Spatial Component in Business Intelligence System for Advanced Threat and Risk Analysis

Mirza Ponjavic and Almir Karabegovic

Abstract. This paper shows an innovative approach for implementation business intelligence systems in advanced threat and risk analysis using spatial component. It demonstrates how to improve intelligence of complete information system by involving spatial extension. Most of business data in data warehouses are often spatial per se, and without using this component, analysis missing very important dimension of the data nature. From other side, frequent problem in enterprise data warehouse is creating relations between tables which come from different sources and without any common attributes; that could be very easily solved by spatial relations. This paradigm of spatialization assumes changing overall system architecture, from data storage, via retrieving to its presentation mechanism. Particular benefit of this approach for threat and risk analysis is effective utilization of location data, advanced spatial analysis techniques and more variety in data visualization. Examples of organizations which need such system are intelligence agencies, emergence services or epidemiology centers.

1 Introduction

Business Intelligence (BI) mainly considers computer-based techniques to support better business decision-making [1]. It uses operations for identifying, extracting, and analyzing business data and offers functions of online analytical processing (OLAP), analytics, data mining, predictive analytics and reporting [2][3]. Business Intelligence data and analysis has more and more importance for business development, but while most of business data has location as a component, few businesses take full advantage of spatial and location analysis.

Location can be described by an address, a geographic region or a tracking route, that can be presented, managed and analyzed interactively in a GIS [4]. Recognized spatial relationships, patterns and trends can answer the sophisticated

Mirza Ponjavic
University of Sarajevo, Faculty of Civil Engineering, Department of Geodesy, Bosnia and Herzegovina

Almir Karabegovic
University of Sarajevo, Faculty of Electrical Engineering, Department for Computer Science and Informatics, Bosnia and Herzegovina

M. Mach-Król and T. Pełech-Pilichowski (eds.), *Advances in Business ICT*, Advances in Intelligent Systems and Computing 257,
DOI: 10.1007/978-3-319-03677-9_7,

questions related to hardly visible and invisible laws applied between phenomena described by specific business data sets [5]. Location awareness can be incorporated into Business Intelligence Systems (BIS) and used in field of risk assessment and risk management extracting the maximum value from GIS and BI integration and traditional and spatial data sets.

2 Spatial Component Improves Business Intelligence of Information Systems

Spatial component improves business intelligence of information systems and brings easy-to-understand visualization to business applications. Most of business data in data warehouses are often spatial per se, and without using this component, analysis missing very important dimension of the data nature. From other side, frequent problem in enterprise data warehouse is creating relations between tables which come from different sources and without any common attributes; that could be very easily solved by spatial relations [6]. This paradigm of spatialization assumes changing overall system architecture, from data storage, via retrieving to its presentation mechanism.

Examples of organizations which need such system are risk analysis centers (RAC), various intelligence agencies, emergence services or epidemiology centers. This paper is focused on the implementation of this system for the purpose of improving the center for the analysis of threats and risks, although the experience gained in its implementation may be fully applied in other domains mentioned. Threat and risk analysis is the process used to obtain quantitative or qualitative measures of risk levels and has focus on quantifying the probability of negative consequences from one or more identified or unknown threat causes.

3 Architecture of Business Intelligence System with Spatial Extension for Risk Analysis

An overview of key architectural components of Business Intelligence System (BIS) with spatial extension for RAC project is shown in Fig. 1. Data warehouse (DWH) is intended to store a large amount of data and considers load strategy involving: extracting data from data sources (operational systems), moving it into data warehouse structures, structuring the data for analysis purposes, and moving it into reporting structures (data marts). The architecture includes the process required to handle and manage the following daily operations: data acquisition, data buffering, transformation and loading data into data staging area within extract, transform and load (ETL) management processes [2].

Specific data from each information service (data source) can be masked and propagated into the appropriate data marts, which are subsets of the data in RAC data warehouse. The data marts contain aggregated (summary) data from heterogeneous information services (e.g. Oracle database, MS SQL Server, MySQL, Excel) at the particular level of hierarchy. In this way, it provides more effective data structuring and eliminates the need to aggregate data when executing a query or analysis by end users. This leads to better performance and avoids the redundancy of data.

