Geoportal as Interface for Data Warehouse and Business Intelligence Information System

Almir Karabegovic and Mirza Ponjavic

Abstract. There is increasing interest of organization for advanced presentation and data analysis for public users. This paper shows how to integrate data from enterprise data warehouse with spatial data warehouse, publish them together to online interactive map, and enable public users to perform analysis in simple web interface. As case study is used Business Intelligence System for Investors, where data comes from different sources, different levels, structured and unstructured. This approach has three phases: creating spatial data warehouse, implementing ETL (extract, transform and load) procedure for data from different sources (spatial and non-spatial) and, finally, designing interface for performing data analysis. The fact, that this is a public site, where users are not known in advanced and not trained, calls for importance of usability design and self-evident interface. Investors are not willing to invest any time in learning the basics of a system. Geographic information providers need geoportals to enable access to spatial data and services via the Internet; and it is a first step in creating Spatial Data Infrastructure (SDI).

1 Introduction

Business users constantly search for new and better ways for improving data warehousing capabilities. Many of them already use existing capabilities to strengthen analytics and business intelligence (BI). They have covered dimensions who, what, when, and why, but, there is only rare answer for where.

The first law of geography according to Waldo Tobler is "Everything is related to everything else, but near things are more related than distant things" [1]. But standard data warehouse cannot answer on the following questions that arise as a result of just this law. How far workers would travel from their house to job? How to choose the best location for a new dam? Optimize delivery route to meet

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

Mirza Ponjavic
University of Sarajevo, Faculty of Civil Engineering,
Department of Geodesy, 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_2,

changing customer demands and conditions? See which parcels and building are in potential flooding areas?

To answer those and other crucial business questions, we need access to true location intelligence – the kind of intelligence geospatial analysis can deliver [2] [3]. Location Intelligence (LI) has the capacity to organize and understand complex phenomena, through the use of geographic relationships, which are inherent in all information. By combining geographic- and location-related data with other business data, organizations can gain critical insights, make better decisions and optimize important processes and applications. Location Intelligence offers opportunity that organizations streamline their business processes and customer relationships to improve performance and results.

The growing availability of geospatial data and the demand for better analytic insight have helped to move location analysis from limited departmental implementations into enterprise-wide environments; from the hands of geographic information system (GIS) experts to IT organizations for deployment across businesses. Most of data in enterprise data warehouses (EDW) have a location reference. It makes possible that every business can enhance its business analytics with location intelligence as shown in Fig. 1.

The fact is that many of those companies cannot have a complete view of their business because their data are not integrated. Instead, their location data are stored in multiple departmental data marts (silos) that drive up costs, cause redundant data, and most important, does not utilize the richness of their data warehouse [4].

2 Use Case: Investment Promotion

2.1 Business Problem Background

Foreign Investment Promotion Agency (FIPA) of Bosnia and Herzegovina (BiH) is a state agency established with the mission to attract and maximize the flow of foreign direct investment into Bosnia and Herzegovina, and encourage existing foreign investors to further expand and develop their businesses in BiH, as well as facilitate the interaction between public and private sectors. It has an active role in policy advocacy in order to contribute to continually improving environment for business investments and economic development, and to promote a positive image of Bosnia and Herzegovina as a country that is attractive to foreign investors.

The project aims attracting and retaining cross-border investments and to provide better access to available land- related information. International best practice and academic research clearly suggest that easy access to land-related information is a key issue for domestic businesses and international investors. Better access to land-related information is clearly associated with greater government effectiveness and better quality of public services and will ultimately increase levels of

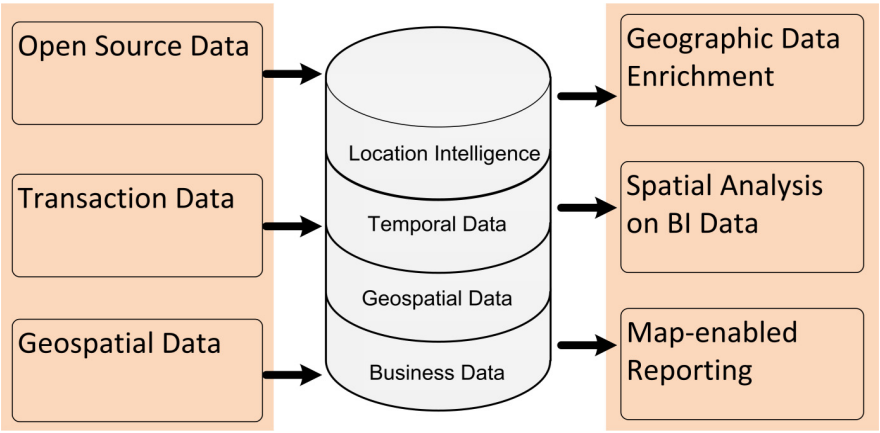

Fig. 1. Location intelligence as Business Intelligence with geospatial capabilities

