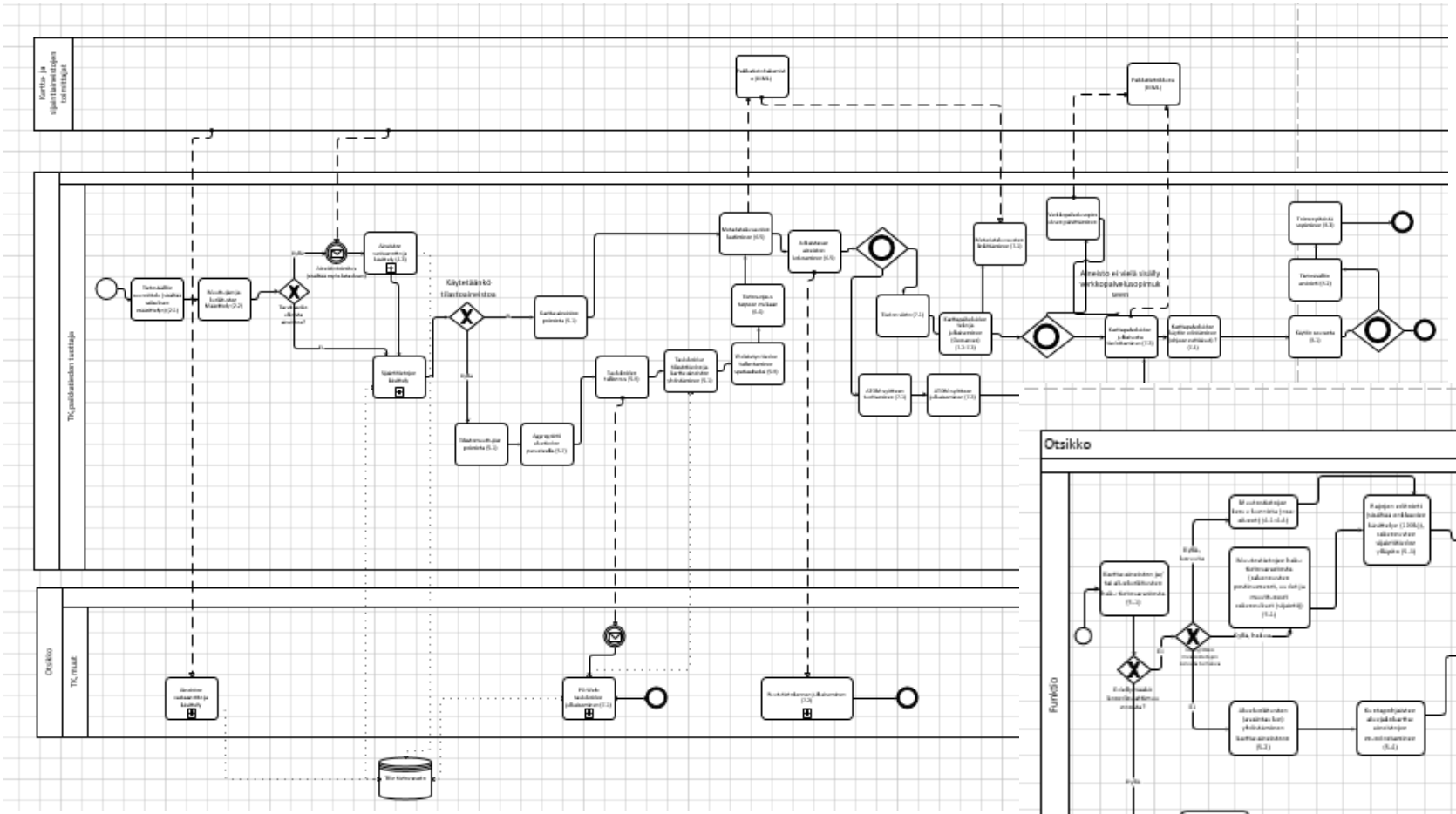
IDENTIFICATION OF PHASES



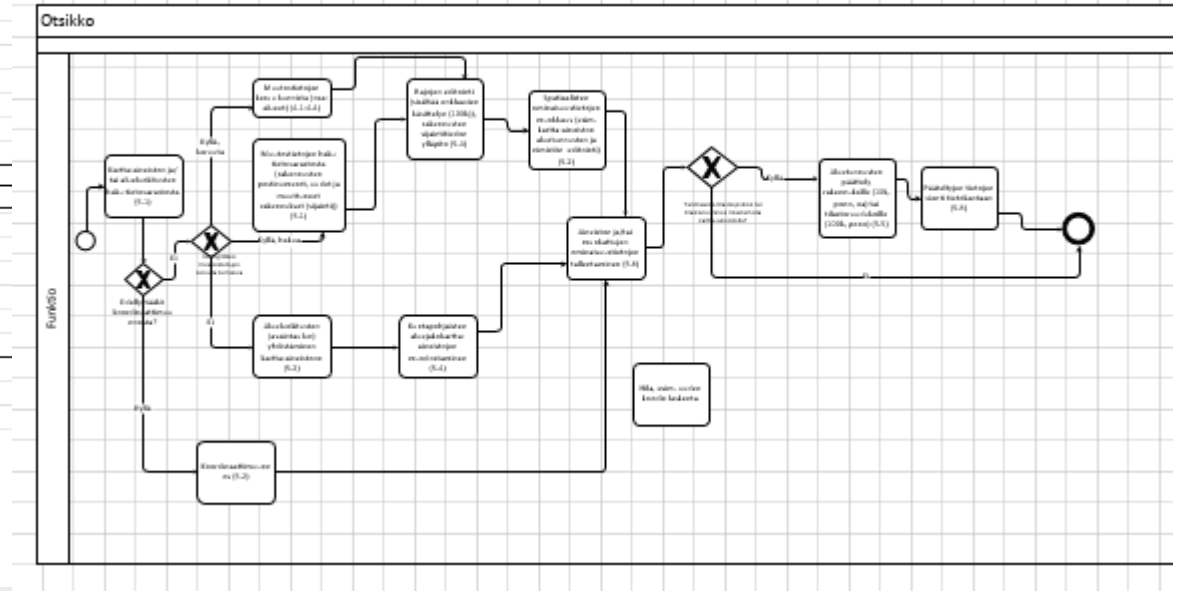