Data warehouse is a central source of consolidated and transformed current and historical data used by various professionals (analysts) for business analysis, data exploration and decision support [2]. This database can be accessed directly from an application level, which consists of GIS analytical machine (thick client) and the Report Server (BI Server). They are used by RAC analysts who have access to all data in read-only mode in order to implement the analytical operations.

GIS analytical machine is a generic set of analytical tools integrated through dedicated applications that are customized in terms of localization, workflow and typical data sets used for analysis. Also, in addition to direct access and read data from the data warehouse, they allow the entry of data which come from other sources, e.g. news media (Open Source Data) through the available forms. GIS tools provides spatial data analysis that can be conducted in order to indicate and analyse of spatial trends [7].

The analysis results are presented in corresponding form by the report generator (BI Server). Reports are available to all analysts and other authorized users of the system, and the representation of spatial phenomena can be achieved by using integrated components of spatial engine, application server and viewer (e.g. Oracle Spatial, WebLogic, MapViewer) consolidating results with a set of base geospatial data in the background.

4 Case Study: Location Intelligence System for Risk Analysis Center

4.1 The Aim of the Project

Building a functional and efficient integrated system for border management at the national and international level implies the establishment of the Center for Risk Analysis as its key parts. Bosnia and Herzegovina conducts activities related to these issues along with other duties essential to the process of liberalization of visa regime with the European Union. The concept of Integrated Border Management (IBM) involves coordination and cooperation of all state agencies and bodies involved in cross-border activities, in order to ensure maximum efficiency and effectiveness of border management.

The general objective is to create a safe and reliable system framework for risk assessment and spatial risk analysis which includes establishment of: basic functionality of the Center for Risk Analysis (CAR), communication channels and protocols for data exchange with professional services, or information services (IS) which supply the center with the source data, and functional information services or agencies and bodies for the collection, primary data processing and propagating it towards the CAR.

