Site-Oriented Statistics and its Geoinformatic Potential

Petr Klauda[1] | *Czech Statistical Office, Prague, Czech Republic*

Abstract

A promising method used by many European statistical institutes, including the Czech Statistical Office, is the consistent orientation towards the addressed model of statistical surveys. This approach can be viewed as a way of standardization of most statistical tasks, which adds quality to the output and makes it more comprehensible and accessible for citizens. The existence of a nationwide spatial identification of addresses and buildings in the Czech Republic bears a significant potential for the application of geoinformatic technologies and their use in various areas of statistical service. The said data basis, which is unique in Europe, opens new qualitative possibilities for the collection, processing, analysis and presentation of statistical data, including the discovering of mutual connections.

Keywords

Address data, spatial data, GRID, spatial analysis, statistical georeports

JEL code

Z19

1 SPATIAL DATA AND STATISTICS

Every statistical task is based on data which can be delimited in terms of subject-matter, time and territory, thus answering three basic questions: What? When? and Where? Also every output from statistical tasks can therefore be delimited in terms of subject-matter, time and territory and the informative value depending on the form of such output is related to the answering of the above questions.

In this respect, spatial data create a certain spatial framework by means of which statistical data may be processed, analyzed and presented in the respective spatial context. This spatial framework can be based on physical real-world objects (e.g. localization of buildings and addresses, delimitation of network of streets) or abstract real-world objects (administrative boundaries, territorial, statistical and other units), or may be created artificially (GRID shapefiles).

2 ADDRESS-ORIENTED STATISTICAL DATA

At present, address-oriented collection and recording of statistical data is done by nationwide population censes, selective household surveys, agronomic censes, statistical researches in building industry and statistical registers. Therefore, the collected data belong to the area of demography, social, economic and agricultural statistics and to the area of recording of economic units, buildings and flats, including their technical and economic characteristics. Direct and exact geographical identification using the address provides for these statistics a detailed spatial dimension which can be further elaborated (Udržalová, 2007).

[1] E-mail: petr.klauda@czso.cz.

Placing statistical data in a spatial framework allows us to apply analytical methods using map algebra and geostatistics, including methods such as the interpolation method ("Kriging"), quadrant method, kernel method, etc. Furthermore, if we relate the data directly to a localized place or building, the results may be even more interesting. Information that is precisely related to a certain location has a number of advantages:

- Possibility to make calculations for any administrative, territorial, residential, statistical and other units or for GRIDs of any size;
- Possibility of a very accurate (detailed) spatial analysis and presentation;
- Possibility to select different regional levels (accuracy) when presenting aggregated data (different GRID sizes);
- Possibility to precisely select the optimum presentation level (spatial detail) in order to maximize the informative capacity of the given indicator (phenomenon) while maintaining the maximum possible degree of accuracy (detail) and ensuring the protection of personal data;
- Possibility to dynamically generate aggregated statistical information for any selection of spatial area (statistical georeports);
- Possibility to create clusters and buffer zones with given characteristics.

However, the address-oriented approach cannot be applied to all statistical data. Several basic conditions must be met, ensuring that their collection in such detailed form is possible and meaningful, including particularly the following:

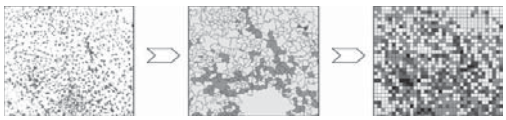
- The source of information must be identifiable in space through its address;
- Area data or sufficiently representative data must be collected (selective surveys do not meet this condition);
- The costs of an address-oriented data collection must be adequate to the value of its output;
- The observed indicator must be meaningfully interpretable also on the micro level (address);
- The use of this method must be based on detailed knowledge of the given data collection methodology;
- The spatial detail (size of the GRID cell) must be selected taking into account the nature of the given indicator, the extent and accuracy of obtained data and the necessity to secure full anonymity.

3 GRID REGULAR POLYGON SHAPEFILES

As it has already been said, when complying with the above conditions, address-oriented data may be very easily presented and analyzed using the GRID regular polygon shapefiles, which is the latest statistical phenomenon gaining its importance mainly in the Europ-wide context, where international comparability of statistical data must be ensured. GRID shapefiles are defined by geographic origin, geometric shape (squares, hexagons), surface area and the size of the elementary cell, which defines mainly the spatial resolution (accuracy) of the presented or analyzed data. The advantages of a regular polygon shapefile are very significant for statistical purposes and can be summarized as follows (Czech Statistical Office and HP s. r. o., 2010):

- Independence of the administrative structure which is changing over time and territory;
- Comparable spatial units of equal size;
- Easy anonymization of presented data;
- Possibility to hierarchize spatial presentations — different GRID sizes (macro, mezzo, micro level, etc.);
- Possibility to apply sophisticated geo-analytical methods (map algebra, geostatistics).

Figure 1 Different forms of presentation of address-oriented data (point density → administrative units → GRID)



Source: Own construction

GRID shapefiles also eliminate the basic problems and disadvantages of administrative and similarly defined spatial units, namely:

- Irregularity;
- Instability over time (changes of boundaries);
- Lack of internal homogeneity (for example, built-up areas, where most socioeconomic phenomena take place, represent only 1.7 % of the total area of the Czech Republic).