IDENTIFICATION OF TASKS IN GSBPM



8th version ☺
Main statistical geospatial related production process

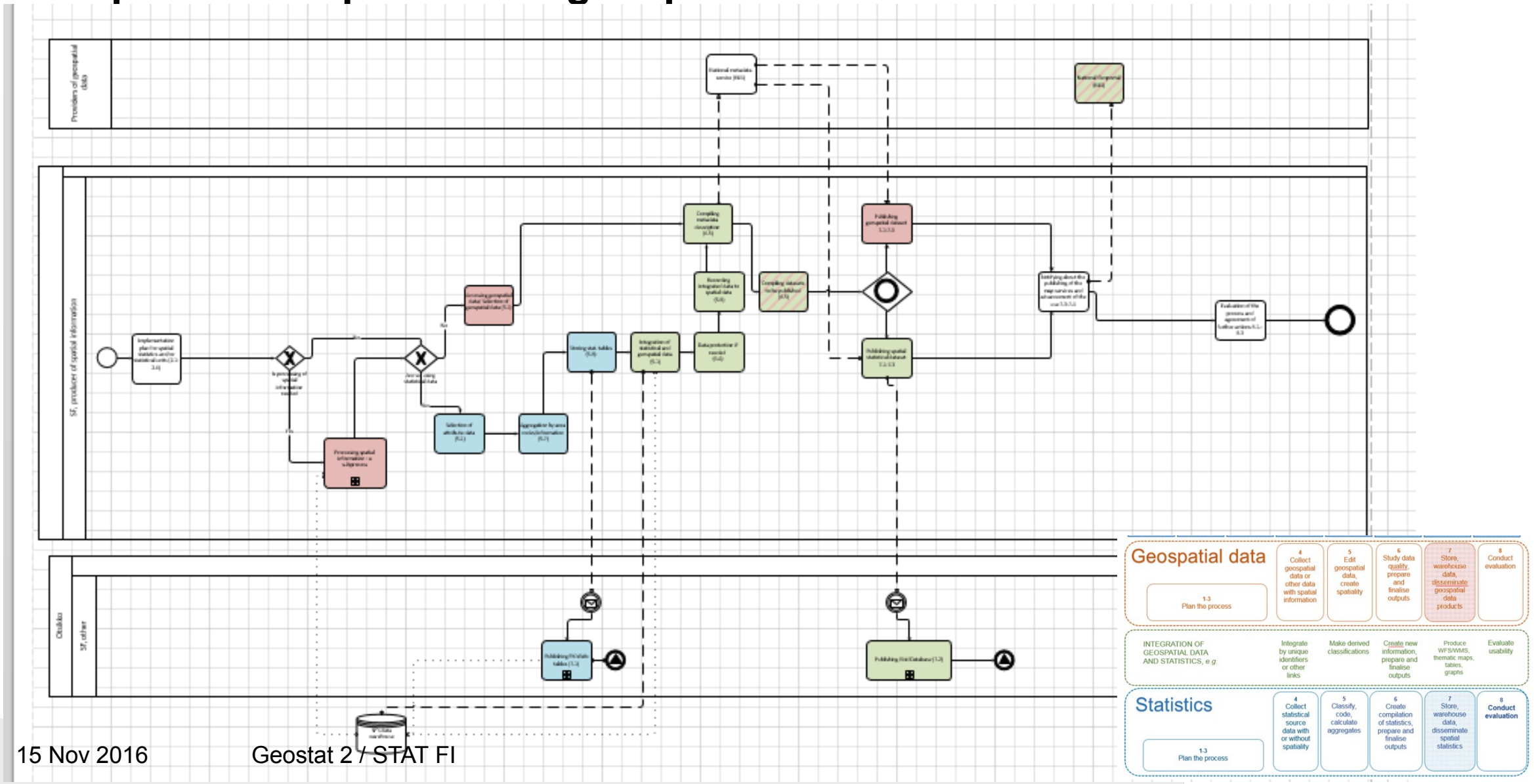


Sub-processes to maintain points and polygons (related to the main process)

