United Nations ECE/CES/2010/13

Economic and Social Council

Distr.: General
30 March 2010

Original: English

Economic Commission for Europe

Conference of European Statisticians

Fifty-eighth plenary session
Paris, 8–10 June 2010
Item 6 of the provisional agenda
Spatial statistics

Establishing a national hierarchical grid in Slovenia — lessons learned and future challenges

Note by the Statistical Office of the Republic of Slovenia

Summary

Managing and planning various human activities in the environment or monitoring the trends of different phenomena in space and time requires a wide range of spatial statistical data and and adequate response from data providers. Long time series and register-oriented databases managed by the Statistical Office of the Republic of Slovenia or other authorities were recognised as a valuable support for these tasks and allow to use contemporary means of performing the spatial (statistical) analyses, i.e. to geographical information systems.

The dissemination of spatial statistical data in geographical information systems format is simply a different form of dissemination already in practice (e.g. digital or paper tables) and since the data can be put into a geographical information system, grids represent a great potential for both the users and statistical offices. The Statistical office of the Republic of Slovenia therefore decided to explore the possibilities of managing its own data on grids, which consequently led to the establishment of a national hierarchical grid system and participation in various international initiatives regarding geostatistics.

I. Introduction

1. The increasing impact of human society on the environment demands a strategically planned management while considering the principles of sustainable development. Geographical information systems (GIS) have offered a new perspective on this issue and coincidently opened a new dimension in understanding of dissemination of spatial statistical data. To meet the requirements of the growing community of spatial data users, statistical offices are urged to adopt new means and formats of spatial data dissemination where grids proved to be a widely applicable solution, among others. The Statistical Office of the Republic of Slovenia (SORS) has already explored several sources for creating grid-based statistics and main conclusions and challenges are presented here.

2. Grid-based statistics are derived both from polygons (e.g. enumeration areas) and from point located data. Register-oriented statistics in Slovenia offered a good foundation for creating grid statistics of high resolution. The Register of Spatial Units — initiated by SORS and now managed by the Surveying and Mapping Authority of the Republic of Slovenia — was the first step towards a sound territorial division which enabled first georeferencing (point locating) of statistical data (1971 Population and Housing Census) in Slovenia. These 1971 Census data were used for the establishment of the Central Population Register (CPR) and for the very first time personal identification numbers were assigned to the people residing in Slovenia[1], which is important for later facilitation of entering the data from some registers. Although, at the time, these data could be stored only in tables and not really managed graphically as is the case today by means of GIS, it was decided to permanently preserve the spatial references of the highest possible (or acceptable) positional accuracy.

3. This far-sighted decision became very relevant when the graphical part of the Register of Spatial Units was completed in 1995. The data stored in tables did have their spatial reference but before that it was very difficult or even impossible to analyse them by means of GIS on the entire national territory. In practical terms, this means that from 1995 on population data captured in the 1971 Census, for example, could be graphically presented for each person on a map as accurately as to their house of permanent residence or to the corresponding enumeration area. When SORS started to handle spatial statistical data on grids, the point located data from various registers were considered as most applicable, but recently some methods are being tested on how to improve the positional accuracy of polygon data while point locating them and aggregating them to grids as described further on in this document. Statistical data from the 1971, 1981, 1991 and 2002 censuses, together with the data from the CPR, thus offer an important historical picture of how various spatial phenomena have changed over the last forty years.

4. Preparing the databases applicable to grids was an important step, but there was still a need to exchange experience regarding the handling of statistical data on grids or geostatistics in general. A joint project with Statistics Austria and participation in the European Forum for Geostatistics significantly improved the understanding of creation, analysing and dissemination of grid statistics. International cooperation is essential in order to harmonise these processes and the European Statistical System's ESSnet project GEOSTAT (and its iterations) is expected to provide widely acceptable and applicable solutions.

