UNITED NATIONS SECRETARIAT
Department of Economic and Social Affairs
Statistics Division

ESA/STAT/AC.279/P7
October 2013
English only

United Nations Expert Group on the Integration of Statistical and Geospatial Information
First Meeting
New York, 30 October - 1 November 2013

Agenda: Item 6

Some practices of Geospatial coding and referencing statistics in Finland and in Europe [1]

Prepared by Finland

[1] This document is being produced without formal editing

Marja Tammilehto-Luode

16 October 2013

Some practices of Geospatial coding and referencing statistics in Finland and in Europe

1 Introduction

Since 1990, population and housing census has been conducted entirely on the basis of register data in Finland. The register-based population census system is built around a set of basic registers, which contain comprehensive data on the units that are to be described in the population census. These registers cover all people resident in Finland, the buildings and dwellings in the country, as well as all enterprises and their establishments. All statistical units can be linked to one another with the help of the identification systems: persons can be linked to families and household-dwelling units, to the dwelling and building in which they live, and to the employer for whom they are working. Similarly, all units can be located on the map using the map coordinates of the centroids of buildings.

This paper presents an overview on how Statistics Finland uses centralised databases for integrating statistical units to map coordinates and other geospatial data. The paper also contains some information about efforts towards national and European standards, which are relevant for integration of statistics and geospatial information. Finally, it includes an overview of major challenges in the field of spatial statistics.

2 Capabilities for geocoding

Building coordinates and addresses

Since 1987, all census data have been annually produced using data of administrative registers in Finland. The register of Buildings and Dwellings (RBD) plays a central role for geocoding census data. In fact, location information of the buildings in our statistical system enables geocoding of different types of datasets. The statistical units only need to be linked to buildings through different links, such as personal identity code, property identifier (or address) or e.g. Business ID.

During the 1970 census, the first efforts to assign map coordinates to centroids of each building were made. The coverage of map coordinates of buildings have improved gradually, and currently the coverage for inhabited buildings is over 99% and around 92% for workplaces. The RDB also contains information about addresses and dwellings of each building. The address information is used to define postal code areas. The address information is used also as a secondary source for “geocoding” and for quality checking of data.

There are several other sources of address information in Finland. However, the most important source of address information is the register of Buildings and Dwellings. Another relevant source of address information is a national database maintained jointly by the Finnish Land Survey, the Finnish Transport Agency and municipalities. It contains precise and accurate data

on the location of all roads and streets in Finland, as well as their most important physical features (DigiRoad).

Grid net, boundaries of territories and other spatial data

In this context, data refers mainly to vector data, which has typology and are used as a source for geographic layers.

An important geospatial dataset for grid-based statistics is a grid net, covering the whole country (and the whole of Europe, see EFGS website). Statistics Finland provides three sets of grid nets: by 250 m x 250 grid cells and 1 km x 1 km grid cells and by 5km x 5km grid cells. Grid nets were established according to the standard national ETRS89 reference system with a common origin and an unambiguous identity code for each grid cell, including an assignment of each grid cell to the corresponding municipality or postal code area.

Statistics Finland also maintains boundaries of municipal sub-areas and postal code areas for statistical purposes. Municipal sub-areas are defined by municipality authorities on a voluntary basis, and stored and updated for statistical purposes by Statistics Finland. Postal code areas are created with the help of spatial analysis tools by using the postal codes of each building.

The National Land Survey of Finland (NMA) provides municipal boundaries. In this case, collaboration includes not only standard naming and coding of municipalities, the NMA also provides boundaries of different scales and for different generalised levels (e.g. with a coastline or without a coastline) for different purposes in the statistical production flow. Statistics Finland makes then available different kind of administrative and statistical areas (divided by municipalities). The role of the NMA in providing geospatial information for the NSI has increased gradually because of a new open data policy and because of the implementation of the INSPIRE directive.

All these spatial territories can be linked to statistics by their identity codes and names. However, there is also a direct link via map coordinates of centroids of buildings.

Nomenclature of administrative units

In Finland as in many European countries, regional and local administrations are traditionally quite strong. There is a constant demand for basic statistics by administrative areas. Standards for coding and naming of these areas are available and are used quite consistently in official registers, statistics and maps. Statistics Finland plays an important role in maintaining standard codes and names of all administrative territories of Finland. These codes are an important indirect way of geocoding for all kinds of statistics.

