Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 945–948

ISBN 978-83-60810-51-4

Innovation in Business Intelligence Systems: Spatial Component for Advanced Threat and Risk Analysis

Mirza Ponjavic
University of Sarajevo,
Faculty of Civil Engineering,
Department of Geodesy,
Patriotske lige 30, 71000 Sarajevo,
Bosnia and Herzegovina
Email: mirza.ponjavic@gf.unsa.ba

Almir Karabegovic
University of Sarajevo, Faculty of
Electrical Engineering, Department for
Computer Science and Informatics,
Kampus Univerziteta, 71000 Sarajevo,
Bosnia and Herzegovina
Email: almir.karabegovic@etf.unsa.ba

***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 relation. 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.**

I. 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 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.

This work is supported by Gauss Tuzla, Center for Geospatial Research Sarajevo

II. Spatial Component Improves Business Intelligence of Information Systems

Spatial component improves business intelligence of information systems and bring 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 relation [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.

III. Architecture of Business Intelligence Systems with Spatial Extension

An overview of key architectural components of Business Intelligence System (BIS) with spatial extension for RAC project is shown in Fig. 1. Data warehouse 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 in 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 application provide spatial data analysis, conducted for the purpose of indication and analysis of spatial trends [7]. 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.

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.

IV. Case Study

A. The Aim of the Project

Building a functional and efficient integrated system of 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.

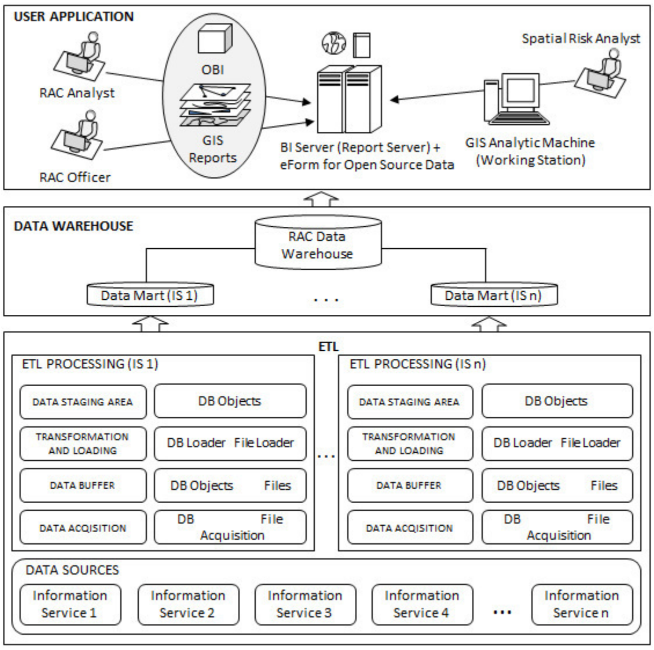

Fig. 1 Architecture of the business intelligence system with spatial extension for RAC

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 processing of data and propagating it towards the CAR.

B. 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 systems for spatial analysis of risk is directly related and dependent on the functionality of individual IS's. It is implemented through the phases of defining the scope, development, testing and deployment of this system.

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

For a number of types of analyzes that are 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 data and analytical techniques. Therefore, the data propagated from the information services (agencies) are geocoded, that is 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 crossing, which 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, road 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 (important institutions, values and indices of economic development), and also Web services offering background geodata (Microsoft Bing satellite images, street maps from Open Street Map web service).

User application environment, which includes the integration of GIS (MapInfo Professional), Oracle Business Intelligence with WebLogic / MapViewer AS (Fig. 3) covers a broad analytical functionality (geospatial analysis, business intelligence, reporting, publishing, creating a dashboard), which allows implementation of various types spatial analysis (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].

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.