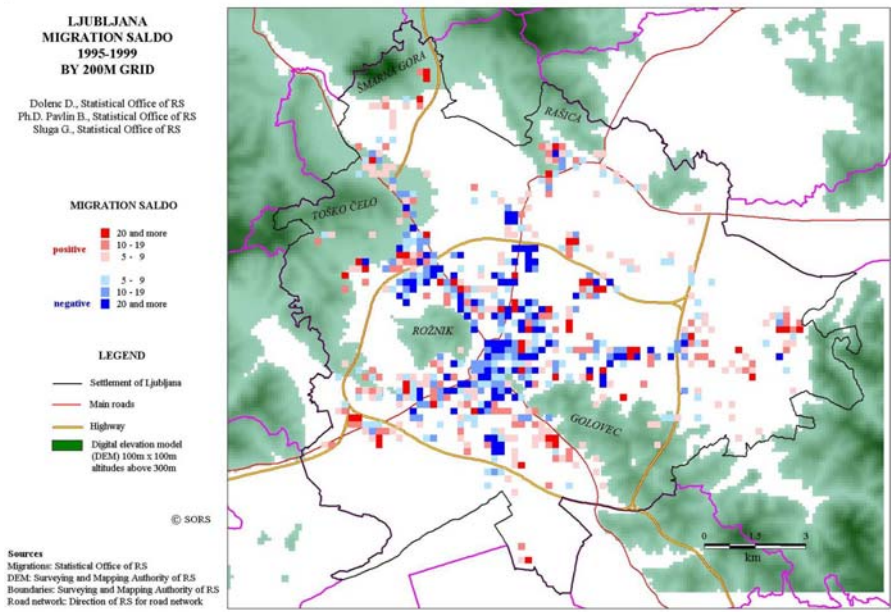
[1] Oblak Flander, A.: Opportunities and Challenges of a Register-Based Census of Population and Housing – the Case in Slovenia. Seminar on Registers in Statistics — methodology and quality, Helsinki, 2007.

5. While learning new perspectives and methods in understanding and handling the grid data, SORS initiated the establishment of a national hierarchical grid system in 2008. Up-to-date information flow and exemplary cooperation among various public authorities in the country promises its wide applicability and possibility of integration into various spatial data infrastructures. After surmounting the first challenges, the definition of the system of data sets being disseminated on grids together with corresponding statistical disclosure control is still ongoing, but work on these issues has already brought some relevant results which are also indicated in this document.

II. Acquiring experience

6. SORS has been involved in the handling of spatial statistical data on grids since the early 1990s with first results of these spatial statistical analyses presented at the end of the decade (Picture 1).

Picture 1
Ljubljana migration balance 1995–1999 on 200m grid



7. Since then, there has been an increase in user demands for statistical data in GIS format, which convinced SORS to further explore the advantages of handling statistical data on grids together with dissemination of such data. Lessons learned from individual case studies and dissemination resulted in the intention to establish a national hierarchical grid system. This grid system was, from the very beginning, understood to be a part of the Register of Spatial Units managed by the Surveying and Mapping Authority of the Republic of Slovenia

8. Grids are commonly described as spatial units that:

(a) Are independent from political or administrative areas;

(b) Do not change in time;

(c) Are evenly distributed (comparable);

(d) Are valuable for micro to macro analyses;

(e) May be manipulated with standard GIS tools;

(f) Are easy to generate from point-based data;

(g) Offer better solutions regarding statistical disclosure control;

(h) Enable better small area estimations;

(i) Enable better sampling.

9. The above-mentioned characteristics were practically applied and tested in a common project between Statistics Austria and SORS in 2005, supported by the European Union (EU) Phare Programme. By then, Statistics Austria had already established systematic dissemination of spatial statistical data together with a corresponding data protection policy on grids and, as a neighbouring country, it was very suitable for presenting different socio-economic data on seamless grid maps for both countries. Cross-border analysis was made on extended Austrian grid in the Universal Transverse Mercator Coordinate System (UTM) projection into which the georeferenced data were aggregated. Since only UTM zone 33 was used, the western part of Austria was excluded. The exchange of knowledge and experience continued with active participation in the European Grid Club and Nordic Forum for Geostatistics, which later became the European Forum for Geostatistics. Being a member of the European Forum for Geostatistics, SORS is also a co-partner in the ESSnet project GEOSTAT. The expert group of European geostatisticians proved to be an excellent platform for discussion of the relevant issues regarding the grids and geostatistics in general.

10. Considering the situation regarding the statistical spatial data infrastructure in Slovenia and studying the experience of several European statistical offices with grids, SORS took advantage of a long-term register-oriented statistical tradition to aggregate spatial statistical data of high positional accuracy to grids.

III. National hierarchical grid system

11. In 2008, SORS decided to launch an initiative for establishing a national hierarchical grid system in Slovenia. Three institutions agreed to cooperate: SORS provided methodological support and the Geodetic Institute of Slovenia together with the Surveying and Mapping Authority of the Republic of Slovenia provided technical support.

12. The purpose of the joint project was to:

(a) Create square grid vector layers with seven different basic sizes of grid cells;

(b) Define the grid cell nomenclature in accordance with the hierarchical structure of grid cells;

(c) Define the origo of the hierarchical grid system;

(d) Define the grid cells both in the previous (D48/GK) and the present (D96/TM) national coordinate system.