investment. The World Bank Group's and the European Bank for Reconstruction and Development's Business Environment Survey (BEEPS) database identifies access to land as one of the major concerns for businesses. The European Union's research shows that making various forms of key land and property related information and information on practices, procedures and of relevant laws and regulations available and easy to access through European Union Land Information System (EULIS) is associated with larger average of business and investment opportunities ultimately leading to an improved business environment and investment climate. It also encourages a spread of best practices in presenting land-related data to businesses, establishing basis for comparing performance of localities and stimulates competition among localities in preparing land-related data in a digitalized form and making them available to the public users.

In Bosnia and Herzegovina, various projects have been implemented related to land registration and administration, and land construction resulting in digitalized form of land-related data in different institutions and levels of government. However, this land-related information is scattered in different ministries and agencies at different levels of government, and this makes it difficult for either public or private sector users to obtain land-related information easily and at an appropriate cost and timeframe.

The goal is therefore to build a platform using advanced GIS and web-based database technologies that would make key land-related information needed for investors available and easy to access. The resulting interactive map is intended to become a comprehensive source of land related information relevant for businesses and investors that is easy to view on-line and that can serve a wide range of public and private sector end users too.

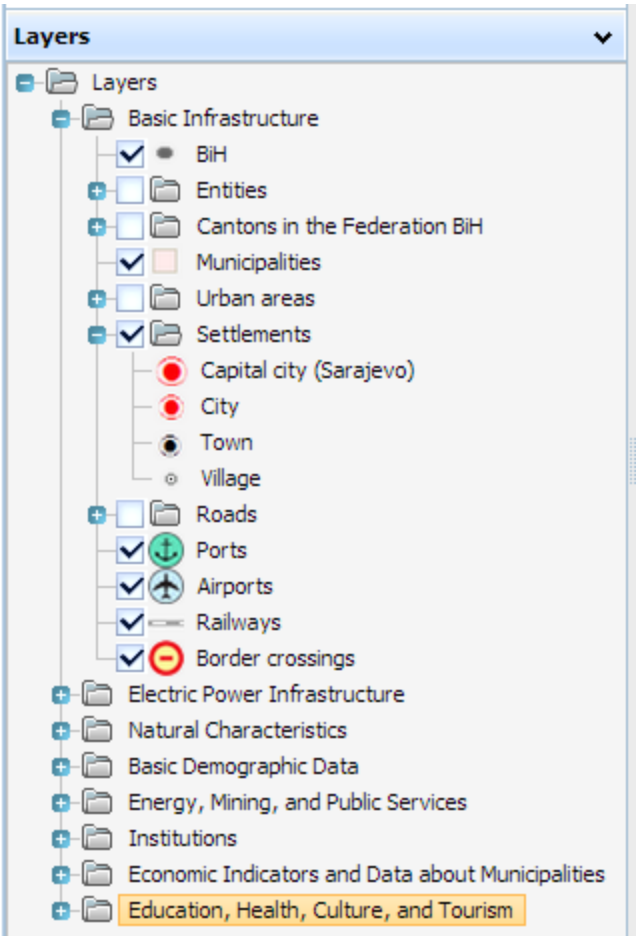

Fig. 2. Layer list with pictogram and legend for the geoportal

Initially, the following layers and data are available within the interactive map, as shown in Fig. 2: basic infrastructure (administrative boundaries, cities, roads, ports, airports, rail, and border crossings); detailed infrastructures (electricity, telecommunication, gas and water supply); natural resources/ environment (climatic zones, land use/ land cover, precipitation, soils, forest canopy coverage, elevation/ digital elevation model, water bodies); basic demographic data per municipality (population disaggregated by gender, age, education, employment, labor force availability); economic data (business entities with addresses, gross domestic product, gross investment, industry data for power plants, mining, manufacturing); special economic zones (business and industrial zones, technological parks and incubators, localities available for investment projects); institutions (business registration courts, customs offices, academic institutions, objects of cultural and

historical importance, top touristic locations); and other available data, images or links relevant for businesses and investors.

