

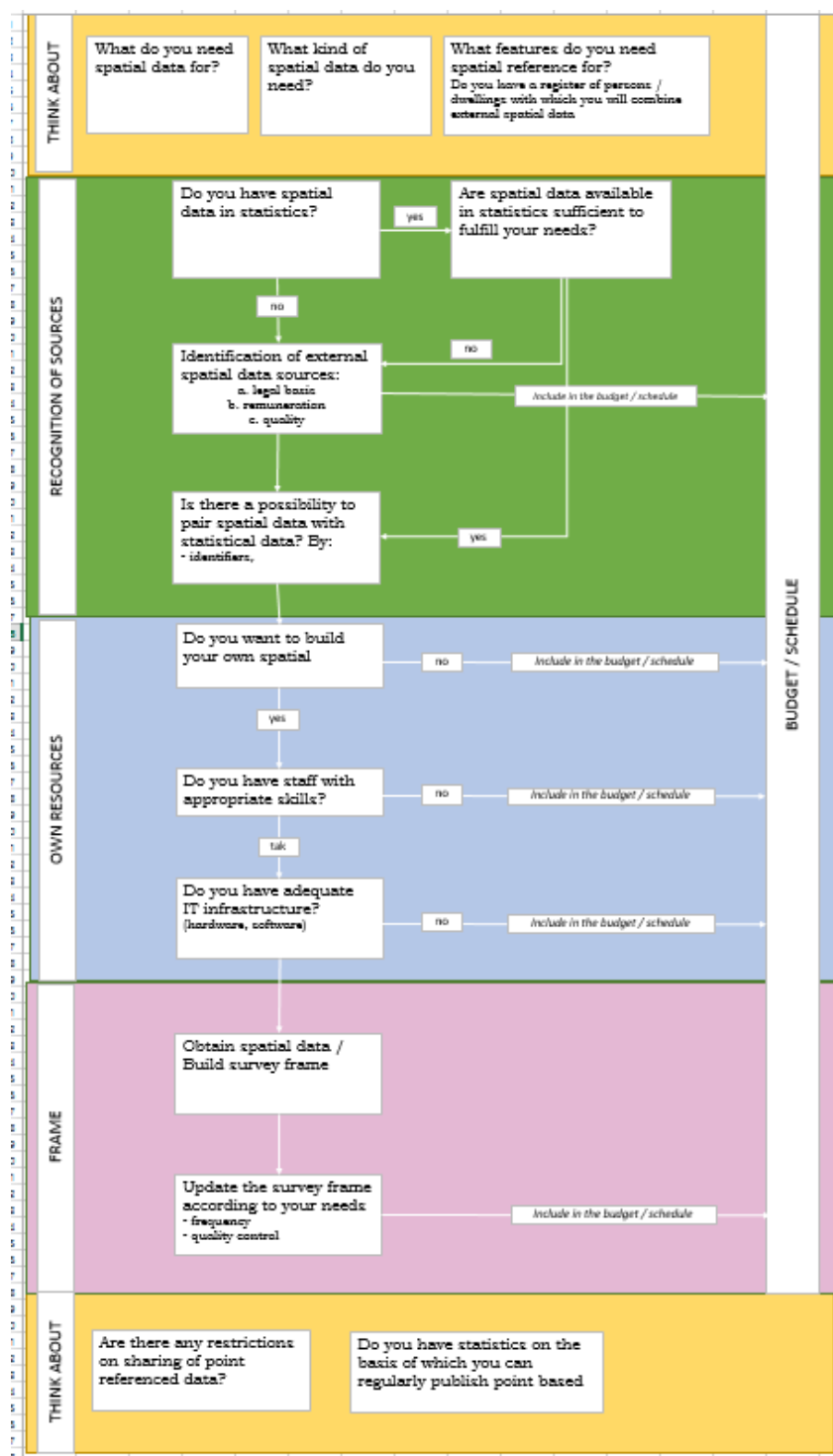
## C. Integrating geospatial and statistical information

112. Statistical data is almost always related to a certain physical space, like a municipality, a state, a country a region, etc. Each level is useful for different actors and different kinds of decisions. Many of those decisions are conditioned by physical elements from the environment, and beyond that, they will have an impact on it. Location and amounts of natural resources, soil types, weather conditions, communications infrastructure, facilities are examples of geographic information which are indispensable elements to fully understand the figures that official statistics generates.

113. The geospatial and statistical data integration landscape is complex. The Global Statistical Geospatial Framework ([GSGF - UN GGIM](#)) and initiatives such as GEOSTAT projects (Eurostat) are vital for developing a consistent and systematic approach to linking geospatial and statistical data.