13. The seven basic grids are 100m, 200m, 500m, 1 000m, 2 500m, 5 000m and 10 000m grid. The smallest grid cell size 100 m x 100 m was defined considering the user needs for spatial statistical data of high resolution and compliance with other spatial databases in Slovenia.

14. To solve the problem of converting the data from one coordinate system to another, it was decided to create square grid vector layers first in D96/TM and then to transform

them into D48/GK where grid cells from both coordinate systems share the same cell ID. Transformed grid cells in the old coordinate system D48/GK insignificantly loose their square shape but the same cell still covers the same area. Therefore, all official spatial statistical data or user's own spatial data which are mostly still in the previous D48/GK coordinate system can simply be aggregated to grids in D48/GK and then transformed to D96/TM using the cell IDs. The origo of the hierarchical grid system is defined as an intersection between meridian 15^{o} east and a parallel at 5 000 000 m distance north from the equator. 500 000 m was added to the coordinate of the 15^{o} meridian to avoid the negative values west of this meridian. The coordinate of the parallel is 0 m. Y axis is parallel to the equator and has a positive direction towards the east whereas X axis is parallel to the central meridian (15^{o}) and has a positive direction towards the north. The coordinate (500 000, 0) thus represents the origo of the grid system and coincides with the origo of the present D96/TM coordinate system, which simplifies the linking of different spatial data in the territory of Slovenia.

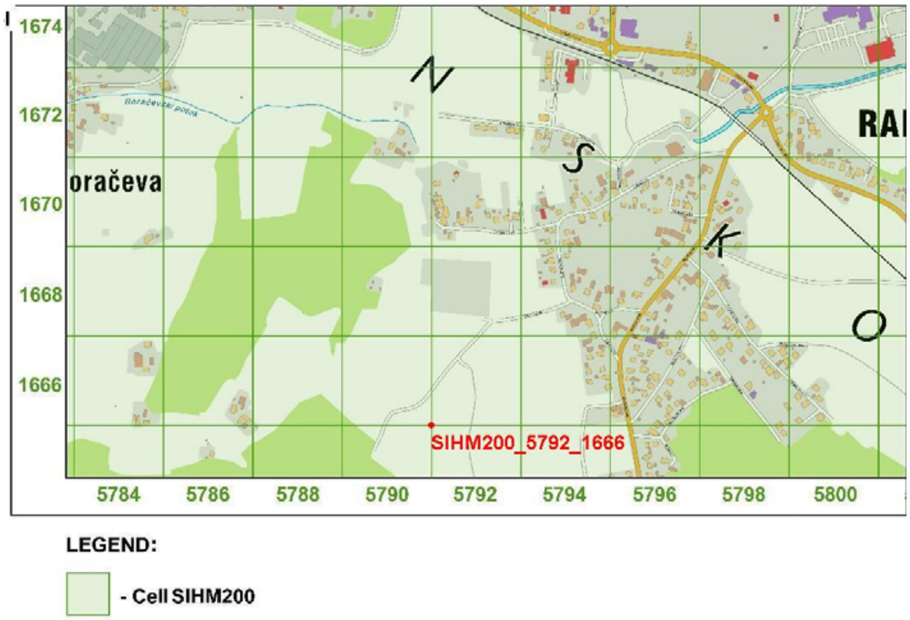
15. The grid cell nomenclature (e.g. SIHM200_5792_1666) consists of: SIHM (Slovene hierarchical grid) with the added size of the grid cell in meters_Y coordinate_X coordinate where both coordinates are truncated by the number of redundant zeroes. The cell ID thus describes both the size and position of the cell.

Table 1

Example of shortening the records of cell IDs.

Complete cell ID record	Shortened cell ID record
SIHM100_463000_103700	SIHM100_4630_1037
SIHM200_463000_103600	SIHM200_4630_1036
SIHM500_463000_103500	SIHM500_4630_1035
SIHM1000_463000_103000	SIHM1000_463_103
SIHM2500_462500_102500	SIHM2500_4625_1025
SIHM5000_460000_100000	SIHM5000_460_100
SIHM10000_460000_100000	SIHM10000_46_10

Picture 2
Example of cell nomenclature for 200 m grid



IV. Spatial statistical data in grid form (bottom-up method)

16. Georeferenced (point located) or geocoded (polygons, e.g. administrative units) spatial statistical data can be aggregated to grids in Slovenia. The sources for grid data are different registers and four population censuses (1971, 1981, 1991, 2002) where register-based data and the Census 2002 data are georeferenced, while other census data are geocoded.