The task was to analyze and design the new system which will answer their goal. It is recognized that main obstacle was missing of any infrastructure for collaboration and sharing data between different agencies that produce spatial data. Searching for the best practices it is found that existing technologies convergence like web portals, data warehousing and location intelligence could offer to develop the new concept.

3 Geoportals

Geoportal is a web service platform for advanced application development, viewing, and editing of geospatial and business information in a service-oriented architecture. In this use case, solution was based on the Oracle Spatial database and Oracle Middleware Map Viewer web client platform.

This kind of portals can dramatically expand the availability of location based data to non-expert users for re-view, editing, and analysis. It enables fast and efficient creation and configuration of tailor-made, intuitive geodata applications for broad bases of users who require geospatial information integrated with business intelligence [5] [6]. As a result, task oriented, intuitive applications are now available to end users via internet without any software installation on client computers.

From other side, next-generation BI capabilities enable IT and business professionals to effectively leverage spatial analytics, improve system performance, and enhance management of complex BI environments. It means that technologies: Spatial Mapping capabilities with Business Intelligence analytic capabilities, work together to make better business decisions with automated, integrated location intelligence. This means that spatial information needs data appliances that can handle the volume and process it with BI technology and customers are demanding location-based data analytics.

Based on previous statements, it is possible to conclude that geoportal is a type of web portal used to find and access geographic information and associated geographic services (display, editing, analysis, etc.) via the Internet. Geoportals are important for effective use of GIS and a key element of Spatial Data Infrastructure (SDI).

Geographic information providers, including government agencies and commercial sources, use geoportals to publish descriptions (geospatial metadata) of their geographic information. Geographic information consumers, professional or casual, use geoportals to search and access the information they need. Thus geo-portals serve an increasingly important role in the sharing of geographic information and can avoid duplicated efforts, inconsistencies, delays, confusion, and wasted resources.

Recently, there has been a proliferation of geoportals for sharing of geographic information based on region or theme. Examples include the INSPIRE, or

Infra-structure for Spatial Information in the European Community geoportal, and UNSDI, the United Nations Spatial Data Infrastructure.

Modern web-based geoportals include 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. There are many use case examples of fields that can use advantages of such solutions like Government Healthcare (Disease Outbreak Tracking), Retail (Trade Area Analysis) or Marketing (Location-based Marketing Effectiveness) [7] [8].

3.1 Used Technology

In this project, it is chosen Oracle technology, because its database performance, enterprise data warehousing and analytics, which makes it easier to move location data into EDW for analysis.

This is a robust solution combining database native geospatial capabilities, geospatial services, and integrated analytics. It stores geospatial data directly in database environment so it is managed with business data. This provides consistency and integration and allows analytical tools to access geospatial data along with the rich data in data warehouse.

Geospatial technology is an in-database capability, and there are no extra costs or data marts required to use the capability. It integrates geospatial data and functions in data warehouse, which add benefits of in-database processing, including

