Principle 3:

Common geographies for dissemination of statistics

Poland & Canada

Janusz Dygaszewicz

Statistics Poland

Reference materials

Primary:

- Ortophotomap,
- Cadastral Data,
- **Administrative division borders,**

Secondary:

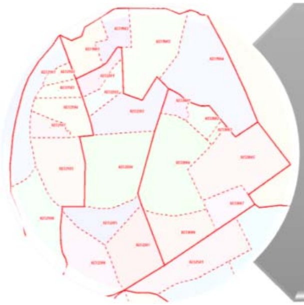
- LPIS (Land Parcel Identification System),
- Road and street network (commercial),
- **Geographical Names,**
- Topographic Data Base.

Spatial address database – consistency issues

With the use of the reference materials obtained, both from geodetic and statistical resources, it was possible to conduct field surveys and develop sampling frames for surveys and censuses, comprising statistical address points and their spatial reference.

Statistical address points

- address points for dwelling locations (only)



Statistical division boundaries

- Statistical regions
- enumeration areas

Benefits from geocoding frames

- Geocoding frames for surveys allows publishing survey results on maps in any spatial division:
    - administrative division
    - statistical division
    - grid
    - any chosen area
- Collecting data in statistical surveys with the precision of **x,y** coordinates will allow a broad use of geostatistical analyses for handling outputs of statistical information.

Principle 3: Common geographies for dissemination of statistics

Principle Definition

The key purposes and benefits of a common dissemination geography:

1. To enhance the dissemination of integrated statistics and spatially enabled data within and among national statistical programmes and other institutions. This can be achieved by promoting stakeholder participation in the elaboration and subsequent of common geographic classifications (a classification constituted of one or more common dissemination geographies). This should also promote increased data consistency and efficiency among participating NSOs, NMAs, NGIAs, international and regional organizations and NGO Open Data institutions.

2. To enable basic comparative statistical reporting, geostatistical analyses and visualization at different scales: the sub-national, national, regional (e.g. European Union, North America, and Africa), institutional (e.g. la Francophonie, G20) and the global scales.

Principle 3: Common geographies for dissemination of statistics

Principle goals and objectives:

1. **The goal is** to enable the consistency and comparability of integrated statistical and geospatial data.
2. That common dissemination geography **be collaboratively assessed and acknowledged** by interested stakeholders prior to the adoption (by NSOs, NMAs, NGIAs, international and regional organizations and other important and key institutions, e.g. NGO-VGI, Open Geospatial Consortium, private sector).
3. That participating NSOs, NMAs, NGIAs, international, regional and NGOs endeavour to integrate acknowledged common dissemination geographies (objects) within existing and emergent statistical geospatial infrastructures.

Principle 3: Common geographies for dissemination of statistics

Principle goals and objectives:

4. NSOs, NMAs and NGIAs that adopt a common dissemination geography are encouraged to move forward and begin producing comparable and integrated social, economic and environmental data, indicators and other information from the integrated statistical geospatial infrastructure. Two objectives may be attained:

   1- an enhanced capacity to produce data and indicators for domestic/national purposes, and;

   2- to meet the monitoring and reporting data and indicator requirements of global initiatives (e.g. the 2020 Round of Population Censuses, and the 2030 Sustainable Development Goals).

5. To acknowledge the continuing need for relevant country-specific dissemination geographies. Proposed and adopted common dissemination geographies should be viewed as congruent and adjuncts to the existing administrative and statistical geographies maintained by NSOs, NMAs and NGIAs.

Principle 3: Common geographies for dissemination of statistics

Principle goals and objectives:

6. The availability of metadata and related documentation on delineation methodologies for any new common dissemination geography is key to ensure the ongoing coherence of the concepts and interpretability of the delineation methodologies used to maintain common dissemination geographies – especially for the continued use by stakeholders. Adherence to existing and the evolving international and national statistical and geospatial metadata standards is a means to achieving this goal.

7. To enable the concordance, where applicable, between common dissemination geographies and established national administrative and statistical geographies (further enabling both comparative statistics and geospatial analysis).