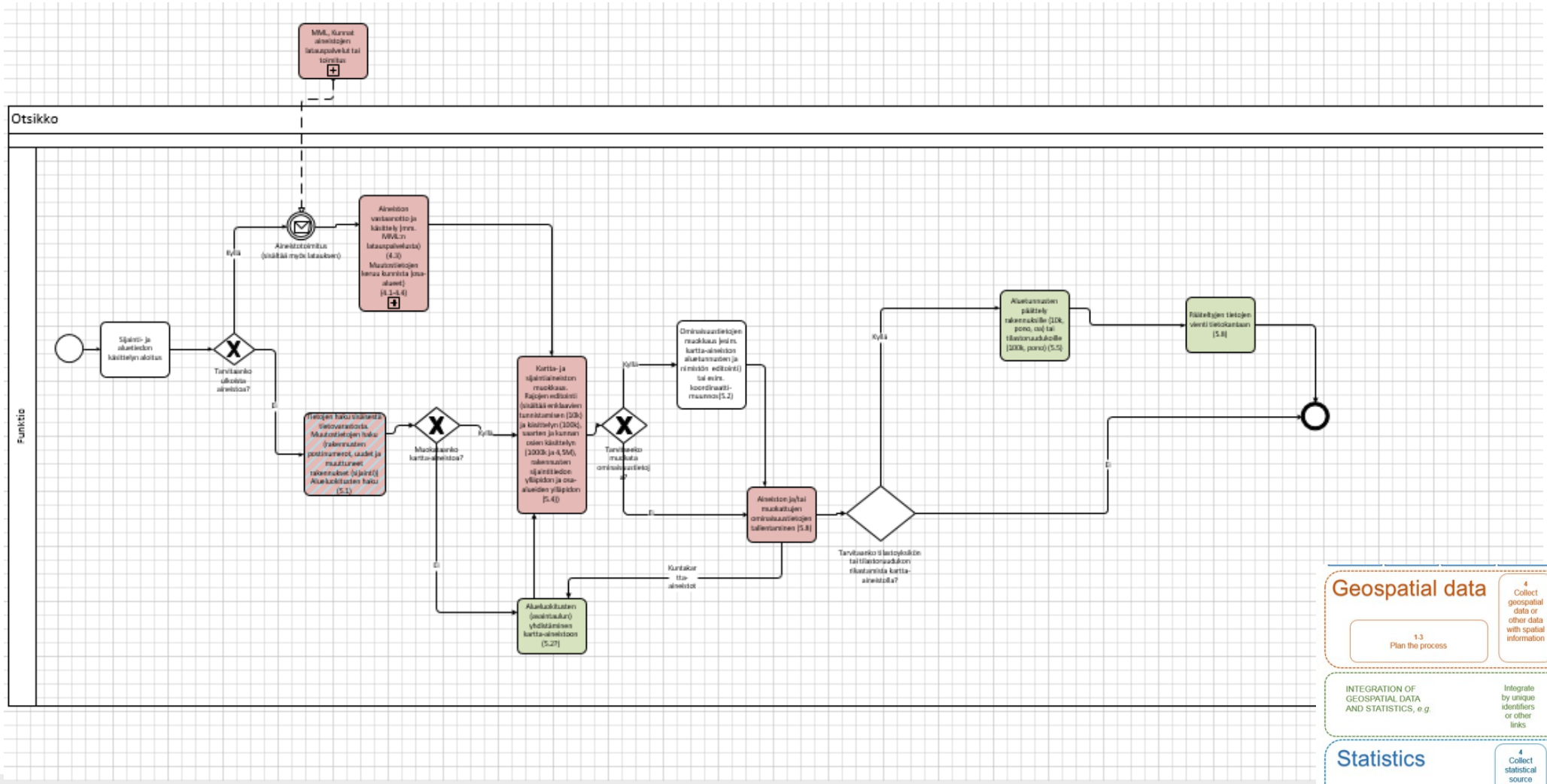


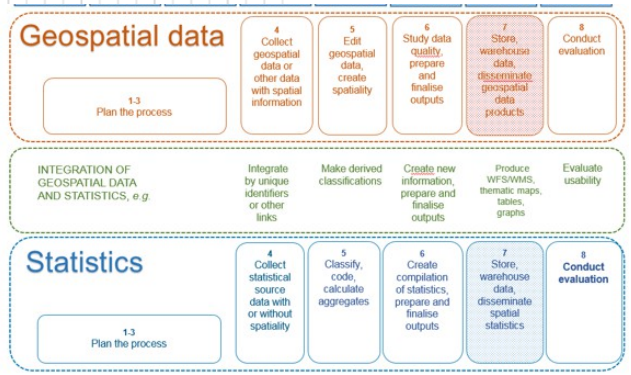
Statistics Finland

11th version ☺
Main production process of geospatial statistics with dimensions



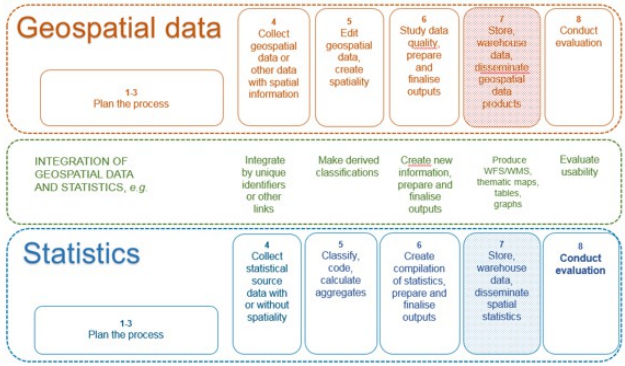
Sub-processes to maintain points and polygons (related to the main process) with dimensions