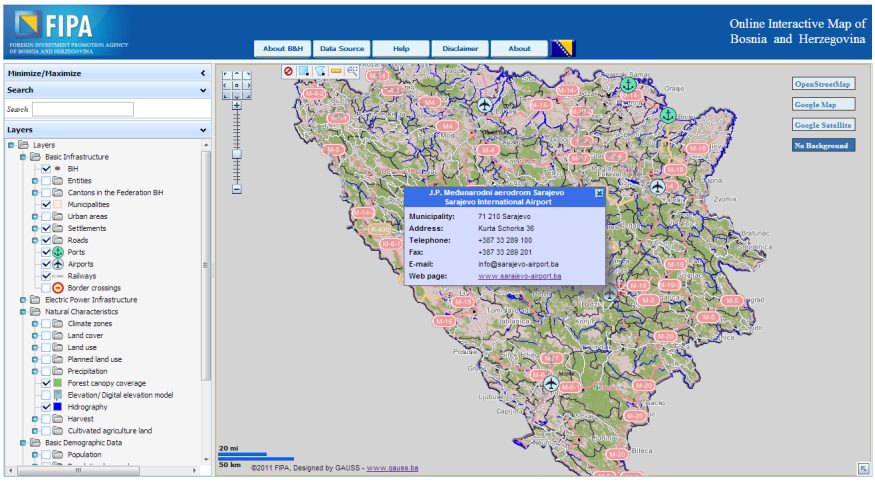

Fig. 3. Geoportal interface present data from spatial data warehouse

efficiency and cost effectiveness (reduce the cost of geospatial analysis by consolidating geospatial data marts, eliminating data redundancy and inconsistencies across applications, all of which leads to lower total cost of ownership); speed (speeds results by making data available for analysis sooner; there is no need to move data among systems); enhanced analytics (making better use of your location data enhances your analytics and BI); and scalability (can grow with business as data volumes increase).

3.2 The Interface

One of the most attractive and useful capabilities of geoportal is visualizing large amounts of information interactively. This ability to create multiple perspectives enhances a viewer's perceptive abilities to understand the phenomenon being studied.

Human-computer interface design focuses on how information is provided to and captured from users, and should provide a uniform structure for finding, viewing, and invoking the different components of a system. It actually defines the ways in which users interact with an information system. This is the reason we dedicated a large part of the work just for the geoportal interface, especially because it allows not only the interaction with the spatial data warehouse, but also present a kind of decision support system, as shown in Fig. 3.

4 Data Warehousing

Generally, a data warehouse is a large database designed to support the decision making needs of an organization as shown in Fig. 4. A data warehouse has been defined as a subject-oriented, integrated, time variant, nonvolatile collection of data that support a company's decision making process [9].

While data warehouses look at many types and dimensions of data, many are lacking in the spatial context of the data, such as an address, postal code, or provider location. By using technology that integrates this spatial component with the data warehouse, an organization can unlock hidden potential in their data and see hidden relationships and patterns in data, in essence data mining by geography.

In practice, there is evidence that spatially enabling database benefit an organization with more organized data structure; better integration of disparate data; new, spatially enabled analysis; reduced decision cycle time; and improved decisions. The spatial data warehouse extends the usefulness of online analytical processing (OLAP) systems. OLAP systems are used by decision makers to interrogate the data warehouse. The data for analysis with OLAP are accessed through metadata that document data source, frequency of update, and location of data. The data returned from the queries are represented as "multidimensional," although their form may be maintained as relational [10].

Spatial data warehouse, like Oracle BI, provides both a data model to the data warehouse and a geographic analysis engine for OLAP, which allows users to

store spatial data inside the data warehouse [11]. It offers data transformation and manipulation, a spatial storage engine, robust data access mechanisms, and a broad range of analytical tools and methods that are designed to facilitate spatial analysis.

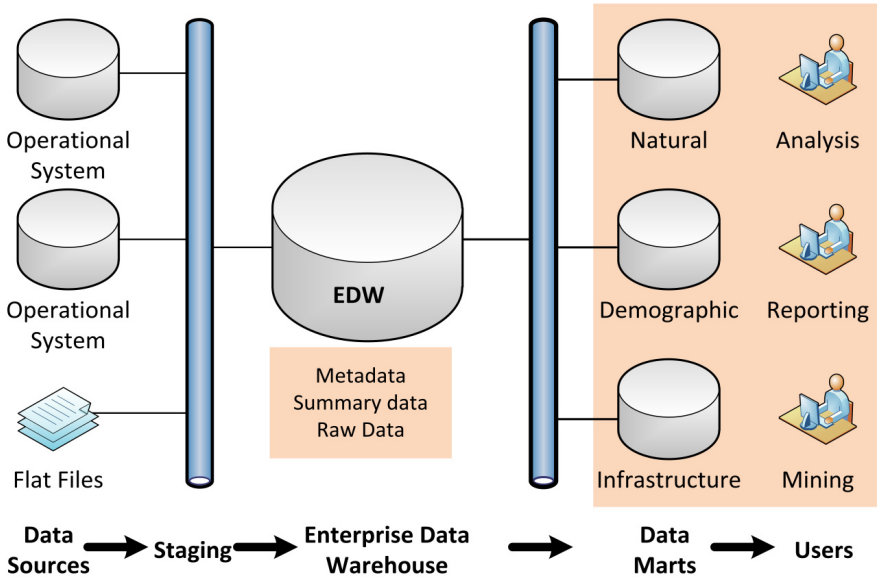

Fig. 4. Architecture of Business Intelligence System with Data Warehousing

5 Decision Support System with Spatial

Spatial data in BiH are characterized with fragmentation and lack of adequate availability of data, inconsistency, redundancy in collecting, insufficient use of standards, lack of coordination, and restrictions on data distribution. This situation makes it difficult to identify, access and use of existing spatial data in the country.