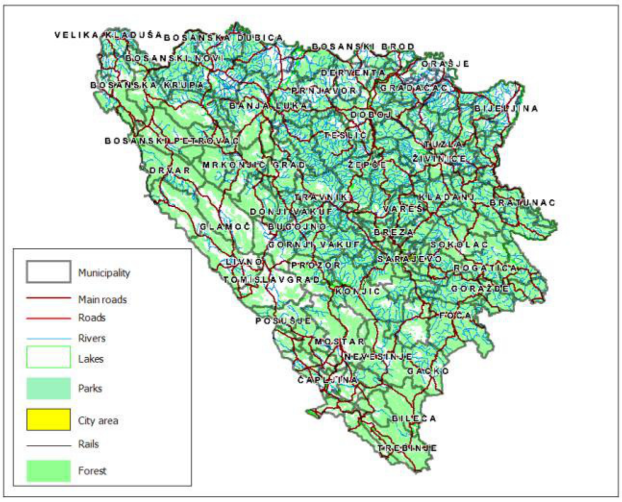

Fig. 2 Background geospatial data: basic cartographic detail

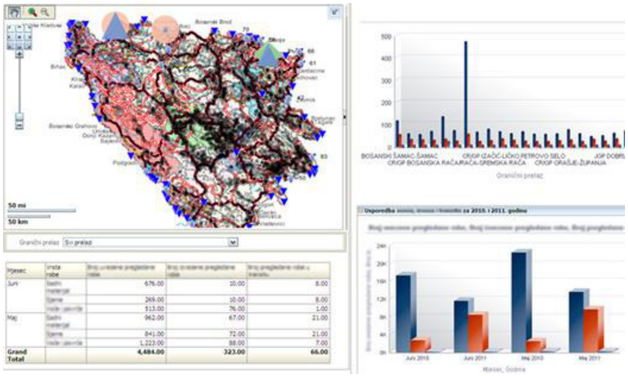

Fig. 3 Business intelligence dashboard for outlier detection

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). Foci of disease are generated using the hot spot and cluster analysis.

The second step is the selection of analytical method based on the entered data on the phenomenon and the available data sets in a 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 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:

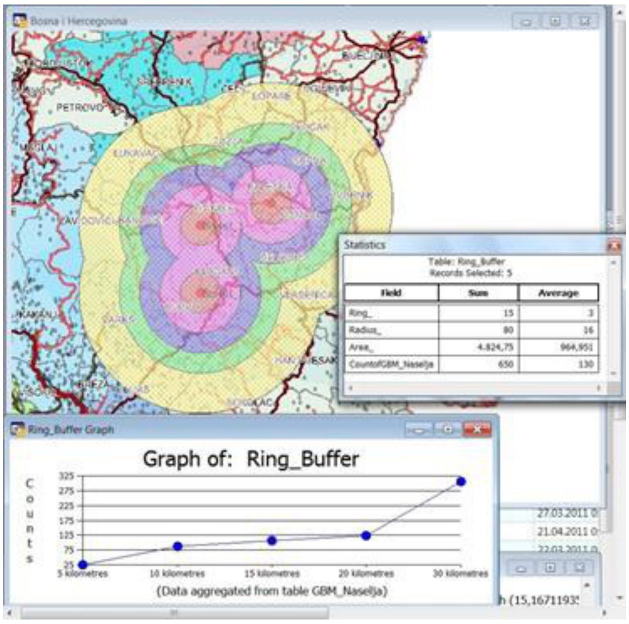

Fig. 4 Spatial reasoning: hot spot buffering and overlapping with the road network layer

- 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 phenomena, but it just says that there is a spatial relationship between certain 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. 4) it is concluded that the disease is caused by consuming dairy products and that is transmitted through infected sheep.

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

V. ADVANTAGES AND BENEFITS OF SUGGESTED SOLUTION

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. 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.

VI. 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.

REFERENCES

[1] J.N.D. Gupta, and S. Sharma, “Intelligent Enterprises of the 21st Century”, IGI Global, 2004
[2] L.T. Moss, and S. Atre, “Business intelligence roadmap: the complete project lifecycle for decision-support applications”, Addison-Wesley, 2003
[3] G. Curtis, and D. Cobham, “Business Information Systems: Analysis, Design and Practice”, Pearson Education, 2008
[4] P. Rigaux, M. Scholl, and A. Voisard, “Spatial Databases - With Application to GIS”, Morgan Kaufmann, 2002
[5] S. Shekhar, and S. Chawla, “Spatial Databases: A Tour”, Prentice Hall, 2003
[6] A. Karabegovic, M. Ponjavic, “Integration and Interoperability of Spatial Data in Spatial Decision Support System Environment”, MIPRO IEEE Croatia Conference, Opatija, Croatia, May 2010
[7] M. Batty, and P.A. Longley, “Advanced Spatial Analysis: The CASA Book of GIS”, ESRI Press, 2003
[8] M.F. Worboys, M.Duckham, “GIS: A computing perspective”, Taylor and Francis, 2004
[9] M. de Smith, M. Goodchild, P. Longley, “Geospatial Analysis: A Comprenhensive Guide to Principles, Techniques and Software Tools”, Matador, 2010
[10] R. Haining, “Spatial Data Analysis: Theory and Practice”, Cambridge University Press, 2003
[11] P.A. Burrough, M.C. Donnell, “Principles of Geographical Information Systems”, Oxford University Press, 1998