4 ADDRESS-ORIENTED DATA COLLECTION

An important prerequisite for the implementation of the above-described concept is the minimization of costs of address-oriented statistical surveys. One of the methods of cost reduction is the optimization of possible field research. The application of geoinformatic techniques and address-oriented geodata may support both logistic and qualitative aspects of these assignments. The solution of geographically oriented optimization tasks aimed at selecting a sufficiently representative sample of respondents, division of territorial districts to individual inquirers or optimization of inquirers' travels to perform field research are some of possible examples of optimization measures. Also, printed or on-line dynamic maps of the researched districts can be provided to inquirers for their better orientation. All these measures can help to achieve a better efficiency and quality of the collection, mainly by reducing administrative work and ensuring maximum productivity of the work performed by individual inquirers.

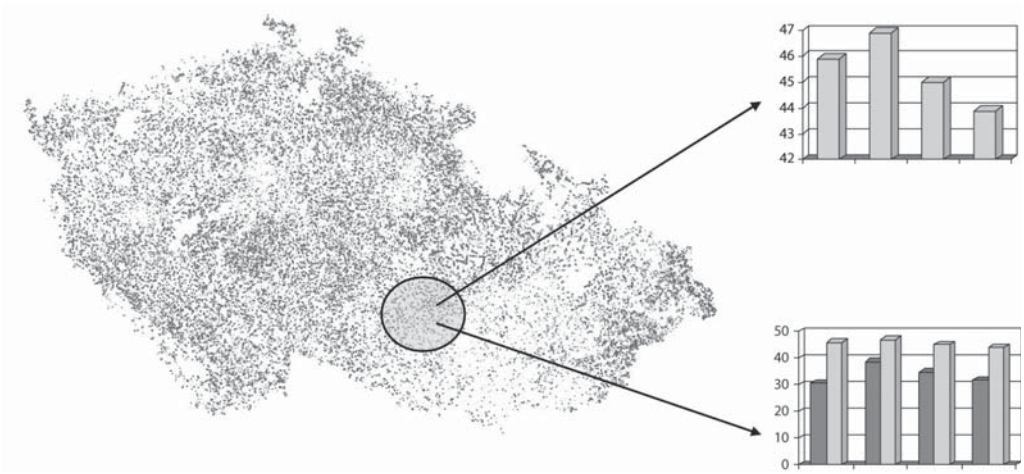
Address-oriented data collection, particularly with nationwide coverage, will always be a very time consuming and costly task. In this respect, statistical registers, being address-oriented and nationwide in their very nature, as well as other administrative data sources including the necessary data are very valuable.

5 USE OF ADDRESS-ORIENTED DATA — STATISTICAL GEOREPORTS

Site-oriented data localized in space can be used in very interesting user-focused services (statistical georeports), which can provide a statistical view on the user delimited area independently on any spatial structures (administrative or other area units etc.). These outputs are applicable in many areas and tasks. Some examples are here (Klauda, 2005):

- Geomarketing (territorial support of advertising campaigns, mapping of customers, customer and retail center network optimization, territorial analysis of the influence of competitors, revealing of geodemographic connections, etc.);
- Zoning (selection of suitable localities for investment and residential development including the analysis of their parameters and properties taking into account the demographic characteristics of the population at the given site);
- State administration and self-administration (optimization of the network of schools, bus and railway transport — analysis of commuter outflows to work and schools, modeling of spatial aspects in the labour market, etc.);
- Crisis management (flood plans, integrated rescue systems);
- Environment (modeling of the influence of emissions and radiation produced by industrial plants on their environment and the population, noise maps, radon risks in residential areas);
- Insurance services (calculation of insured amounts based on spatial aspects — houses in flood areas or highly polluted areas, classification of insured persons according to local conditions, etc.).

Figure 2 Statistical georeports — statistical data aggregations according to spatial selection



Source: Own construction

CONCLUSION

The Czech Statistical Office works with spatial data in selected fields of applied statistics, in databases and statistical registers on a long-term basis. Since 1999 it has been developing its own geographic activities in the field of collection, processing and presentation of data in an efficient cooperation with the Ministry of the Interior and the Czech Geodetic and Cadastral Office. The Czech Statistical Office contributes to the creation and implementation of a national geoinformatic infrastructure, increases awareness and facilitates access to geodata and geoinformation for all uses and gathers feedback.

Thanks to that, it has a wide database of statistical and spatial data, including the basic technological equipment and important know-how, allowing it to manipulate with address-oriented data also in spatial representation. However, it does not have a sufficient geoinformatic infrastructure necessary for standard implementation in statistical assignments and outputs. This situation should change radically if we succeed in creating such infrastructure including a statistical geoportal developed in accordance with Directive 2007/2/EC of the European Parliament and of the Council establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). This project should also fulfill the vision of gradual transformation of the existing forms of statistics to site-oriented statistics (geo-statistics) and the related offer of new user services.

References

CZECH STATISTICAL OFFICE and HP s. r. o. *Studie proveditelnosti: StatGIS, Fáze II: Hlavní studie (Feasibility Study: StatGIS, Phase II: Main Study).* Czech Statistical Office and HP s. r. o.: Prague, 2010.

KLAUDA, P. *Prostorové systémy pro podporu rozhodování* [Diplomová práce] *(Spatial systems to support decision-making,* Master's Thesis*).* Prague, 2005.

UDRŽALOVÁ, Z. Adresní místa. (Address Sites). In *ArcRevue* 2/2007, pp. 5–7.