As per definition, this geoportal provides an entry point to access all data (geo-spatial data, remote sensing, information and services), and could be used to discovery, view, download, and transformation. We suggest building geoportal on three levels, web services platform, enterprise geoportal, and finally spatial data warehouse as shown in Fig. 5.

Web services platform contains Web GIS services, like Discovery service, means CSW (Catalogue Service Web), Viewing service, means WMS (Web Map Service), WCS (Web Coverage Service) and Download service, means WFS (Web Feature Service). Here also could be implemented geoprocessing services, open web services and tracking services. Enterprise geoportal contains catalog services, like Search, Channels, Link Browser Map, Download, and Collaboration. Finally, third level is data warehouse where data are stored.

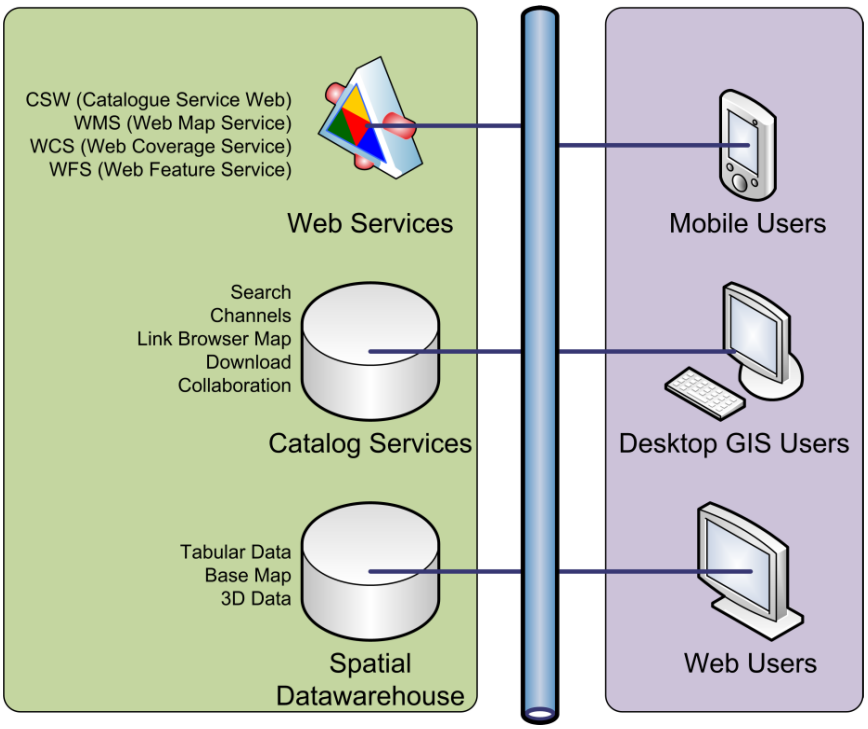

Fig. 5. Suggested architecture for geoportal based on web services

These services allow access to spatial information from different sources of local, national and global level in an interoperable manner and for a wider range of users to access relevant, harmonized and quality geographic information for the purposes of decision-making organizations and individuals.

We got information system that supports business decision making activities and serve the management and planning levels of organizations, show structural and non-structural data changing. Basically, this data warehouse represent knowledge based system, and we have interactive system intended to help decision makers compile useful information from a combination of raw data, documents, and personal knowledge, with their business models to identify and make decisions.

Summarizing previous statements we can see that it sounds like real decision support system. This decision support geoportal gather and present information like inventories of information assets (including legacy and relational data sources, cubes, data warehouses, and data marts); comparative statistic and demography figures between time points; and historic and projected economic indicators and natural characteristics based on statistic assumptions.

Generally, spatial-temporal domain is complex and characterized by a large amount of data as shown in Fig. 6. For advanced treat of data time series or huge amount data as spatial data usually are, there is need for advanced technique like

data mining or fuzzy clustering are. Fuzzy clustering methods allow classification of the data, when there is no a priori information about data set or content is not known. In particular, the fuzzy methods allow identifying data in more flexible manner, assigning to each datum degree of membership to all classes. [12]

The design criterion was data security because distribution of information via geoportal does not imply access to production databases, but replicated databases located in data marts.