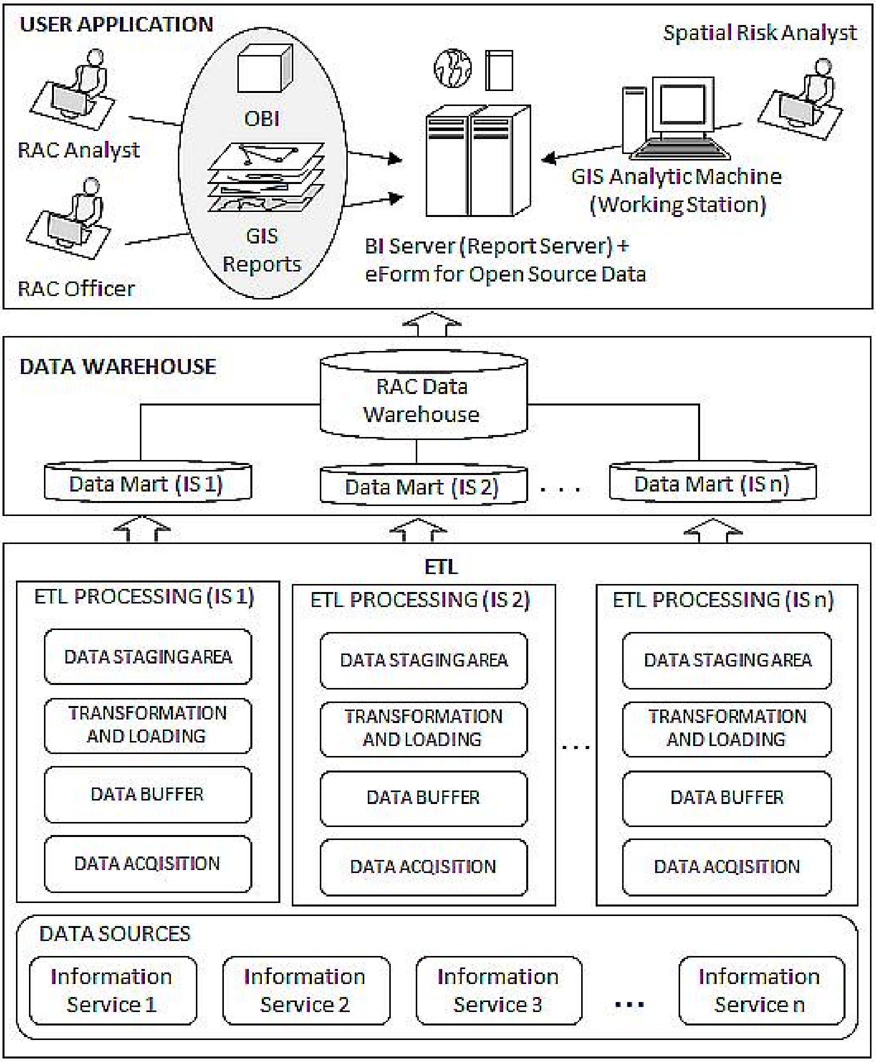


Fig. 1. Architecture of the business intelligence system with spatial extension for RAC

4.2 The System Implementation

System design is based on a central database to which data are propagated from information sources responsible for collecting, selecting and masking (protected) data relevant for risk analysis (Fig. 1). These data are prepared and transformed in a way that enable the production of quality reports of potential threats and risks that could endanger human lives, material goods or social order. Information Services (IS) are the various state institutions, agencies, professional services or local centers for risk analysis with its own organizational, informational, technological and technical characteristics. Integration of the spatial risk analysis system is directly related and dependent on the functionality of individual IS components. It is implemented through the phases of defining the scope, development, testing and deployment of this system.

The phase of defining the scope includes the following activities: information collection, analysis of the existing state and system design. During this phase, the final architecture is defined as logical and physical foundation on which to build the entire system of risk analysis. One of the major goals at this stage is surely evaluation of the integration of agencies (information services) in the proposed architecture (Fig. 1), in terms of compatibility of the existing systems with the future architecture considering equipment, databases, communications, and the specific needs of each agency.

During this phase, it is developed logical and physical architecture of the business intelligence system (BIS) as a configuration map data sources needed to finally form a single repository. After defining the logical configuration, it is defined the data flow from the source to the application level, as well as the general data access. Part of data propagated towards the center for risk analysis are masked by using special algorithms for encryption. Decryption key for unmasking the data should have only agencies that own ID data for conducting operational tasks.

Therefore, based on analysis of existing systems in some agencies with the specification of the available equipment, and other verified operational resources, as well as their operational and analytical needs is defined:

- logical system architecture,
- ata flow from the source to the user level,
- data sources (existing systems) through relational diagrams,
- the way of masking certain protected data before sending it to the CAR and
- other key elements for building the system.

Finally, the detailed design of all procedures and processes for the project is documented at the end of the design phase of the system. In this sense, through the final design solution is necessary to consider and address the following questions:

- the amount of data in a central database,
- extract / transform / load (ETL) of the data,
- data warehouse (DWH) design / database / collection process and design data mart's including: defining the dimensional model with the presentation in the form of diagrams, the convention on the level of object names, relationships between objects, the logical level of metadata and define the factual and shared database tables,
- data manipulation: security, access, updating, refreshing, replication, editing, archiving, backup, disaster recovery of the data,
- system management,
- testing the system and training of operating personnel and
- move into production mode, and system maintenance.

The development phase includes activities of development all the components that make up a complete system for risk analysis, their documentation, development of test scenarios and test data and training materials preparation. The testing phase includes activities of checking the readiness for testing and conducting of the final tests.

The deployment phase will be conducted when:

- all necessary equipment, software and communication network is installed and ready for use, and when the test environment is set up,
- programs and scripts for retrieval, transformation, load and refresh the original data in DWH are developed and tested individually in a test environment with appropriate examples of data,
- measuring loads in the process of loading data are executed,
- sample report, together with some ad-hoc queries are developed in the test environment and when the validity of results is verified and
- policy of access to data in the central database is established.

After the programs and scripts are developed and tested, it is performed the final inspection and testing of network resources and equipment, and overall functionality of the integrated system (before switching the system into production mode with real data). The transition from the test environment to production mode, means that production central database (data warehouse) is created and the processes of extracting, transforming and cleansing data is made in real systems with real data. At this stage, the development team and the operational staff conducts an initial uploading and procedure of the first data refreshing in the risk analysis system.