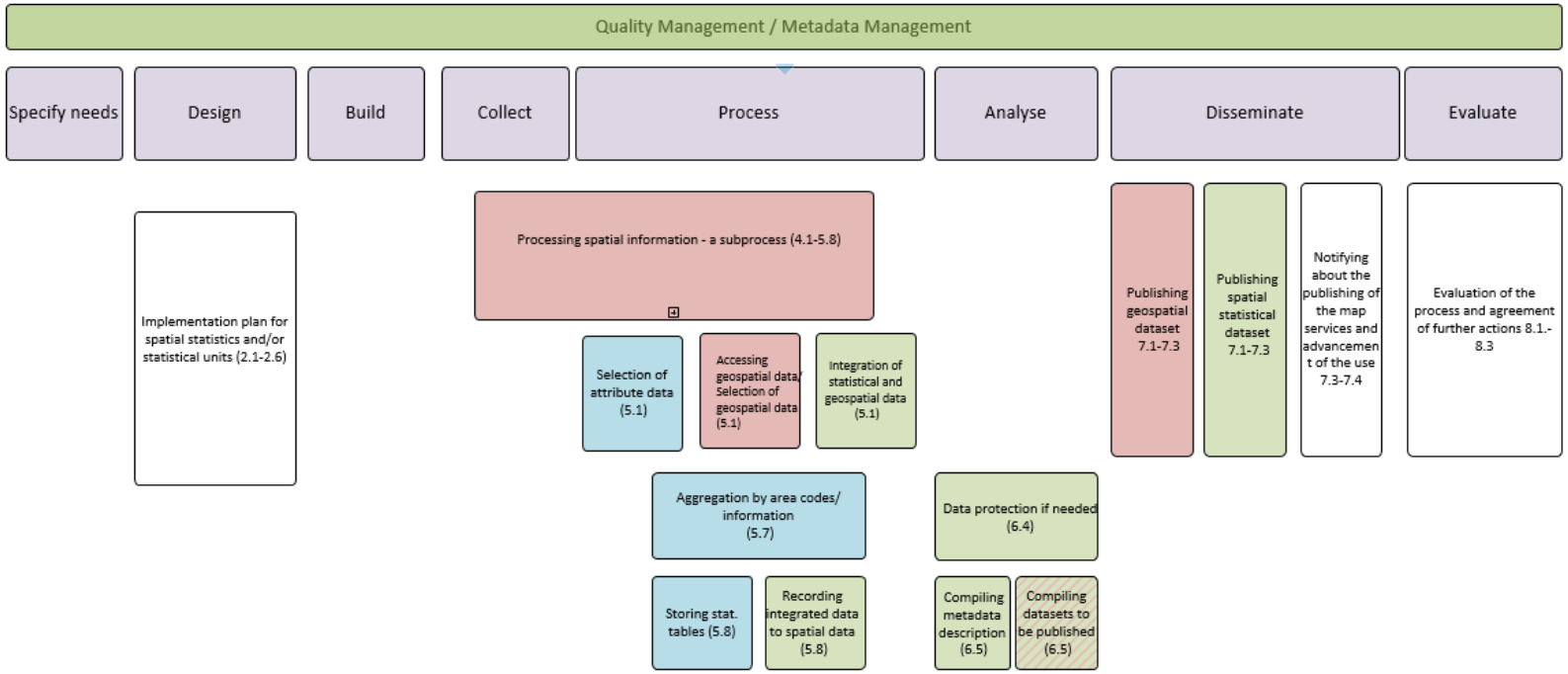


A list of sub-processes and decission points

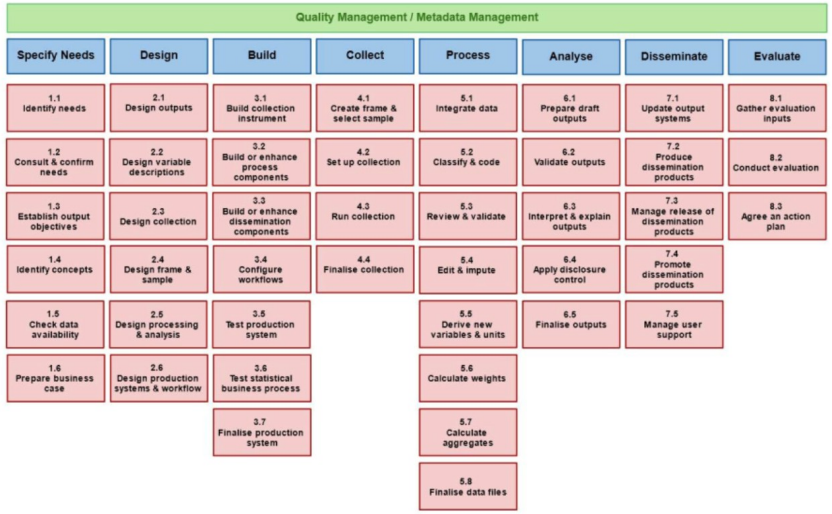
			GSBPM	INPUT	TASKS	OUTPUT
	Implementation plan for spatial statistics and/or statistical units		(2.1-2.6)			
X	Is processing of spatial information needed					
		Processing spatial information - a subprocess	(4.1-5.8)	Map data (Point and polygon features)		
X	Are we using statistical data					
		No				
		Accessing geospatial data/ Selection of geospatial data	(5.1)	Selection of geospatial data from a database, a disc or a file geodatabase	Selection of geospatial data to be published	Geospatial datasets with topography on a disc / geospatial datasets by gis-formats on a disc
		Yes				
		Selection of attribute data	(5.1)	Point-based/Unit level attribute data	Formalizing access to data warehouse	Access to database and datasets available
		Aggregation by area codes/information	(5.7)	Access to database and datasets available	Statistical information processed and classifyed according to definitions	Statistical datasets aggregated and compiled to database tables
		Storing stat. tables	(5.8)	Statistical data aggregated to database tables	Storing statistical tables	Aggregated statistical tables extracted from unit level data
		Integration of statistical and geospatial data	(5.1)	Aggregated statistical tables extracted from unit level data *and Geospatial datasets with topography on a disc / geospatial datasets by gis-formats on a disc*	Integration of statistical and geospatial data	Spatial statistical datase
		Data protection if needed	(6.4)	Spatial statistical dataset	Recording/modifying dataset according to confidentiality rules	Spatial statistical datase