All European countries have defined three hierarchical levels of territories for delivering statistics to the European Union. These territories are mainly based on administrative areas (NUTS).

Statistics Finland has a central database for nomenclature of administrative units. It includes also classification of nomenclatures for different use,

nomenclatures of different time periods and links between different nomenclatures.

3 Spatial reference system in Statistics Finland

Data warehouse

Centralised maintenance of statistical units including nomenclatures and spatial units enables most efficient and unified production processes of different kinds of statistics and statistical services. Updated and uniform data are available for every user. Since 2005, all census-type data in Statistics Finland is maintained in one SQL database. Another centralised uniform database is in preparation for business statistics, and it will be put into operation this autumn.

In the census "data warehouse", the key source for all territorial statistics is "the building table". The building table includes the location of all buildings by map coordinates but also official codes of all territorial units to which each building belongs. This means e.g. municipal code, postal code, urban area and even grid code. The map coordinates are based on information from the Register of Buildings and Dwellings, and so are municipal codes, but some territorial units are deduced from available boundary information. These territorial units include e.g. municipal subareas, postal code areas or urban and rural areas, which are regularly used in statistical production. All statistical data/attribute data in the same database can be linked to the "building table", from which necessary geographic information is used to compile statistics by different domains and combinations. In general, a user of the database has all territorial units available to aggregate data by different territories if only she or he is able to link the data to buildings.

The use of "the building table" is also a starting point of geocoding enterprises' local units (establishments).

For the time being, no spatial components of SQL are used.

Geodatabase

Parallel to the warehouse, Statistics Finland maintains a Geodatabase with all kinds of vector and raster data about statistical area boundaries used for statistical production or for spatial analysis purposes. Boundary data and other kinds of map data are used for deducing statistical areas of buildings, for visualising statistics on maps but also for integrating statistics and geospatial for geospatial statistics and for spatial analysis. At the moment, the Geodatabase is maintained by ArcGIS technology.

All geodata, including building coordinates, follows a major European terrestial reference system (ETRS89), which is gradually being implemented in all European countries, standardising the framework for geospatial data. The implementation of ETRS89 in Finland is ETRS-TM35FIN, which is a common reference system for all geospatial data in the national geoportal.

Statistical grids, important spatial statistics

Grid-based statistics is the most used "geospatial statistics" in research but also in the public sector and private enterprises in Finland. A grid cell georeference is the map coordinate of the lower left corner of each grid cell and/or an identity code assigned on the geospatial grid net. Statistical variables are calculated and displayed by georeferenced grids.

In the framework of the INSPIRE directive, Statistics Finland released earlier this year basic grid-based population data for free. Until now, grid-based statistics have only been for sale as a fixed product (Grid Database) or through a commissioned agreement, individually according to a customer's order.

Grid-based statistics are delivered in a table format (e.g. a CSV files) and/or in a geospatial format (e.g. suitable for MapInfo or ArcGIS software). According to Inspire specifications, some population grid data are now available also via WMS and WFS services (Inspire data).

Statistical grids are also used for delineation of localities (to measure the degree of urbanisation in municipalities) and to maintain national spatial classification of urban and rural areas.

4 Benefits of a point-based geospatial statistical framework

There are several advantages of point-based, as accurate as possible, spatial framework for statistics.

Points, as building coordinates or addresses (or grids) do not change in terms of space as boundaries of administration change. Tracking the changes in space/in regional statistics and especially maintaining and updating regional time series, when administrative borders change is less resource demanding and likely of better quality when there are updated georeferenced point-based data available, instead of polygon-based data (e.g. by administrative areas).

Point-based reference system is the best source for grid-based statistics, and it has gained wide recognition also in the EU as a complementary system for disseminating statistics. Grid-based statistics have great potential for comparable territorial statistics and statistical time series, not only inside a country but also between countries. A grid net is a good potential framework for harmonising the datasets of different kinds of territorial units, when data need to be combined by their location. Grid-based statistics are well suited for spatial analysis purposes.

Data by points or by standard sized grids can be compiled flexibly by small or large areas or by areas defined by natural boundaries, distances or other spatial factors. Data by points or grids describing actual distribution of population can be used to build relevant and comparable functional areas, such as urban and rural areas.