114. The Global Statistical Geospatial Framework (GSGF) is a high-level framework which facilitates consistent production and integration approaches for geo-statistical information. It is generic and permits application of the framework principles to the local circumstance of individual countries.

115. An example of good practice of assessing the maturity and capability of organizations for spatial statistics is the decision tree developed in the GEOSTAT 2 project. It is a path of practical dynamic questions to be answered before embarking on integrating geospatial data with statistical data as shown in the figure below.



**Figure 3. Decisions related to integrating geospatial data**

116. Developing a coherent and systematic approach to linking statistical and geospatial data is likely to take some time. The best way to achieve consistent integration is through having a common method of geospatially enabling statistical and administrative data and integrating this with geospatial information through an internationally agreed the Global Statistical Geospatial Framework, which enables comparisons within and between countries.

117. Integration can take place at any stage of the statistical production process, as described by the GSBPM. The integration includes geocoding of statistics, spatial analysis, and creating statistical maps. As part of the integration process the following steps may occur:

- Geocoding statistical information at unit-record level
- Processing and manipulation of statistical information using spatial analysis techniques with the purpose of selecting information or deriving new information with a focus on their spatial characteristics, e.g. buffering around spatial features
- Supporting a more efficient and flexible statistical production process with geospatial information e.g. for surveying and sampling, field operation
- Combining statistical end products with geospatial information in statistical maps

- Improving the quality of existing statistical products adopting spatial models, e.g. commuting information by calculating journey times based on detailed transport networks.

118. All statistical phenomena that can be associated to a location are in principle relevant for the integration of statistical and geospatial information. Location in this context means the location of the most individual observation at unit record level. In most cases the location will be a point with coordinates or a precise address. However, other spatial reference frameworks like lines or polygons are relevant as well representing e.g. road segments or areas with a certain land cover.

119. Integration of geographical data with statistical data aims to improve the value of the statistical information that is being produced. Geographic information systems (GIS) as far as it is possible should be used at all stages (inventory, preparation, progress, monitoring, dissemination of results) of the geospatial integration. Wherever it is possible, data should be collected with reference to an address point - results can then be disseminated using any desired spatial divisions. GIS technology should be considered only at a level appropriate to the skills and resources available and constitute an integral part of the overall work of a national statistical organization.

#### "The 10 Level Model"

120. An example of a detailed practical model is "The 10 Level Model" developed by CSO Poland. It can be used to better understand and develop a statistical and geodetic reference framework as a standard of geospatial data production.

121. Recently Statistics Poland worked on the project which aim was the improvement of the use of administrative sources. As a result of the Polish experience, the methodology of assessing the usability of administrative data sources has been elaborated. Nevertheless, quality assessment should be conducted separately for each register, taking into account its possible use. The methodology of assessing register quality will include a few sections. For purpose of statistical division based on geodetic division the section regarding information about the quality of spatial data register will be the most important. The issues included under this section will enable assessing the overall quality of data sets, and the quality of data which they include. Within this area two criteria have been distinguished, i.e. accuracy and comparability which will be measured by specific indicators.

122. Polish methodology of assessing spatial data sets could be a proposal of standard for other countries and statistical organisations which want to harmonise statistical and geodetic divisions to receive better quality of statistical geospatial data and analysis.

123. The opportunities arising from integration of geospatial data with statistical information include:

- integrating the sets of data from different sources of data ex. administrative data educational data, mobile data with geospatial information
- increased added value of statistical and spatial data
- better quality of geospatial data, integrated statistical and spatial data
- better interoperability of sets of data, possibility of analyses, easier methods for linked data sets
- better quality and different spatial analyses
- many cartographic methods of presenting data
- new kind of services and data for users' needs
- flexibility to use statistical information by external users
- useful for policy and decision makers, especially for regional policy makers
- useful for scientific purposes
- useful in environmental protection
- enhanced collaboration between mapping agencies with statistical institutions with maintenance timeliness sets of data and systems.

124. The challenges include:

- sources of data, quality of data
- format files and reference system of the data sets from different sources
- differences in classification of territorial units among countries outside EU
- the reference time of the data
- aggregation level of the data sets
- budget restrictions
- legal issues
- expertise and knowledge
- confidentiality
- accessing skills in GIS and other geospatial related areas
- the need for additional technology to prepare geospatial data and to publish results
- standardization of identifiers (or other information by which it is possible to link statistical information to geospatial information)
- collaboration between mapping agencies with statistical institutions and other institutions (scientists)
- testing/exploration of data source before getting further details.
- secure ways to exchange the data sets
- secure ways for processing data sets
- close cooperation is needed with geospatial data providers (administration, mainly National Mapping Agencies, other organisations).