17. Generally, polygon data are treated as more convenient for the top-down method (downscaling method) where, for instance, the population data for municipalities are disaggregated to grids using some auxiliary data such as land use and land cover. This method can be applied in situations where no spatial statistical data of higher positional accuracy are available. Nevertheless, the nature of geocoded census data in Slovenia due to their acceptable positional accuracy still enables the application of the bottom-up method.

18. The 1971, 1981 and 1991 census data are geocoded to spatial districts. Spatial districts are the smallest administrative units in Slovenia and, as they already cover a certain area, their aggregation to grids is different from aggregation of georeferenced data. The process of converting geocoded data into georeferenced data in practice means defining the centre of gravity (centroid) in the form of a coordinate for a particular phenomenon in a particular spatial district. Centroids of spatial districts in Slovenia already meet this requirement since they mostly coincide with the area of the highest population density in that particular spatial district. The centroids are thus defined by the location of significant objects, e.g. schools. Spatial districts without significant objects obtain their centroids from significant natural objects, i.e.:

(a) Centre of gravity of densely built-up area of the spatial district;

(b) Centre of gravity of all buildings in the spatial district when buildings are scattered;

(c) Centre of gravity of the spatial district when there are no buildings in the spatial district.

19. Any territorial change of the spatial district consequently means a change of its centroid. Despite this, the centroids of spatial districts were additionally examined and corrected where necessary, since the population distribution has changed over the past decades significantly in some areas. The correction performed was based on the present state of the centroids of buildings where the information of the construction year of buildings was used to select only buildings which existed and were populated in a particular census period. In other situations where, for instance, the number of working places is presented, the centroid of the spatial district can be corrected according to the centre of gravity of all buildings with business activity. Additionally, the position accuracy of 1981 and 1991 census data on population when aggregated to grids can be examined with georeferenced data from the Central Population Register already available for those periods.

20. Different from georeferenced data, geocoded data determine the grid cell size according to their average area. Spatial districts in densely populated urban areas cover smaller areas and in rarely populated areas, larger areas. Several spatial analyses indicated that the census data can be aggregated to grids with the cell size 100 m x 100 m or 200 m x 200 m for high population density areas, 500 m x 500 m for those of medium population density and 1 km x 1 km for low population density areas. Table 1 compares the area of spatial districts to an area of 100 m, 200 m, 500 m and 1 km grids together with the number of population. Approximately 47% of the population can be directly aggregated to 100 m, 200 m or 500 m grids, thus ensuring high resolution spatial data in densely populated areas.

Table 2
Comparison between spatial districts and grids

Area of spatial districts in km^2	*% of all spatial districts*	*% of total population*	*% of national territory*	*Population / km^2*
area ≤ 0,01	11,80	10,85	0,05	20 078
0,01 < area ≤ 0,04	14,28	15,70	0,27	5 770
0,04 < area ≤ 0,25	19,40	20,85	1,90	1 095
0,25 < area ≤ 1	22,15	20,14	11,18	179
area > 1	32,36	32,46	86,60	37

21. Since there is a possibility that a grid cell of a particular size does not cover most of the spatial district area, i.e. the spatial district is divided into several grid cells, a second check was performed where the centroid of the spatial district was assigned to the grid cell which captures the majority (at least 75%) of the population of that spatial district or, where even this was not possible, a grid cell of greater size was used.

22. The applied methodology suggests that it is highly recommendable to store the census (or other) data together with spatial references of the highest possible positional accuracy if legally and technically possible. Positional accuracy can be improved when relevant spatial objects (e.g. buildings) to which most of the data are related are assigned to coordinates of their point location, and this only has to be done once. A great advantage of geocoded data transformed in this way is that they can be aggregated to an optional grid system regardless of its cartographic projection or coordinate system, but, of course, still considering the appropriate grid cell sizes. Additionally, these data also acquire all the advantages of the grid data mentioned above.

23. The location of georeferenced data in Slovenia is determined by coordinates of centroids of buildings or parts of buildings with house number(s). These centroids are kept

in the Register of Spatial Units and, due to common identifiers, the centroids can be linked to individual data from different registers. Relevant registers applicable for aggregation of data included in grids are:

(a) The Central Population Register with reliable data since 1981 regarding positional accuracy;

(b) The Slovenian Business Register;

(c) The Tax Register; and

(d) The Statistical Register of Employment.

The Register of Dwellings and the Register of Real Estate are planned to be realised in the near future.

24. Georeferenced data from listed registers can not just be aggregated to grids of optional cartographic projection or coordinate system: they also enable data aggregation to optional grid cell sizes. Considering many advantages of register-oriented statistics, the register-based census on the other hand means the loss of some important information on e.g. daily migrations, real location of working places, etc..