4.3 The System Using: Business Intelligence with Spatial Application in Risk Analysis of Epidemic Infectious Disease

For a number of analyses types conducted at the Center for Risk Analysis (e.g., trafficking, tracking shipments of plant origin...) there is a need for spatial presentation, or for the use of spatial analytical techniques. Therefore, the data propagated from the information services (or agencies) are geocoded and referenced in the spatial domain. Analytical capabilities directly dependent on available data, but also the level of detail displayed. At least, a phenomena or its trend can be displayed to the level of settlements, streets or border crossings. This is enabled by means of incorporated background (base) map data for the area of Bosnia and Herzegovina and the wider region. These data include: basic cartographic detail (cities, roads networks, ports, airports, railway network, administrative boundaries, border crossings) (Fig. 2), environmental data (climate, land use, precipitation, soil, forests, DTM, water bodies), utilities (power supply, telecommunication), demographic data and descriptive statistics (population, education, employment, agricultural yields...), economic data (administration bodies and development indices), and also Web services offering background geodata (satellite images, street maps).

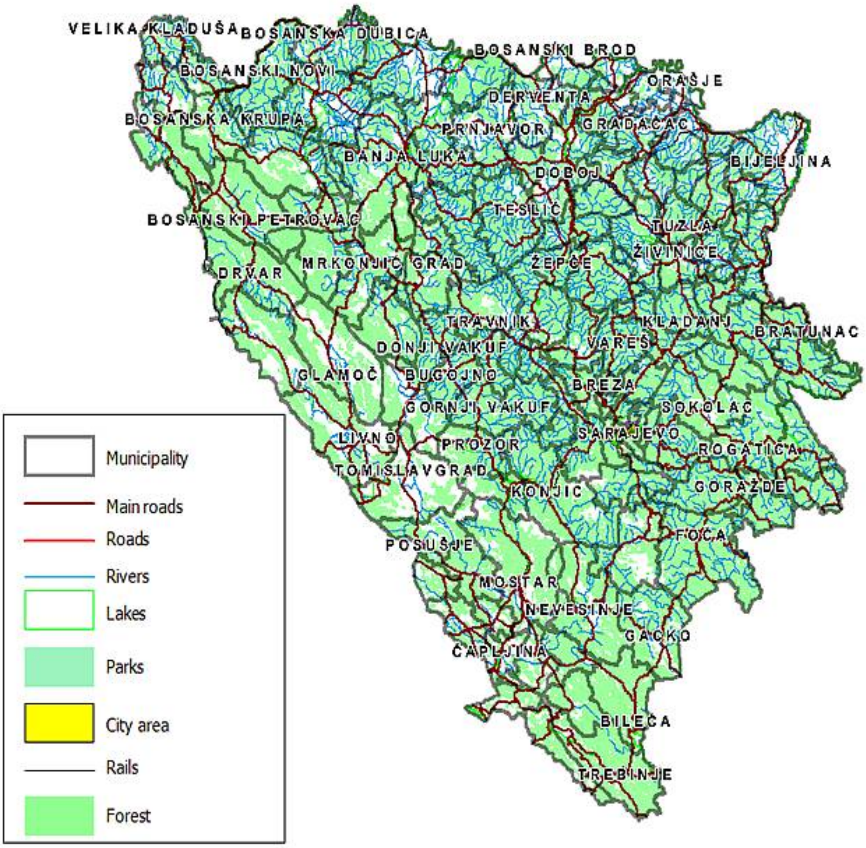


Fig. 2. Background geospatial data: basic cartographic details

User application environment, which includes the integration of GIS, Business Intelligence with Map Viewer (web application) (Fig. 3) covers a broad analytical functionality (geospatial analysis, business intelligence, reporting, publishing, creating a dashboard), which allows implementation of various types spatial data analysis from routine data presentation (eg. inspection of entries and exits at border crossings as shown in Fig. 4) to complex exploration techniques and concepts (geostatistics, clustering, gridding, network analysis, spatial queries, spatial data mining...).

The methodology of these analytical concepts is independent of the architecture and implementation of the system [8] that gives full freedom of analysts to create different models and scenarios during the analytical process. This allows the application of different methodological approaches that may include iterative stages such as framing the question, formulating the approach for addressing the problem, data acquisition, selecting appropriate methods and tools for analysis, and delivering the results and conclusions [9].