		Recording integrated data to spatial data	(5.8)	Spatial statistical dataset	Recording/modifying dataset according to confidentiality rules	Spatial dataset available on disc
	Compiling metadata description		(6.5)	Information on geospatial and statistical data content	Compiling metadata description	Joint metadata descripti
	Compiling datasets to be published		(6.5)	Spatial statistical or spatial dataset available on disc	Compiling spatial statiscal or spatial dataset as predefined	Dataset ready for publishing



The sub-processes of production of geospatial statistics placed in the GSBPM



… a way to communicate



Statistics Finland

Stat Fi findings

- The GSBPM seems to be able to cover geospatial related statistical production – however,
- Use of descriptions of the GSBPM requires raising the level of abstraction when interpreting phases - a risk from the coherence point of view

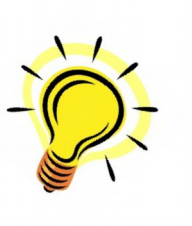
→Our opinion is that the GSBPM model does *not* need
new sub-processes for covering geospatiality
but the documentation needs to have a broader scope

A suggestion of further work after the Finnish contribution

1. **Study the national processes of spatial statistics, Finland** + *other countries*
2. Recognise mutual phases
3. Identify and agree mutual phases against the GSBPM model
4. First ideas of items that should be included in the GSBPM descriptions
5. Evaluate draft additional texts with other interested parts
6. Study the potentials of other descriptive methods (e.g. GSIM)

How to run the same kind work at other NSOs

- A workshop…
- ”10 steps” instructions …
- Several telemeetings...