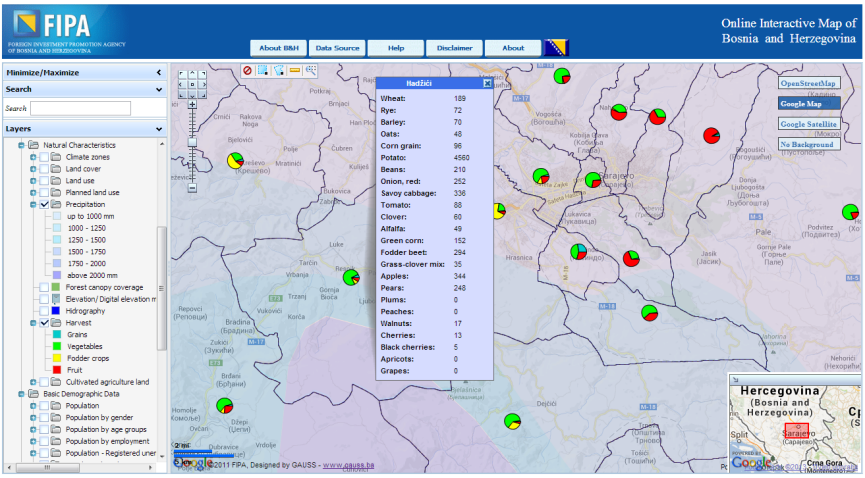

Fig. 6. Decision making in geoportal with available data from Data Warehouse

Implementation of the geoportal took following activities and deliveries: System analysis and design (design of system and software, as well as data models); Geoportal implementation and stuff training (policies and procedures for recovery, including failover functionality).

Authors recognized that there is need to research the extent to which open source software can support the development of geoportals as front-end of spatial data infrastructure especially in compliance with the INSPIRE directive [13].

5.1 Use Cases and Examples

There are more scenarios how users can make analysis and search for data in Online Interactive Map. Probably most popular methods are:

1. searching by mouse click on map
2. searching by Search form
3. predefined queries

Searches are initiated by clicking in the Search box and entering the search criteria. In most cases, these terms are "begins with" searches; meaning that user do not need to spell out the complete search criteria (Fig. 7.). As user entering search

criteria, it initiates the search in real time and the results appear in the information pane immediately below the search criteria. In background of application is integrated complete functionality of searching technique based on metadata or on parts of the original texts represented in databases (such as attributes or locations).