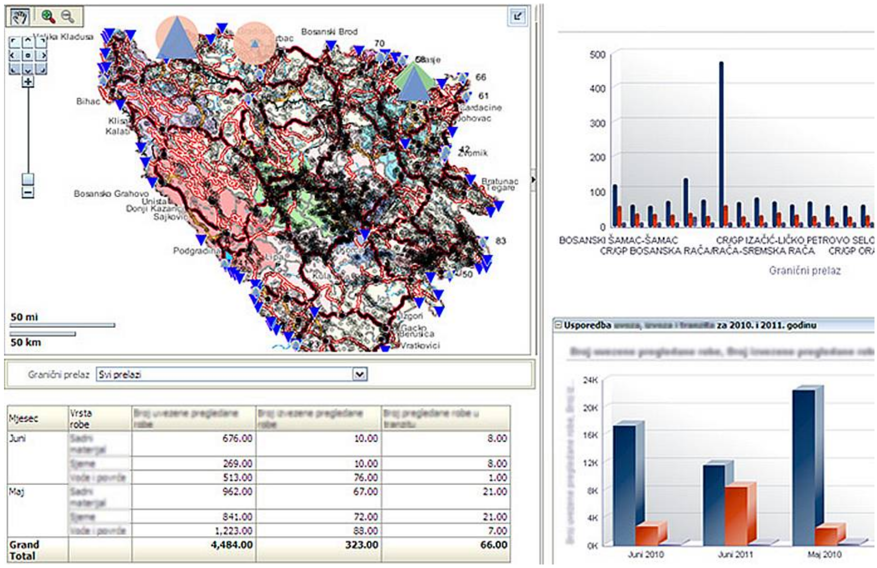


Fig. 3. Business intelligence dashboard for outlier detection

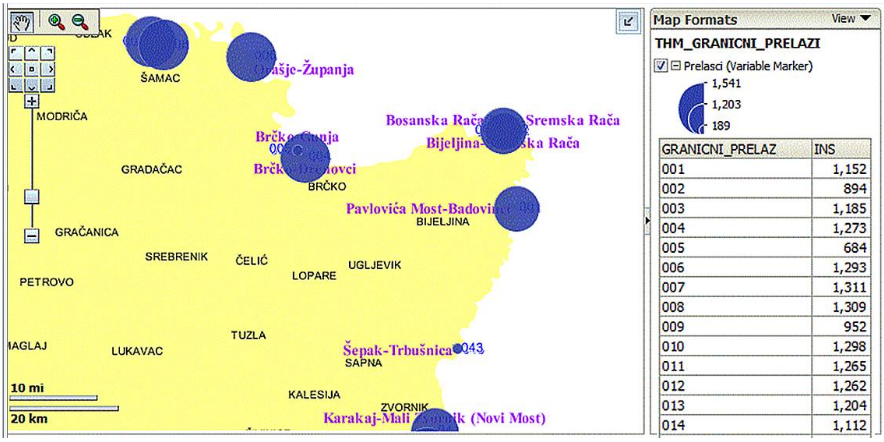


Fig. 4. Inspection of entries and exits at border crossings

One of the test scenarios used to conduct the analytical process and generate reports from the BI system is risk analysis of epidemic infectious disease to monitor the appearance of brucellosis caused by Brucella melitensis bacteria.

It is transmitted by ingesting infected food, direct contact with an infected animal, or inhalation of aerosols. Brucellosis primarily occurs through occupational exposure (e.g. exposure to sheep), but also by consumption of unpasteurized milk products.

This phenomenon is identified on the basis of information received from public health services. The task (set in this scenario) is to determine:

- spatial foci (hot spots) of disease,
- spatial trend of expansion,
- source and cause of the phenomenon
- measures to control epidemics and future prevention.

In short, this analytical process is conducted through five methodological steps [9].

The first step is the entering of external data (on the phenomenon with disease indications and number of patients) in the system. The phenomenon is registered at the locations (marked by symbol) that are found through address system search engine (based on the known address of infected individuals) (Fig. 5). Foci of disease are generated using the hot spot and cluster analysis.