Point-based data or grids are suitable for integration of data to the road network in order to carry out accessibility studies.

5 National and European cooperation and standards

In Finland, the most important organisations to guide the development of nation spatial data infrastructure are the Advisory Committee on Information Management in Public Administration (JUHTA) and the National Council for Geographic Information. Statistics Finland has a representative in both of these organisations. JUHTA has been set up by the Ministry of Finance to promote cooperation in information management between the State and the municipalities. The Committee's responsibility is e.g. to set standards and draw up guidelines. It has set several national standards and recommendations in the field of statistics and geospatial information (http://www.jhs-suositukset.fi/web/guest/jhs/recommendations) and the work is an ongoing process.

The National Council for Geographic Information, where Statistics Finland is also represented, is responsible for monitoring the development of the national spatial data infrastructure (NSDI) and the implementation of the INSPIRE directive (http://inspire.jrc.ec.europa.eu/). It is responsible for the Finnish National Spatial Data Strategy (Location; The unifying factor) and it has published e.g. practical guidelines for the harmonisation of core geographic datasets (Improving operational efficiency with geographic information).

The national geographic information strategy emphasises interoperability and harmonisation of geographic datasets: "A prerequisite for effective use of geographic information at all levels of society is that the datasets are comprehensive and mutually interoperable, possess integrity in logical and technical terms and, above all, are readily available. To achieve interoperability common standards and recommendations need to be applied to the maintenance and management of geographic datasets in widespread use."

Once a year, Eurostat (the European Statistical office) organises a Working Party of GIS for Statistics, to which representatives from both NSIs and NMAs are invited. Based on discussion in the working party and in the European Forum for Geostatistics Eurostat launched the GEOSTAT project (Statistics Finland is a partner) to prepare guidelines for the creation of national population grids and to make recommendations for standards for comparable European grid-based population datasets. The second phase of the project that started in 2010, comes to an end this year, delivering manuals (with case studies) that will set ground for harmonised population grids for the whole of Europe.

In Europe, the Inspire directive regulates NMAs and NSIs to follow harmonised specifications for data delivery of certain data themes. The directive, its first draft and now its implementation, have opened up new and more concrete ways of cooperation between NSIs and NMAs on a national and European level. Statistic Finland has published new kinds of statistical services and has committed to continue cooperation with the NMA in a larger project in 2014-2015.

6 Conclusions and Challenges for the future

The traditional way of geocoding is not widely used in Statistics Finland. However, in the data warehouse of census-type data, georeferencing is implemented by buildings comprising, not only map coordinates but also codes of other territories to which buildings belong. The indirect geocodes are standardised and collected directly from other register data or deduced from available geospatial information (boundary data).

In general, geocoding of buildings with address information is used mainly as a complementary method of georeferencing in the statistical process. However, the use of addresses and other geospatial information might improve the quality of different datasets. At the moment, spatial statistics are not part of official statistics and the use of geospatial information in a statistical production process is not exploited throughout.

Nomenclatures of territorial units are quite standardised. Along with the INSPIRE directive some common standards for European geospatial information were introduced. The GEOSTAT project has made an effort to draft some European standards for grid-based statistics. However, spatial statistics in general are missing standards. There are standards for geospatial information as well as standards for statistics (including regional statistics) but usually neither of them apply well to spatial statistics.

Geospatial information traditionally describes the physical characteristics of our environment and has no data protection concerns. Socioeconomic statistics, as small area statistics (which is usually demanded for spatial analysis), often includes confidential information, which need to be handled one way or another. Integration of geospatial data and socioeconomic statistical data may increase the problem. There is a need for common rules and spatially orientated methods to avoid the risk of disclosure in geospatial statistics.

The compilation of core datasets with consistent georeferences is not yet part of the production of official statistics. In Finland, grid-based statistics, as well as all different types of small area statistics, together with utilisation of geospatial information in general, have been developed for chargeable services. The official statistics are defined to cover statistics by major administrative areas. Maintenance of good quality spatial information and a concern with data security in small area statistics, together with a compilation relevant services by geospatial data demand resources. Diminishing budgets in governmental institutes (including NSIs) and a demand for open data (including geospatial data) cause conflict, which needs to be solved. In this kind of a situation, it is extremely important to increase understanding of the importance of spatial information among public institutes.