PROBLEMS AND POTENTIAL SOLUTIONS FOR THE IMPLEMENTATION OF GIS WITHIN THE BULGARIAN STATISTICAL SYSTEM

Alexander Kotsev*
Stelian Dimitrov**

* Regional Statistics Division
National Statistical Institute of the Republic of Bulgaria
2, Panayot Volov Str.
Sofia 1504
Bulgaria
akotsev@gea.uni-sofia.bg, akotzev@nsi.bg

** Sofia University "St. Kliment Ohridski"
Faculty of Geology and Geography
Cartography and GIS Department
15, Tzar Osvoboditel Blvd.
Sofia 1504
Bulgaria
stelian@gea.uni-sofia.bg

ABSTRACT

Geographic information systems (GIS) are a very important part of any national statistical system. Their integration within the system multiplies the value of data and poses new analytical perspectives, especially for the regional and urban statistical divisions. Since geoinformation solutions are very expensive and time consuming to develop, it is very important that the existing problems and obstacles concerning their development are correctly identified and solved. The purpose of this paper is to make an attempt to shortlist the major issues concerning the efficient and practical use of GIS solutions within the Bulgarian statistical system. We consider most of the obstacles mentioned here common to most statistical systems in South-East Europe.

INTRODUCTION

Accurate, up-to-date and reliable information is vital for the activities, concerning management and planning of a region's human and natural resources and for dealing with regional development decisions that have a spatial context (Klosterman, 1995). A comprehensive information base could reduce uncertainty and improve decision-making process in order to make it more effective.

In transitional countries, such as Bulgaria, the data barriers are still obvious due to both institutional and technical reasons. As institutional issues are being recognized and governments start to invest millions of euro in collecting data, the data management and usage are still far from satisfactory level. Information on various aspects of regional development – social, economic and environmental – is originally collected for different purposes, at different scales, at different time frames and with different underlying assumptions about the nature of the phenomena. This creates technical difficulties for the integration of social and environmental data, and explains why most of the plans for regional development, emerging from the process of EU integration, are not more than whish lists.

In recent years, Geographic Information Systems have become an important tool for regional and urban research. As about 80-90% of data collected and used for regional and environmental information systems are related to geography (Huxhold, 1991). Up-to-date and reliable information is vital for GIS, and it provides an integrated computing environment for social and environmental data integration. It is widely recognized that GIS provide a large range of analytical capabilities to operate on topological relationships or spatial aspects of geographical data, on non-spatial attributes of such data, or on non-spatial and spatial attributes combined. GIS facilitate the integration of disparate data sets, creation of new and derivative data sets, and development and analysis of spatially explicit variables. Furthermore, the integration of GIS with spatial statistical analysis has the potential to become a powerful analytical toolbox, enabling regional and social scientists to gain fundamental insight into the nature of spatial structures of regional development (Brown, 1996).

APPLICATION OF GIS FOR STATISTICAL PURPOSES

GIS are considered to be very important for use within the statistical institutions worldwide. Their ability to integrate data in spatial context multiplies the value of information. By using GIS data are referenced to a low territorial level. GIS are also frequently used for derivation of spatially relevant variables. The functionality of a GIS generally varies from simple visualization of data (mapping) to complex spatial analyses and spatial modelling. Some of the classical applications of GIS are listed below.

Data visualization

Representation of territorially referenced data through thematic maps is considered to be very efficient. Visualization is an important component of any effort to understand, analyze or explain the distribution of phenomena on the surface of the earth, and will become increasingly important as volumes of digital spatial data become more unmanageable (Buttenfiled and Mackaness, 1991).

Maps are typically created within the statistical offices to illustrate the distribution of phenomena over a given territory. Visual representation of data, by using maps is easy

to understand, and helps users in the process of revealing differences and similarities across space. According to Robinson (1953) for most of the people by far the largest share of knowledge of wide-area geographical relationships results from looking at prepared maps.

Geographic information systems are frequently incorrectly referred to as a set of tools for digital mapping. Their potential spreads far beyond simple mapping. According to Maguire (1991) it is the ability to analyze geographical patterns and relationships which differentiates GIS from computer cartography. The power of GIS comes from the integration of various data in territorial context and their ability to 'spatially' analyze data. Visualization in GIS is only the most suitable way to represent geographic data. It is arguable whether more attention needs to be paid on visualization research and the principles of good map design (Buttenfiled and Mackaness, 1991) in the context of a GIS. GIS software products are anyway quite frequently used for simple map creation.