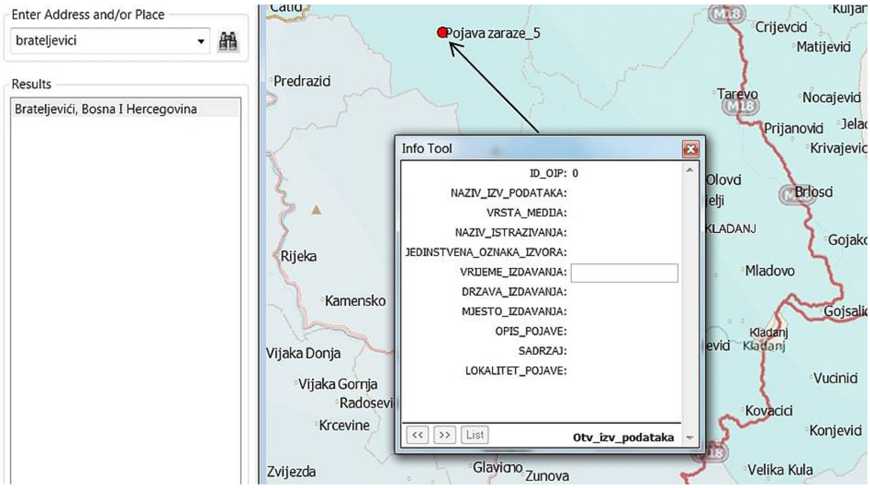


Fig. 5. Entering of external data with location of disease phenomenon

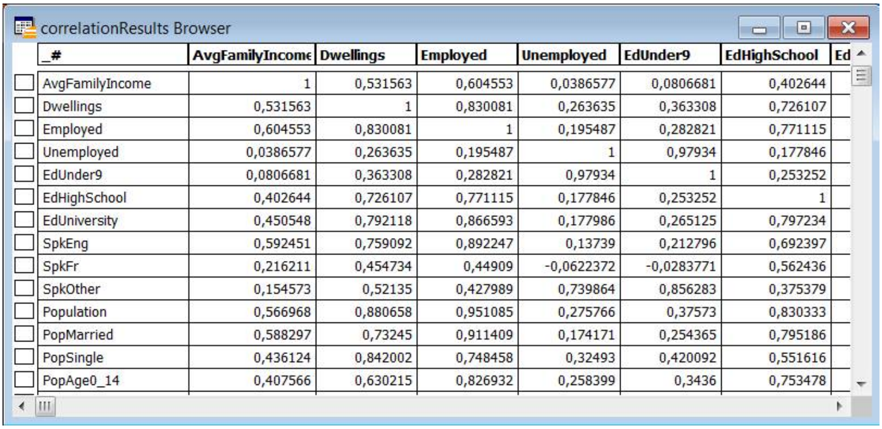
The second step is the selection of analytical method based on the entered data on the phenomenon and the available data sets in DWH (propagated from information services). The proposed method here is spatial autocorrelation, which examines the relationship between numerical grids. The matrix obtained by the method (Fig. 6) allows the investigation of correlation levels between groups of numerical grids (i.e. sets of attribute data from different information sources, e.g. records of foreigner residence registration, or tracking animals and plants shipments...). The elements of this matrix are the coefficients with values ranging from -1 to +1, where values close to 1 describes the attributes with a high degree of correlation (spatial dependence), and those that tend to 0 describe pairs of attributes with a low degree of correlation (spatial independence) [9].

The third step is the implementation of the analytical process, where the matrix analysis identified the following spatial relationships:

- there was increased presence of male foreigners with primary education (nomadic cattle breeder) and temporary residence nearby the hot spots for the previous period and
- there was migration of sheep herds nearby the sites of infection and a number of their border crossings.

The higher value of correlation coefficient does not imply a cause of the phenomenon, but it just says that there is a spatial relationship between certain spatial phenomena and the next task is to use other methods of analysis to determine the actual cause of infection [10].