8. To ensure the broadest array of potential common dissemination geographies, legal-administrative*, statistical** and more recently initiated integrative*** geography types be considered and eventually be represented:

Principle 3: Common geographies for dissemination of statistics

Principle goals and objectives:

9. To ensure that the evolving national and international data privacy and data quality principles, frameworks and practices are considered and respected in the design of common geographic areas, their adoption and subsequent implementation (e.g. data and indicator reporting, distribution and visualization).

10 To undertake regular and cooperative reviews of any adopted common dissemination geography to ensure continuing relevance to key national, regional and international programs.

Relationship to other principles

- **Accessible and Usable – producing accessible data for common dissemination geographies:**

Common dissemination geographies and linked and aggregated data should be open and accessible to users and stakeholders within and outside of national statistical and geospatial organizations. Data produced by common dissemination geographies is a key factor to achieving basic comparable and interpretable statistics, enabling advanced and exploratory geospatial analysis and a corner stone for producing clear data visualizations.

Relationship to other principles

- **Interoperable data and metadata standards – enabling the sharing of data and information for common dissemination geographies**

A pre-condition to adopting a common geographic dissemination geography is the necessity that it be interoperable within the existing and emergent national integrated statistical geospatial infrastructures. The goal is to facilitate, not hinder, the sharing of data within and among national statistical, international and regional organizations and other institutions. Another pre-condition is the creation and maintenance of metadata based standards: clear and structured descriptions of the meaning of all concepts, data, variables and uses (and limitations).

Relationship to other principles

- **Geocoded unit record data in a data management framework – enabling the linkage and aggregation of data to common dissemination geographies.**

Within each national integrated statistical geospatial data management framework all unit records should (ideally) be linked to a least one type of geographic object (a highly accurate point – *a precise longitude and latitude*, or the smallest polygon entity– *smallest geographic area entity possible*). This will allow data for efficient (simplified) and more accurate (precise) unit record-to-geographic entity linkage within the existing national (integrated) statistical geospatial infrastructure and the allow for the aggregation of data to most current and future common dissemination geographies (sub-national, regional and international).

Relationship to other principles

- **Fundamental geospatial infrastructure and geocoding – maintenance of common dissemination geographic areas within an integrated statistical geospatial infrastructure.**

That common geographic areas and the data produced and visualised by any modality relies on a structured (efficient and accurate) and organized (to standards) integrated statistical and geospatial infrastructure.

Principle optimal implementation and other pathways

- **The ability to be able to consistently disseminate accurate and reliable statistics at a small area geography level.**

In some cases the smallest geographic level for which unit records are coded to and subsequently disseminated for is the census enumeration area or district geography. In other cases, the smallest common geography is defined by the intersection of features such as road network, other physiographic features (e.g. hydrography) and in certain situations selected administrative boundaries (e.g. state or province boundaries). Small area geographies are sometimes defined by cadastral or property lines. These two latter types of small area geographies area often referred to as the *block* geography.

Principle optimal implementation and other pathways

- **The ability to aggregate data accurately from the smallest common dissemination geography to larger/higher level dissemination geographies to meet the widest possible array or user and stakeholder needs.**

This would include aggregations to both higher level administrative and statistical geographies (e.g. electoral areas, census tracts). There is nothing inherently incorrect with a unit record-to-small area geography approach. However, this approach will become less optimal over time as the demand from stakeholders for statistical data for new and additional common higher level geographic areas are made of institutions responsible for integrating statistical data and geospatial infrastructure.

Principle optimal implementation and other pathways

- **An additional optimal implementation strategy is the coding of all unit records-to-a-precise point location.**

This approach allows for the possibility of more flexible and accurate data aggregation. However, a concomitant challenge of the adoption of a more optimal (precise) unit record locational approach is the often implicit expectation that recasting or recoding of older data coded to a more accurate point location is feasible in all cases. In each case/situation the feasibility would need to be assessed and transparently communicated to stakeholders and data users in order manage expectations.

