A Feasibility Study Towards a Common Geographical Base for Statistics across Europe

Marja Tammilehto-Luode
Statistics Finland, Population Statistics
Työpajakatu 13
00580 Helsinki, Finland
marja.tammilehto-luode@stat.fi

Lars Backer
Statistics Sweden, Regional Planning and Natural Resources
BOX 24 300
S-104 51 Stockholm, Sweden
lars.backer@scb.se

Philippe Guiblin
Office for National Statistics
1,Drummond Gate
SW1V 2QQ London, UK
philippe.guiblin@ons.gov.uk

1. Problem description

Regional statistics of the European Union are managed on the basis of the NUTS (Nomenclature des Unites Territoriales Statistiques) system. The system of the NUTS areas, and administrative areas in general, are far from ideal for flexible and comparable regional statistics. The system faces the problem of what has been termed as the "modifiable areal unit problem" (MAUP, for example Openshaw, 1995).

Each country has its own building blocks by which regional statistics are compiled. Administrative areas are not always the only ones. There is a strong need for a system of territorial units with relatively small populations (statistical blocks) that could be used in combination with the NUTS hierarchy.

There are two major geometrical data models. One is based on a regular, usually square, tessellation subdividing the space into cells of regular size and shape and the other is based on a tessellation with an irregular subdivision of space (Frank and Mark 1993).

This study tries to find alternatives to the NUTS system by looking at these two data models side by side in order to find an optimal system for comparable new building blocks for the geo-statistics of the European Union.

2. The objectives of the study

The aim of this study is to draft a theoretical system of relevant building blocks for displaying and analysing comparable regional statistics. The drafted system has to be usable in the contexts of Geographic Information Systems. The drafted building blocks need to be powerful in describing not only structural differences but also processes and their changes between regions (Tandem Consortium 2000).

The aim on one hand is to describe at least one system of regular tessellation (grids) that could be used across the EU and to explore its strengths and weaknesses. On the other hand the aim is to explore ways of optimising irregular tessellation (polygons) based on consistent criteria (Openshaw, Rao, 1995, Martin, 1998). The final purpose of the study is then to explore the problems in the integration of these two systems.

The study is partly funded by Eurostat. The final report will be available in October 2001.

3. The research methods and empirical tests

The study will be carried out in co-operation with the representatives and by using the data of three statistical institutes. Existing knowledge and theories concerning its application will be studied in order to build a prototype of an optimal system of grids and polygons.

Both prototypes, i.e. grid-based and polygon-based systems, will be tested using point-based and polygon-based data. The criteria for a good system are explored using empirical data and iteration processes. Integration of these two systems will be tested. A real user case is made for benchmarking. A delimitation of urban areas will be made using the prototypes.

4. Major challenges and future plans

The purpose of this study is to propose alternative solutions to complement the classical regional statistics by administrative areas. The solutions should be applicable in cross-border situations throughout the whole of Europe, where approaches to the collecting, producing, and displaying of data differ from country to country. The study explores grid-based and polygon-based input areas in order to find an optimal building block for comparable regional statistics.

The current project is a pilot study, which will hopefully serve as a basis for discourse about further development. For example, it is obvious that disclosure control should be integrated into the standardisation of spatial units for a European Geo-statistical System. It is also obvious that user needs should direct the accuracy and scale of the optimised building blocks.

By and large, the harmonisation of statistics has so far concentrated on the concepts, definitions and classifications of data. This study endeavours to widen the scope of harmonisation towards spatial comparability in terms of regional statistics.

REFERENCE

Frank, A.; Mark, D.,1993: Language issues for GIS. Geographical Information Systems. Vo lume 1: Principles. Edited D. Maguire, M. Goodchild and D. Rhind. pp. 147-163.

Martin, D., 1998: Optimizing census geography: the separation of collection and output geographies. International Journal of Geographical Information Science, 1998, VOL.12, NO.7, 673-685.

Openshaw, S.; Rao, L., 1995: Algorithms for reengineering 1991 Census geography. Environment and Planning, volume 27, pp. 425-446.

Tammilehto-Luode, M.; Backer, L.; Rogstad, L.A.; Mark, D.,2000: Grid data and area delimitation by definition towards a better European territorial statistical system. Statistical Journal of the United Nations ECE 17 (2000) 109-117.

Tandem Consortium 2000: A Management Summary. Tandem_GIS. A feasibility study towards a common geographical base for statistics across EU. Compiled by Lars Backer, Statistics Sweden. Unpublished.

RESUME

Le projet "Tandem GIS" a pour objectif de mettre en place une recherche méthodologique et pratiq ue ayant pour but la définition d'unités géographiques statistiques de base mieux appropriées à la dissemination et la comparabilité des statistiques des différents pays de l'Union Européenne.

Deux approches sont considerées en parallèle et comparées selon des critères statistiques visant à re specter une meilleure cohérence et homogénéité:

- une approche basée sur l'exploitation de données géo-referencées à un très bas niveau de d éfinition et sur l'exploitation de données aggrégées sur des grilles carrées de faibles mailles;
- une approche basée sur l'exploitation de données aggrégées sur des petites régions géograp hiques (polygones) et sur les moyens d'optimiser differents schémas d'ag