The fourth step is spatial reasoning and inference on the cause of the phenomena [11]. On the basis of detailed exploratory spatial analysis which included: mapping the spread of infection, buffering, overlapping with the road network and analyzing the data records of border crossings (Fig. 7) it is concluded that the disease is caused by consuming dairy products and that is transmitted through infected sheep.


correlationResults Browser

_#	AvgFamilyIncome	Dwellings	Employed	Unemployed	EdUnder9	EdHighSchool	Ed
AvgFamilyIncome	1	0,531563	0,604553	0,0386577	0,0806681	0,402644	
Dwellings	0,531563	1	0,830081	0,263635	0,363308	0,726107	
Employed	0,604553	0,830081	1	0,195487	0,282821	0,771115	
Unemployed	0,0386577	0,263635	0,195487	1	0,97934	0,177846	
EdUnder9	0,0806681	0,363308	0,282821	0,97934	1	0,253252	
EdHighSchool	0,402644	0,726107	0,771115	0,177846	0,253252	1	
EdUniversity	0,450548	0,792118	0,866593	0,177986	0,265125	0,797234	
SpkEng	0,592451	0,759092	0,892247	0,13739	0,212796	0,692397	
SpkFr	0,216211	0,454734	0,44909	-0,0622372	-0,0283771	0,562436	
SpkOther	0,154573	0,52135	0,427989	0,739864	0,856283	0,375379	
Population	0,566968	0,880658	0,951085	0,275766	0,37573	0,830333	
PopMarried	0,588297	0,73245	0,911409	0,174171	0,254365	0,795186	
PopSingle	0,436124	0,842002	0,748458	0,32493	0,420092	0,551616	
PopAge0_14	0,407566	0,630215	0,826932	0,258399	0,3436	0,753478	

Fig. 6. Results of spatial autocorrelation applied on data sets in DWH

The fifth step includes presentation of the results of the analysis and generation reports using GIS and BI Publisher tools.

5 Advantages and Benefits of the Approach

BI system for advanced analysis of threats and risks has been designed and implemented as a dynamic and flexible framework with a variety of possibilities for improvement and expansion in all segments. This solution will certainly contribute to the quality and consistency of IBM in Bosnia and Herzegovina.

Basic functionality and capabilities of the system can be further expanded and enriched by adding new data sources from existing agencies and other relevant institutions, by consolidation of existing transactional systems, regulation procedures and templates for making the required reports, analysis, etc. As well, this system can be expanded with special geoportal that would allow other institutions and agencies access to thematic spatial data from the data warehouse for risk analysis at the local level in other segments of risk assesment. The geoportal can allow direct access to raw data in multiple formats, complete metadata, online visualization tools so users can create maps with data in the portal, automated provenance linkages across users, datasets and created maps, commenting mechanisms to discuss data quality and interpretation, and sharing or exporting created maps in various formats. This empowers BI solution with complementary technologies including spatial ETL, data visualization, and geographic information systems [12].