Current research on visualization in spatial data sets is focused on the need for visualization tools for higher-dimensional spatial data sets (Fotheringham, 1999), since most of the contemporary spatial data sets have many attributes and it is inadequate to use simple univariate or bivariate visualization. Readability is the major concern for using more than two dimensions when visualizing complex spatial databases.

Territorial Referencing

One of the main advantages of GIS technology is that it allows the referencing of data to a very low territorial level. The possible spatial level of detail only depends on the scale and detail of the input map source and attribute (tabular) data being used.

Digital spatial data is nowadays readily available, both in raster and vector format, though issues of copyright and access still exist (Bond, 2001). Confidentiality issues also arise when going to a very low territorial level. They will be discussed further later in this paper.

Attempts are now made within the Bulgarian NSI for the development of two interrelated registers – one addressing the changes that take place in the administrative breakdown of the country and another – maintaining addresses and buildings. Both registers are urgently needed since changes in the listed areas are very dynamic and changes have to be appropriately tracked. GIS are one of the main tools that need to be incorporated during the development stage of both registers. On the other hand having standardized information about settlements and street addresses will ease classic geoinformation procedures like geocoding and will lead to enrichment of the geoinformation capacity of the institution.

Data Derivation

Spatial objects, contained in a database can be queried and contemporary GIS software products allow the extraction of synthesized quantitative characteristics of those objects. Lengths, areas, distances, etc., are frequently derived from spatial databases to characterize the distribution of certain phenomena over space.

Attempts are now made within the Bulgarian NSI for obtaining the polygonal boundaries of settlements (NUTS5) from the Bulgarian cadastral agency. Once received, they will become part of the settlement register that is now in the process of being developed. By integrating spatial data about settlements within a register, areas and

perimeters of territorial units can be derived for any given moment, covered by the history of the register, starting from NUTS5 bottom up.

Spatial Statistics

GIS and spatial statistical tools are widely used during the variable estimation process. The set of tools known as exploratory spatial data analyses (ESDA, EDA) are becoming increasingly important. According to Anselin (1988, 2000) ESDA are 'the collection of techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models.' Fotheringham (2000) defines ESDA as '… a set of techniques to explore data in order to suggest hypothesis or to examine the existence of outliers.' Some researchers question the quality and meaningfulness of the output of the ESDA process, while others consider ESDA as being very useful during the estimation process (Fotheringham, Anselin).

Good examples of ESDA tools used during the variable estimation process can be found in the documentation of the Eurostat project for comparative urban statistics Urban Audit II (http://europa.eu.int/comm/regional_policy/urban2/urban/audit/).

Spatial Analysis

The true potential value of Geographical Information Systems lies in their ability to analyze spatial data using the techniques of spatial analysis (Goodchild, 1988). Spatial analyses are very important when characterizing complex systems such as the regional and urban ones. Social, economic and environmental characteristics of the territory are modeled during the spatial analysis process. The term spatial analysis is according to Bernhardsen (1999) a set of computer-based techniques, applied to analyze patterns, connections, and possibly, causes of variation over spatial data. Spatial analysis techniques are also frequently used for variable estimation.

PROBLEMS FOR THE IMPLEMENTATION OF GIS

Significant reforms have been taking place in Bulgaria since the beginning of the transition period from planned to market economy. Lack of funds has been a major concern for the past fifteen years. This has had an inevitable impact on the national statistical system. Attempts are made so that expenditures are reduced and the system tuned to the European standards. One of the main objectives of the reform is to produce better data, adequately reflecting the user needs. GIS solutions have the potential to become a major source of revenue for the public sector, including the National Statistical Institute as the main data producer in the country. Detailed cost-benefit analysis for the implementation of GIS must take place, since GIS solutions take sufficient time to develop and are considered to be very expensive. The need of spatial information must be studied so that the user requirements can be met.

GIS in the National Statistical Institute – State of the Art

Experts from the National statistical office work on several GIS-related projects. A section within the population department at the Bulgarian NSI has been established especially to deal with geoinformation matters. Data from the 2001 Population and Housing Census are territorially referenced within the Population, settlements and GIS department. The lowest spatial level being used for this is buildings. Statistical confidentiality issues forbid the dissemination of those territorially referenced data.