V. Grid data protection policy

25. Geographical location of the reporting unit is an attribute which can be used to reveal the reporting unit and its information. Therefore, the dissemination of grid data and spatial statistical data in general requires a selection of acceptable positional accuracy of the data regarding the data protection rules. Since there are no international guidelines on how to manage the data protection on grid data, countries define their own data protection policy, which may consequently aggravate cross-border projects. When establishing the national hierarchical grid system, SORS decided to define a standard set of data for dissemination on grids together with data protection rules. Spatial statistical data were also previously disseminated on grids at a user’s request but not as a standard offer of SORS, and each request has to be discussed by the Data Protection Committee.

26. Demographic statistical data were the first in the system of data sets to be disseminated on grids since they are the data most requested by users. The population data set, already disseminated at the level of settlements (settlements form municipalities, i.e. LAU 2), was a suitable ground and includes:

(a) Number of population;

(b) Sex;

(c) Five-year age groups.

27. These data are not subject to suppression due to statistical confidentiality of the data so it was decided temporarily to apply this practice to grids as well and, since the smallest settlement in Slovenia is smaller than a 100 m grid, there was no argument against this decision. This may change in 2011 as the Data Protection Committee is preparing a new data protection policy for Census 2011 data and it will also cover the dissemination of all non-census spatial statistics. Secondly, adding new attributes to the grid dissemination data set, especially in more sensitive areas such as education or income, also requires suppression of less sensitive data, e.g. numbers of population. Non-suppression of numbers results in a higher data protection threshold for attributes, but it is important to consider that spatial data users are usually more interested in areas where a particular phenomenon appears with greater frequency, so there is a tendency to set at least a small data protection threshold also for numbers.

28. There are several methods available to solve the problem of suppressed cells and, considering the spatial distribution of different phenomena in Slovenia, it was decided to give priority to revealing more data of high resolution — e.g. densely populated areas — rather than reveal more data of the total or losing the higher resolution by joining the cells. Smaller grid cells are joined together only within the existing hierarchical grid system and only when this does not create unnecessary loss of the spatial data resolution — e.g. three suppressed 100 m grid cells would only be joined with the fourth corresponding one into 200 m grid cell if the sum of the number (e.g. population) for newly joined four cells is greater than or equal to the threshold determined for revealing the attributes (e.g. education).

29. The predefined system of data sets is expected to be of great help to users as it provides in advance the information of which data are available on grids and under what conditions. Additionally, it means a smaller burden for SORS as the data are prepared in advance considering the predefined data protection rules.

VI. Conclusion

30. SORS, like many other European national statistical offices, recognised grids as an important complement to existing forms of statistical data dissemination or presentation. Grids are simply treated as a different format to traditional paper or digital tables but with their own characteristics and applicability. Grids present an applicable positional accuracy suitable for different spatial analyses or display of the data, easy handling and assure statistical confidentiality of published statistical data. Long experience with GIS in SORS and adequate development of (statistical) spatial data infrastructure in the country provide a promising basis for an upgrade of the results achieved where defining the system of data sets being disseminated on grids and grid data protection policy are considered as the two greatest challenges of the moment. Therefore, following the global trends in managing the spatial statistical data together with the directive recommendations of the Infrastructure for Spatial Information in the European Community (INSPIRE), SORS is committed to continue to participate in international discussions and projects regarding the grids or geostatistics in general (e.g. European Forum for Geostatistics). It will also strive to further promote the application of grids in Slovenia while encouraging spatial data providers to apply grids also to the process of data capture (e.g. soil, vegetation, meteorology), thus widening the national grid-based data. Development of GIS and grid-based statistics definitely offers a new dimension in understanding the mission of national statistics and the vision of wide accessibility and applicability of statistical data.

VII. References

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

Izdelava hierarhičnih mrež Slovenije – zaključno poročilo ob izvedbi projekta. Geodetski inštitut Slovenije, Ljubljana, 2008.

Oblak Flander, A.: Opportunities and Challenges of a Register-Based Census of Population and Housing – the Case in Slovenia. Seminar on Registers in Statistics – methodology and quality, Helsinki, 2007.

Tammilehto-Luode, M.: Tandem II – Towards a common geographical base for statistics across Europe. Meeting of the WP "Geographical Information System for Statistics", Luxembourg, 2003.

Wonka, E., Kaminger, I., Kuzma, I.: Regionalstatistisches grenzübergreifendes Projekt zwischen Slowenien und Österreich zum Thema Raster. Statistische Nachrichten 1, 2007