→***What the difficulties tell us?***

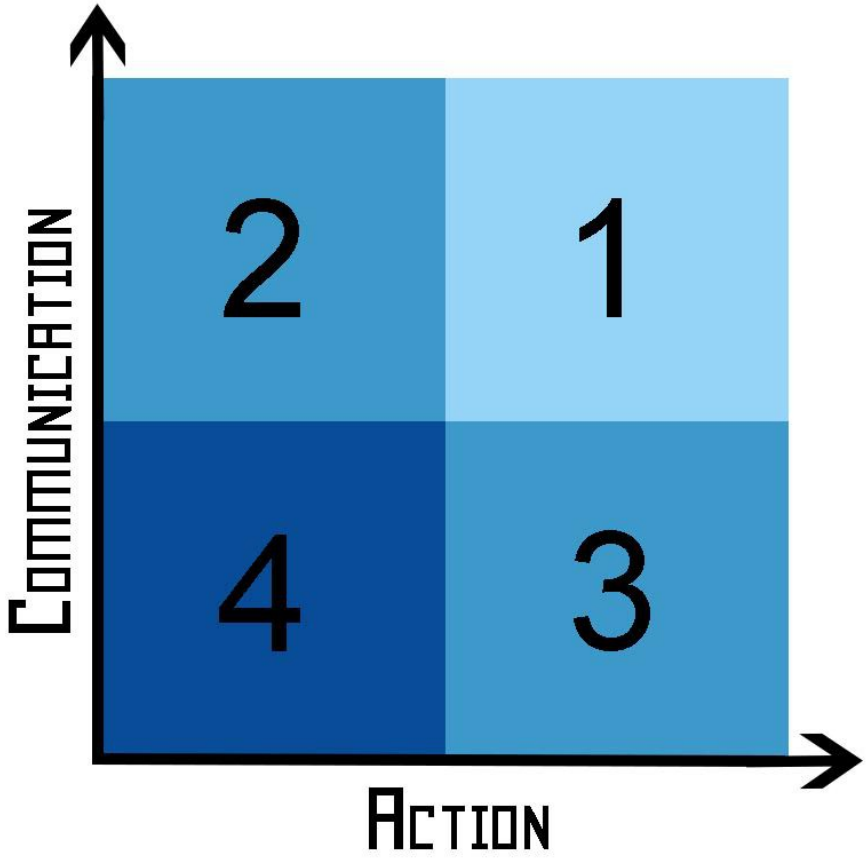
- The GSBPM is on its adaption phase
- It has not been used for geospatial related processes
- The level of in-house support rises as an important factor in testing

A suggestion of further work after the Finnish contribution

1. **Study the national processes of spatial statistics, Finland + *other countries***
2. Recognise mutual phases
3. Identify and agree mutual phases against the GSBPM model
4. First ideas of items that should be included in the GSBPM descriptions
5. Evaluate draft additional texts with other interested parts
6. Study the potentials of other descriptive methods (e.g. GSIM)

Hopefully later

In-house support - *Maturity* of applying the GSBPM



1. The right persons were found, discussion was fruitful, support promoted the work
2. The right persons were found, discussion was partly fruitful but practical results were missing, applying of the GSBPM seemed to be too large workload
3. The GSBPM experts were found but no proper benefits of discussions, no support for the applying
4. The project member could not found support for the work

In Statistics Finland

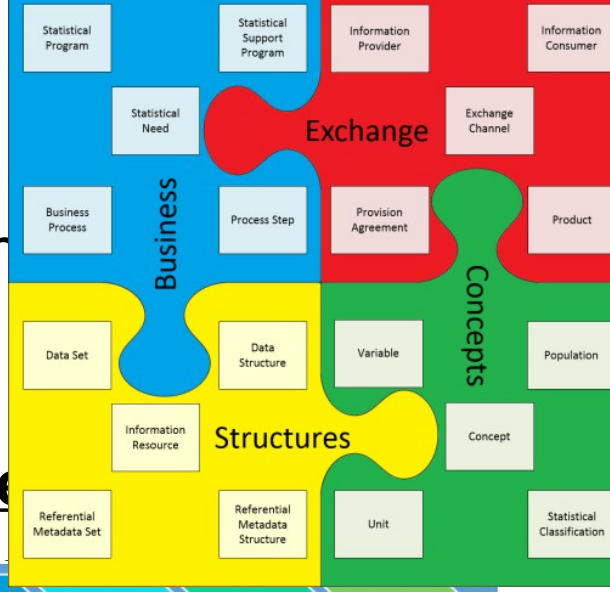
- 3 members of the project group
- Additional geospatial expert
- GSBPM expert