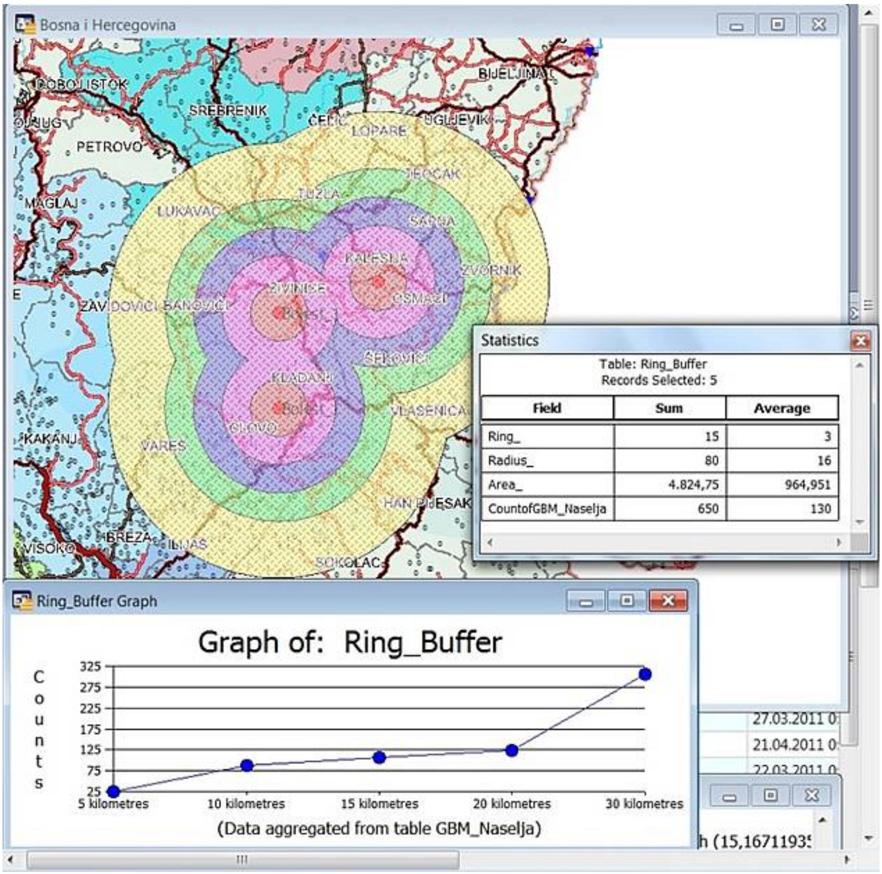


Fig. 7. Spatial reasoning: hot spot buffering and overlapping with the road network layer

For its implementation is a key issue to achieve high level of cooperation between RAC and individual information services (agencies). As well experience and knowledge of operating personnel (analysts and administrators) is very important. Lessons learned through design, implementation and use of this system can be extrapolated to other similar systems, with certain adjustments.

6 Conclusion

The implementation of a robust and complex BI system with a spatial extension, as a support to center for risk analysis, represents a challenge in any sense. Despite the availability of technological capabilities this requires additional innovation to achieve the goal and develop a functional and operational system. This paper shows an innovative approach for implementation business intelligence systems for advanced threat and risk analysis in RAC environment using spatial component. It demonstrates how to improve intelligence of complete information system by involving spatial extension. Particular benefit of this approach for risk analysis is effective utilization of location data, advanced spatial analysis techniques and more variety in data visualization.

By implementing this system at the center for risk analysis has been made the foundation of which is reflected in a modern, flexible and multiple employable system. In its production stage, it is not a closed system, but should be further developed through a series of possible improvements and options that should be recognized by its customers.

One of the proposed improvements is the extension of the system with geoportal for collaboration between the Center for Risk Analysis, and other local agencies (civil protection, forest companies, fire departments, police forces, public health departments, etc.) responsible for assessing the risks and threats to ensure preventive action and rescue of people and property.

In the case study project, we utilized standard data warehousing infrastructure to integrate data from multiple source systems and geospatial data. Data warehousing creates a single point of control for managing processes that cross the entire system, while on the other side geoinformation system (GIS) tools have the potential of exploiting the spatial context of any information to give qualitative and motivated decision-making. Integration of data warehousing, business intelligence capabilities and geoinformation technologies creates an information support tool that assists analysts and decision makers to understand complex spatial patterns, identify threats and reduce risks in dangerous situations from real life.

References

1. Gupta, J.N.D., Sharma, S.: Intelligent Enterprises of the 21st Century. IGI Global (2004)
2. Moss, L.T., Atre, S.: Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley (2003)
3. Curtis, G., Cobham, D.: Business Information Systems: Analysis, Design and Practice. Pearson Education (2008)

4. Rigaux, P., Scholl, M., Voisard, A.: Spatial Databases - With Application to GIS. Morgan Kaufmann (2002)
5. Shekhar, S., Chawla, S.: Spatial Databases: A Tour. Prentice Hall (2003)
6. Karabegovic, A., Ponjavic, M.: Integration and Interoperability of Spatial Data in Spatial Decision Support System Environment. In: MIPRO IEEE Croatia Conference, Opatija, Croatia (2010)
7. Batty, M., Longley, P.A.: Advanced Spatial Analysis: The CASA Book of GIS. ESRI Press (2003)
8. Worboys, M.F., Duckham, M.: GIS: A computing perspective. Taylor and Francis (2004)
9. de Smith, M., Goodchild, M., Longley, P.: Geospatial Analysis: A Comprenhensive Guide to Principles, Techniques and Software Tools. Matador (2010)
10. Haining, R.: Spatial Data Analysis: Theory and Practice. Cambridge University Press (2003)
11. Burrough, P.A., Donnell, M.C.: Principles of Geographical Information Systems. Oxford University Press (1998)
12. Karabegovic, A., Ponjavic, M.: Geoportal as Decision Support System with Spatial Data Warehouse. In: Federated Conference on Computer Science and Information Systems, ABICT 2012, Wraclaw, Poland, pp. 943–946 (2012)