Principle optimal implementation and other pathways

- **An additional optimal practice would be to further link unit records to cadastral related and building/structure data.**

The integration of cadastral and building/structure information would introduce two high value added data into an existing integrated statistical and geospatial data framework at the point location and/or geographic level.

Countries should agree on a unique official reference data per country for geocoding statistics and producing geospatial statistics with clear ownership, defined scales and attributes taking into account statistical requirements. Both administrative data sources and survey information should be geocoded to the same reference system. Countries should make use of these geospatial reference data, mandatory for all public stakeholders at all government and administration levels, and for all public data and administrative tasks (Geostat 2 conclusion).

Principle inputs

- According to the GEOSTAT 2 project geospatial data on address locations, buildings/dwellings and/or cadastral parcels form the complete basis for a point-based geocoding framework for statistics.
- The Global Statistical Geospatial Framework advocates the recognition of fundamental and authoritative geospatial data from the National Spatial Data Infrastructures or other nationally agreed upon sources.

Principle inputs

To preserve collinearity in case of any changes made by the cadastral service appropriate changes must be made on the statistical side. Therefore statistics should consider the possibility of harmonizing statistical division (statistical regions and census enumeration areas) with the cadastral division (cadastral units), taking into account the needs of official statistics.

In reference to the above, Polish proposal of the 10 Level Model could be used to better understand and develop statistical and geodetic reference framework as a step towards better quality of statistical geospatial data production.

„The 10 Level Model”
for harmonization of statistical and geodesy reference framework

Geodetic System	Layers (suitable for geocoding)	Statistical System
+	NUTS1 - Administrative level 1	+
+	NUTS2 - Administrative level 2	+
+	NUTS3 - Administrative level 3	+
+	LAU1 - Administrative level 4	+
+	LAU2 - Administrative level 5	+
Cadastral units Cadastral parcels	INDIVIDUAL UNITS level 6	Statistical regions Enumaration areas
+	POLYGON level 7	?
?	GRID level 8	+
+	LINE level 9	?
+	POINT level 10	+

Proposal of CSO Poland

Principle inputs

- The level of integration of spatial objects used in statistics and geodesy is presented and described below in „the 10 Level Model" for harmonisation of statistical and geodetic reference framework:

- To support works on common geographies for dissemination of statistics the proposed model should be the subject of intensive works in order to overcome existing barriers and as a starting point to make practical progress in the methodology of combining spatial data with statistical data.

External dependencies

Gaps:

- Depending on various traditions, the choice of a dataset for point-based geocoding may vary between countries. In some countries, location data frameworks comprise integrated combinations of address information, building/dwelling data and cadastral parcels. Ideally, these objects are consistently and hierarchically linked to each other, both conceptually and topologically, which enable the inclusion of all three object types in the geocoding infrastructure. Yet, in other countries only building data or address data exists (Geostat 2 conclusion).

External dependencies

- Implementation of “The 10 Level Model” into GSGF could support practical activities of integrating statistics with geospatial data. “The 10 Level Model” could help to communicate and understand the geospatial capability requirements for statistical information. This could support data sharing between institutions, enhancing the interoperability of geospatial and statistical information and building institutional collaboration between geospatial and statistical communities. As a result it is expected to enable new, better quality and more integrated information for analysis and decision-making process.

Community roles

- Fundamental and authoritative geospatial data from the National Spatial Data Infrastructures is typically maintained under the authority or supervision of NMCAs. Local and regional administrations may also be involved in data collection, but in most cases the NMCAs gather and store data from municipalities in centralised repositories. In some countries, NSIs have established a direct collaboration with the municipalities providing location data to statistical offices.

Community roles

- Typically, geospatial data is of good quality if it is regularly used, e.g., for administrative purposes or business activities. Citizens usually have an incentive to provide correct and up-to-date addresses to administrations as they can expect benefits and services in return, such as health care, tax refunds or social benefits. Hence, address registers or other geocodes used for statistics should ideally be the same as for the administration, i.e. one single unique address register for all applications.