The use of GIS within the National Statistical Institute generally varies from population related activities and estimation of environmental indicators to simple desktop mapping. The lack of coordination between the different units and with other relevant institutions is unfortunately noticeable.

The outcome of the integration of geoinformation technologies with classical statistical tools is relatively poor. Users can not benefit from the spatial data that is produced by the statistical institute. Geographic data are practically not being disseminated and are primarily used for internal purposes.

PROBLEMATIC AREAS

The existing problems can be separated into two types: internal and external. Most of the listed issues are interrelated and their solution requires a complex policy vision that has to emerge within the Bulgarian public sector.

Awareness of GIS

Geographic information systems can be considered to be relatively new to transitional countries like Bulgaria. Potential users are not fully aware of the functionality of GIS that they can benefit from. Organizations with potential to become major spatial data producers like the NSI need to be involved in systematic initiatives aiming to raise awareness of the application of geoinformation solutions in different spheres of social life.

Irrelevant Legislation

Probably, the biggest problem for implementing GIS not only within National Statistical Institute, but in the country as whole, is the absence of appropriate normative and legal framework, regarding spatial data, the rules for dealing, distributing and processing such data.

Because of the inertia from the past, map standards of the Warsaw pact are still used in the country, and there are many restrictions for using and producing geographical data. For example the only cartographic projection, which might be used for civil purposes, is called “Projection 1970”, which parameters are still military secret according to the outdated Bulgarian legislation.

Statistical Confidentiality

According to Bond (2000) and other the ideal world GIS would be the one that contains single records that are adequately territorially referenced. The Law on statistics, adopted by the Bulgarian Parliament forbids the dissemination of individual records. The text defining data confidentiality and protection of secret in the Law on statistics has to be improved, since it might be pointless to protect individual data for hospitals, polyclinics, kindergartens, schools, etc.

The statistical confidentiality problem is common to most countries and is considered to be a serious obstacle towards the development of GIS since aggregated data become abstract in terms of spatial reference.

Institutional Capacity

Experts, along with software, hardware and data are an inseparable part of any contemporary GIS. A problem, typical to the country is related to the education and training of the GIS specialists. There is an obvious lack of capacity in the institutions, including the statistical institute to work with contemporary geoinformation systems, and the main reason for this is the training and education. There is only one MSc degree programme in the country (at Sofia University “St. Kliment Ohridski”), and only a few

offered training programs. This is resulting in shortage of qualified specialists and the attempts for self training are still the most preferred form in the Bulgarian institutions.

Spatial Data Infrastructure

The National spatial data infrastructure (NSDI) encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data.

The NSDI in Bulgaria is relatively weak. There is a noticeable lack of coordination in the production and use of data in the country and the region in general. Several conferences and workshops addressing spatial data infrastructure issues took place during the past years without serious outcome. No efforts have been made for the coordinated data production within the country and the region as a whole.

An initiative of the Joint Research Center of the European Commission - the INSPIRE (Infrastructure for Spatial Information in Europe, http://inspire.jrc.it/) addresses matters of geographical data infrastructures across 32 European countries. The major problems outlined in the INSPIRE report for the state of the art of the Bulgarian spatial data infrastructure are the very fragmented GIS sector and the absence of a single national mapping agency in the country.

Software Products and Data Formats

It is clear that there is an important link between the way in which data are structured in a geographical database and the type of functions that can best be employed. Furthermore, restructuring geographical data is time consuming and error prone (Maguire and Dangermond, 1991). Vector data which are the main one used within the statistical offices need to be appropriately structured, so that they are suitable for further analytical procedures. Very importantly, a topological model has to be applied for all spatial objects, so that geometric relations found in the real world are followed accurately and errors are discovered and corrected once after they have occurred.

Harmonization and standardization of data as produced by multiple producers is an urgent issue. Interoperability of the GIS software and the used data formats have been subject to worldwide discussions during the past few years. It is very important when developing time-consuming and expensive systems like GIS. Regulative bodies working towards the standardization of spatial data have been established on regional and global scale. The main organization involved in interoperability activities worldwide is currently the Open GIS Consortium (www.opengis.org), which is a member-driven, non-profit international association that is involved in the development of geoprocessing interoperability computing standards. Very few Bulgarian organizations are unfortunately members of the Open GIS Consortium.