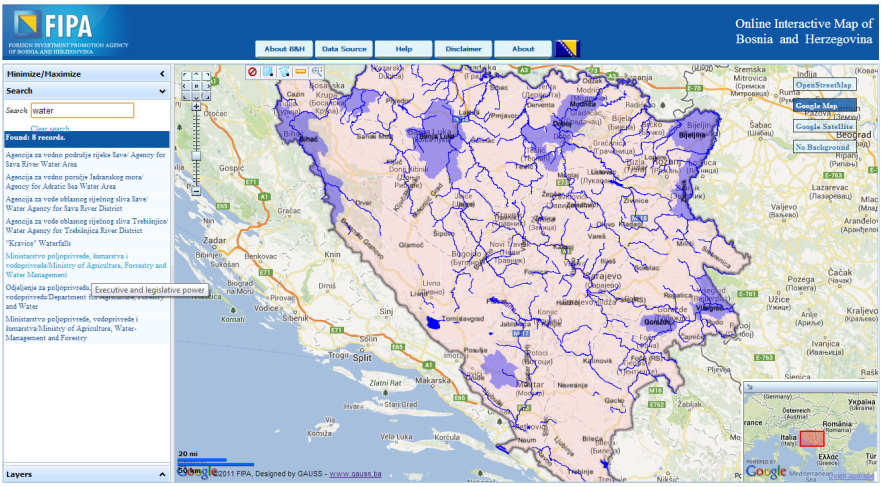

Fig. 7. Basic search for data

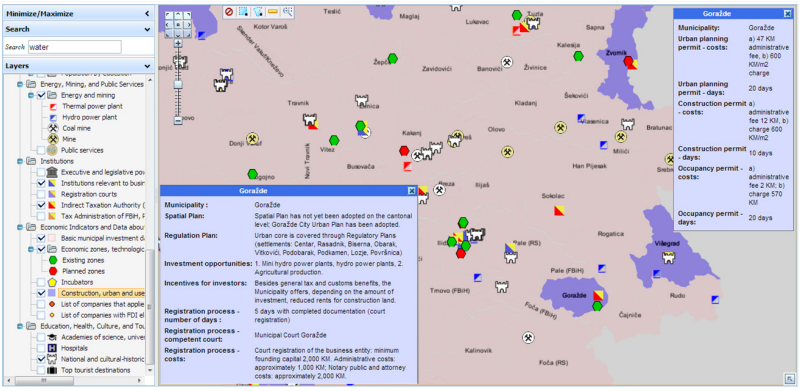

Fig. 8. Drill-down to basic infrastructure data and economic indicators

After that, potential investor can investigate Basic Infrastructure and Economic Indicator (investment data, economic zones, technological parks, incubators or construction, urban and use licenses) to better understand country potentials as shown in Fig. 8.

Using technique called "drill down" user goes from summary information to detailed data (Fig. 9.). In a GUI of GeoPortal, this means clicking on some representation in order to reveal more detail. It involves accessing information in database by starting with a general category and moving through the hierarchy: from

category to table to record to field. Using this technique user performs data analysis on a parent attribute. It is a method of exploring multidimensional data by moving from one level of detail to the next one, where levels depend on the data granularity. Good example is retrieving data about natural characteristics.

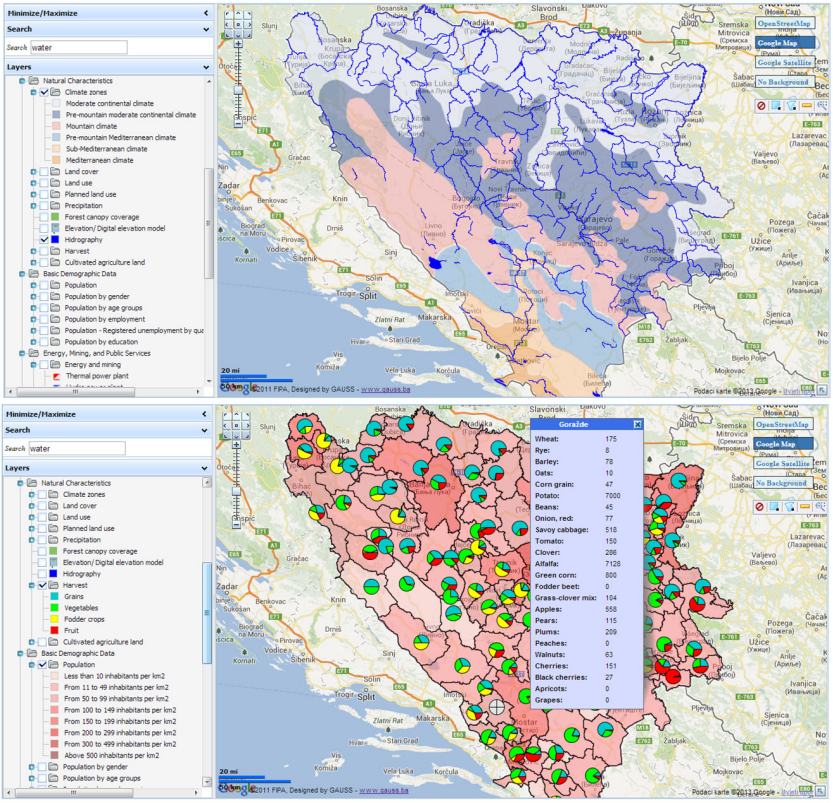

Fig. 9. Exploring data about natural characteristics

Revealing more details about data can include turning on data from other data source like external web services or data from image servers, like ones for orthophoto image presentation for whole country. This technique represent hybrid web visualization which enables integration different data format and from different sources. User can understand geography of location with measurements tools. Measurement panel displays distance, area, or radius depending on which tools in the toolbar are currently in use. Text appears in the following format: [label] [value] [unit of measure] to the right of the information panel. Example of this is presented in Figure 10.