Summary of required standards

Relevant international and national standards, frameworks and infrastructure:

- Common geographic classification, including use of administrative and statistical geographies that are complemented by use of grid type geographies.
- Standards or guidance on the use of geographies for dissemination of data.
- International statistical and geospatial metadata standards.
- Systems and methods to correspond data between geographies. (http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Correspondences)
- National privacy laws and/or agreed privacy principles, in accordance with the United National Fundamental Principles for official Statistics.

Summary of required standards

Quality measures:

- Quality assessment is really important from the point of view of the usefulness of combining spatial and statistical data. In case of the harmonization of statistical division based on geodetic division common geographies will cause that quality of statistical geospatial data and conducted analyses will depend on the quality of input spatial data from external administrative registers. That is why assessment of the overall quality of external data sets, and especially the quality of data which they include is essential.
- Assessing the overall quality of data sets, and the quality of data which they include can be characterized by two criteria:
  - - Accuracy which indicates the extent to which the register reflects real values
  - - Comparability which indicates the degree of the methodological compliance of register data with statistical survey data. This criteria can be measured by specific indicators.
- The quality measures of data in register:

„The 10 Level Model” for harmonization of statistical and geodesy reference framework

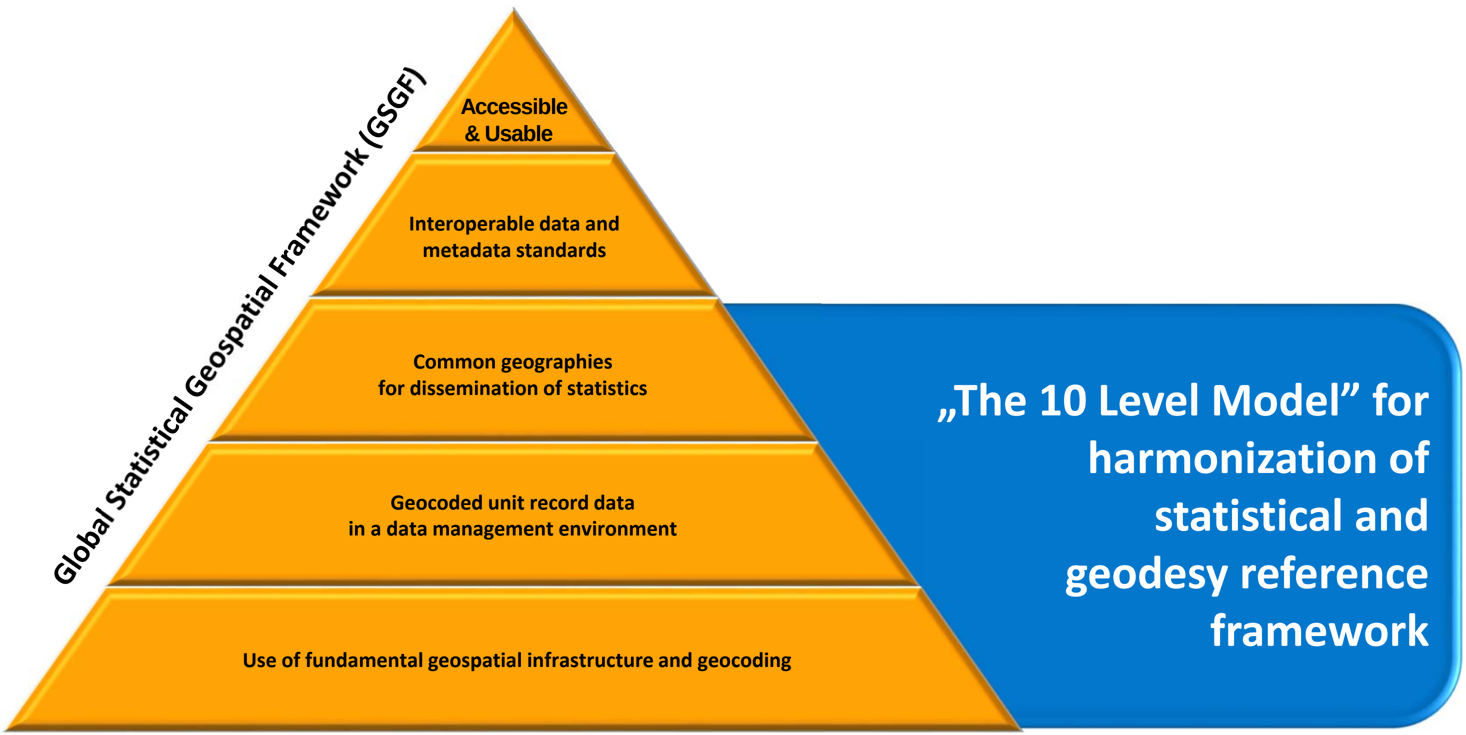
Conclusion:

The question marks in the proposed model (lack of grid on the geodesy side and lack of linear objects on statistical side) should be the subject of intensive works in order to break down existing barriers and as a starting point to make practical progress in the methodology of combining spatial data with statistical data, with particular emphasis on the specifics of environmental phenomena.

Geodetic System	Layers (suitable for geocoding)	Statistical System
+	NUTS1 - Administrative level 1	+
+	NUTS2 - Administrative level 2	+
+	NUTS3 - Administrative level 3	+
+	LAU1 - Administrative level 4	+
+	LAU2 - Administrative level 5	+
Cadastral units Cadastral parcels	INDIVIDUAL UNITS level 6	Statistical regions Enumaration areas
+	POLYGON level 7	?
?	GRID level 8	+
+	LINE level 9	?
+	POINT level 10	+

Proposal of CSO Poland

Implementation of "The 10 level model" into the Global Statistical Geospatial Framework (GSGF)



- Implementation of “The 10 Level Model” into GSGF could support practical activities of integrating statistics with geospatial data. “The 10 Level Model” could help to communicate and understand the geospatial capability requirements for statistical information. This could support data sharing between institutions, enhancing the interoperability of geospatial and statistical information and building institutional collaboration between geospatial and statistical communities. As a result it is expected to enable new, better quality and more integrated information for analysis and decision-making process.

Merging statistics and geospatial information in EU Member States

Common aim of geo-statistical researches:

To development of a geo-statistical division framework for official statistics - with respect to the geodetic division of the country and needs of statistics - related to survey sampling and quality assurance of final statistical products for dissemination of census data and SDG indicators

The current GSBPM model v 5.0 could be modified by adding the following sub-processes:

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

Geospatial components in the GSBPM model

Generic Statistical Business Process Model
UNECE/Eurostat/OECD

Quality Management/Metadata Management

1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.1 a Geocode frame & sample	5.2 Classify & 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 Build or enhance dissemination components	4.2 Set up collection	5.3 Review & validate	6.2 a Prepare spatial analyses & maps	7.2 a Manage spatial analyses & maps using GIS	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.3 Run collection	5.4 Edit & imput	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system	4.3 a Geocode collection	5.5 Derive new variables & units	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.6 Prepare business case	2.5a Design geocoding frame, sample & data collection	3.6 Test statistical business process	4.4 Finalise collection	5.6 Calculate weights	6.5 Finalise outputs	7.5 Manage user support	
	2.6 Design production system & workflow	3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Thank you for your attention

Janusz Dygaszewicz

Statistics Poland

j.dygaszewicz@stat.gov.pl

The GRID

In Europe the GEOSTAT population-grid dataset is promoted as a first example of a European Union (EU) population grid.

The main assumptions of GEOSTAT grids are:

1. uniform coordinate system for all European countries (ETRS-LAEA)
2. reporting at the European level in the 1km x 1km grid NET.

The system of grids with equal-size grid cells promoted in Europe has many advantages:

- grid cells all have the same size allowing for easy comparison;
- grids are stable over time, enabling easier data analysis in time series;
- grids integrate easily with other scientific data (e.g. meteorological information);
- grid systems can be constructed hierarchically in terms of cell size thus matching the study area; and
- grid cells can be assembled to form areas reflecting a specific purpose and study area (mountain regions, water area)