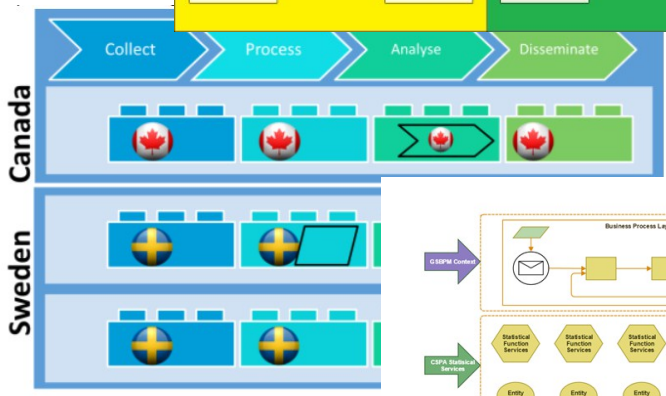
Further work, the GSBPM

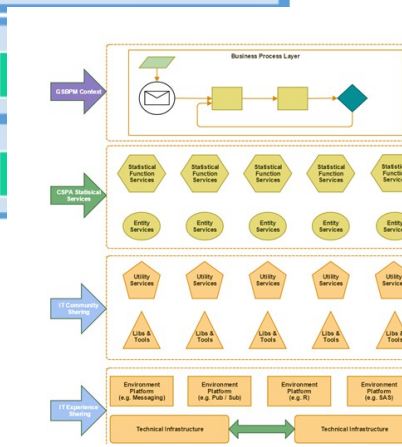
- For managing geospatial dimension - a common view is needed
  - For that, more experiences still needed
  - No easy ways: applying requires to understand geospatial processes in organisations
  - For statisticians point of view important is to have a *common understanding* (in own organisation or in international context)

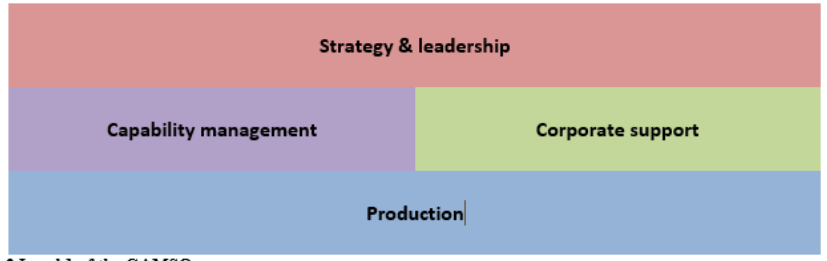
Further work, the other models

- The GSIM, Generic Statistical Information Model for star
  work
  - Are geospatial elements included?
- The CSPA, Common Statistical Production Architecture
  for supporting coherency, for service-based
  production
  - What the services are and from which GSIM elements they are constructed ?
  - From geospatial point of view: e.g. geocoding, services to run geospatial analysis,..
- GAMSO, Generic Activity Model for Statistical Organisations
  for an organisational view
  - An organisational view **also for geospatially related *strategical* objectives**









Conclusions

- The GSGF into ESGF (Geostat 3), the reference frameworks give a structure to focus the developing (by a consistently way in different levels)
- The models are definitely the right way to indicate the geospatial parts of the statistical production processes
  - For a common view new implementing projects are needed
- The models and their documentations should cover the geospatial dimension
- Support for implementation: a gap between the developers of the models and the geospatial statisticians (who wish to apply)
  - Is the community that develops the models focusing only on *abstract and theoretical aspects* more than *supporting the implementation* of the models? Do we need an intermediate actors in between?

Thank You!

Statistics Finland

Rina Tammisto rina.tammisto@stat.fi

European Forum for Geography and Statistics 2016, Paris
15 November 2016