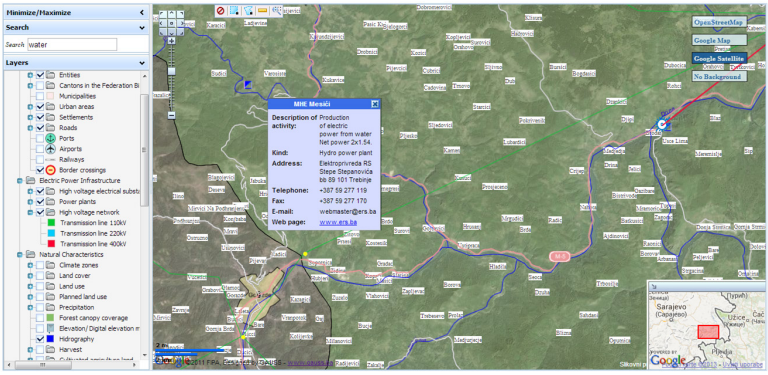

Fig. 10. Hybrid web representation with measurements functionality

At the end of analysis, user can summarize data, download in its own system and print it. Geoportal provides several map printing options. Since the current map extent shown in the active map window will be printed, generally user want first to zoom to the desired area and activate the desired layers before generating the printable output. This option allows generating printable maps in Acrobat PDF format. Also, this option offers more printing options, such as the ability to choose between several templates or scales. For investor most popular tool for final analysis is still Microsoft Excel, and they usually download analysis result in this format and continue with data analysis.

6 Conclusion and Future Work

The primary purpose of the geoportal is distribution and visualization of spatial data over the Internet, but its interactive capabilities could bring its functionality far beyond.

This paper proposed a usage of geoportals for decision making, especially with spatial data warehouses, whose main characteristics are: more organized data structure, better integration of disparate data, new spatially enabled analysis, reduced decision cycle time and improved decisions.

In the use case project, we utilized standard data warehousing infrastructure to integrate data from multiple source systems (different government agencies), enabling a central view across agencies. For agencies which could not accept the approach, we have created web services and catalogued all their data to central geo-portal.

We confirm advances in usage of spatial data warehouse to present spatial with business data together to generate thematic maps, because it present all information consistently, provide a single common data model, restructure the data so that it makes sense to the business users, and delivers excellent query performance, even for complex analytic queries.

One of the objectives in creating geoportal was to make it a single source for consistent spatial information. To achieve this goal there is need to establish cooperative mechanisms with other institutions and organizations involved in the collection, maintenance and distribution of geospatial information in the country, as well as monitoring and adjustment of the primary requirements of users.

As future work, we plan to incorporate case management system, add full collaboration system, and implement full WebGIS editing capabilities for agencies, who are interested to publish their data.

References

1. Tobler, W.: A computer movie simulating urban growth in the Detroit region. Economic Geography (1970)
2. Shekhar, S., Chawla, S.: Spatial Databases: A Tour. Prentice Hall (2003)
3. Worboys, M.F.: GIS: A computing perspective. Taylor and Francis (2004)
4. Batty, M., Longley, P.A.: Advanced Spatial Analysis: The CASA Book of GIS. ESRI Press (2003)
5. Goodchild, M.F., Fu, P., Rich, P.M.: Geographic information sharing: the case of the Geospatial One-Stop portal. Annals of the Association of American Geographers (2007)
6. Maguire, D.J., Longley, P.A.: The emergence of geoportals and their role in spatial data infrastructures. Computers, Environment and Urban Systems (2005)
7. Haining, R.: Spatial Data Analysis: Theory and Practice. Cambridge University Press (2003)
8. Abonyi, J., Feil, B.: Cluster Analysis for Data Mining and System Identification. Springer (2007)
9. Kimball, R., Ross, M.: The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons, Inc. (2002)
10. Moss, L.T., Atre, S.: Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley (2003)
11. Karabegovic, A., Ponjavic, M.: Integration and Interoperability of Spatial Data in Spatial Decision Support System Environment. In: MIPRO IEEE Croatia Conference, Opatija, Croatia (2010)
12. Karabegovic, A., Avdagic, Z., Ponjavic, M.: Fuzzy Clustering in Geospatial Analysis. In: MIPRO IEEE Croatia Conference, Opatija, Croatia (2011)
13. Petrovic, Z., Karabegovic, A., Ponjavic, M.: Spatial Data Infrastructure in Compliance with INSPIRE Directive and International Standards. In: 13th International Multidisciplinary Scientific GeoConference & SGEM 2013, Varna, Bulgaria (2013)