Old formats of spatial data still exist within the country's infrastructure, particularly within the cadastre. Parts of the Bulgarian cadastral system are contained in simple ASCII text files (ZEM files), containing coordinates and attribute values. Specific software is developed for manipulating those outdated files. No contemporary software product like ESRI's Arc/Info of Geomedia's Intergraph can handle these spatial data sources. Problems like this complicate the normal exchange of information between governmental agencies.

Metadata

Metadata is crucial for the normal building up of a countrywide geographic data infrastructure. There is no metadata standard for spatial data being created in Bulgaria. This fact seriously affects data users who can not smoothly exchange and use spatial

information. National standards have to be adopted and their use must be obligatory. The provision of metadata must be well structured and widely accessible. It must also be centrally coordinated so that users can benefit from it to the maximum extent.

An appropriate example of metadata standard is the one developed by the United States Federal Geographic Data Committee (FGDC, http://www.fgdc.gov). Similar standards need to be adopted in Bulgaria because this is the first step towards improving the Bulgarian spatial data infrastructure.

The statistical institute as an organization involved in spatial referencing, by integrating various data must be one of the first to initiate the building up of a countrywide metadata standard.

CONCLUSIONS

As conclusion we can clearly define the following statements:

1. There is a visible lack of capacity within the Bulgarian institutions to implement the geoinformation technologies.

2. The reasons for this are related not only with the reform processes of Bulgarian governmental and public administration, but also with the inertia from the Socialist past, inappropriate legal and normative framework and the lack of educational and training capacity.

3. The implement of GIS not only for the purposes of the National Statistics, is inevitable and the proof for this are the various projects within the country, mainly funded by EU programs.

4. The lack of coordination was outlined above as the major problem in front of the GIS sector in Bulgaria. That inevitably reflects the National statistical institute as an institution incapable of producing own spatial and attribute data itself. A central body independent of Government should be set up to set up a focus and forum for common interest groups in the geographical information area, undertake promotional activities and review progress and submit proposals for developing national policy. Its members should be from all interested groups and it should maintain strong links with the Government (Chorley and Buxton, 1991).

5. A regulative unit in the National Statistical Institute has to be set up to coordinate the efforts of different divisions and to manage the smooth communication of the institution with third party organizations.

REFERENCES

1. Anselin, L. (1999):'Spatial Econometrics'
2. Bernhardsen, T. (1999):'Geographic Information Systems'. John Wiley & Sons, Inc. USA
3. Brown, D. G., (1996): 'Spatial statistics and GIS applied to internal migration in Rwanda, Central Africa'. Practical Handbook of Spatial Statistics, Arlinghaus S. L. (Eds.). New York: CRC.
4. Buttenfield, B., W. Mackaness (1991):'Visualization' In: Maguire D. J., Goodchild M. F., Rhind D W, Geographical Information Systems: principles and applications. Longman, London
5. Bond, D. (2000):'GIS and Spatial Analysis in Regional and Urban Research' Cities and Regions GIS Special, pp. 1-8

6. Chorley, R., R. Buxton, (1991): ’The Government Settings of GIS in the United Kingdom’ In: Maguire D. J., Goodchild M. F., Rhind D. W., Geographical Information Systems: principles and applications. Longman, London
7. Fotheringham, A., et al. (2000):’Quantitative Geography – Perspectives on Spatial Data Analysis’. Oxford University Press Inc.
8. Huxhold W. E., (1991): ‘Introduction to Urban Geographic Information Systems’. New York: Oxford University Press Inc.
9. Klosterman R. E., (1995): ‘The appropriateness of geographic information systems for regional planning in the developing world’. In: Computer, Environment and Urban Systems, vol. 19, No. 1, pp. 1-13.
10. Maguire, D., J. Dangermond, (1991):’The Functionality of GIS’ In: Maguire D. J., Goodchild M. F., Rhind D. W., Geographical Information Systems: principles and applications. Longman, London
11. Rhind D., N. Green, H. Mounsey, J. Wiggins, (1984): ’The Integration of Geographical Data’
12. Robinson A. H. et al., (1953):’Elements of Cartography’. John Wiley & Sons, Inc. USA
13. Vandenbroucke, D., P. Beusen, (2003):’Spatial Data Infrastructures in Bulgaria - State of Play’