

# Encyclopedia of Information Science and Technology

Volume I–V

Mehdi Khosrow-Pour, D.B.A.  
*Information Resources Management association, USA*

pp. 33-37

**IDEA GROUP REFERENCE**  
Hershey – London – Melbourne - Singapore

# **A Web-Geographical Information System to Support Territorial Data Integration**

V. De Antonellis\*, G. Pozzi, F. A. Schreiber, L. Tanca, L. Tosi\*

Dipartimento di Elettronica e Informazione,  
Politecnico di Milano,  
p.zza L. da Vinci 32, 20133 Milano, Italy

## **1 . I N T R O D U C T I O N**

The design of a Web-geographical information system, Web-GIS (Zhong Ren & Ming Hsiang 2003, Worboys & Duckham 2004), strongly requires methodological and operational tools for dealing with information distributed in multiple, autonomous and heterogeneous data sources, and a uniform data publishing methodology and policy over Internet Web sites. In this article we describe the experience of the Politecnico di Milano group in the activities of requirement analysis and conceptual design of the DEAFIN Web-GIS (Schreiber et Al. 2003), whose objective is to provide a common environment for comparison of information about available vacant industrial sites coming from different regional data sources. Heterogeneity and Web availability requirements have been taken into account in the system architecture design; the system is thus conceived as a federated Web-based information system, apt to manage and provide access to all the regional relevant information in an integrated and complete fashion. Furthermore, since the data available by a given region partner can be both spatial and alphanumeric, a Web-GIS is defined for each regional component system.

## **2 . B A C K G R O U N D**

The DEAFIN (Development Agencies and their impact on Foreign direct Investments) project has been launched with the purpose of allowing companies and investors to get a comprehensive information framework about areas located in European regions suited for potential investments. The aim is to make the regional

---

\* Dipartimento di Elettronica per l'Automazione - Universita' di Brescia

\* Agenzia per la Mobilità e Ambiente – Comune di Milano

data about possible investment areas homogenous and comparable, and internationally accessible. Potential investors need both a survey and a detailed view of vacant sites in different locations, in order to compare different opportunities and decide their convenience. Quite naturally, such requirements call for a *Federated Information System* (FIS), which grants local sites a great deal of autonomy while enabling interoperation by means of a global integrated conceptual schema, i.e. the federated data schema. Furthermore, owing to the capillarity of the end user locations and to the need of a simple and widely known interface, *Web-based* access is mandatory. To define the functional specification of the system, the following activities have been carried out:

- *analysis of the requirements* of a distributed Web-based information system relying on a common conceptual database schema of the regional information which was initially (almost completely) available on paper support;
- *conceptual design* of the DEAFIN FIS, centered around the conceptual design of the federated conceptual database schema. The regional databases must be built according to the federated schema and then made accessible via Web. The availability of data on the Web allows potential investors to navigate in the DEAFIN site according to various and customizable criteria, based on a benchmarking model developed within the project.

## **I N F O R M A T I O N   R E Q U I R E M E N T S**

Three regional administrations from Germany, Italy, and Sweden were involved. The project started with a data gathering phase, aimed at collecting requirements about data and processes managed at the partner Public Administrations. A questionnaire-based tool was circulated to collect common information to be stored in the FIS.

The basis of the questionnaire is a set of the *data categories* managed in Public Administration information systems. The relevant data categories concern Land Use Plans (Master and Regional or specific), Territorial Services, Industrial Vacant Sites, Mobility data, Statistical and Social-Economic data, Base Cartography data, and Information on Cadastral Units data. Information on Vacant Industrial Sites is the main focus of the investigation. For each category, the questionnaire collected the data characteristics reported in Table 1.

In general, the data collected at the sites show uniformity with respect to the attention paid to cartographic availability, regulations and laws about reuse of vacant areas, and centralization of resources. In particular, the need exists at each regional site to introduce tools able to treat heterogeneous data, since these data more and more intensively are to come from various data sources, to be mapped into the federated schema. Also the degree of automation is similar, since cartographic systems and basic data management tools are available at the three partners' sites.

DATA CHARACTERISTICS
Availability of the data category
Location of the data source
Support technology and name of products/tools
Data format
Data owner
User roles involved in data access
Restrictions applying to data access
Performances
Maintenance policies
Availability on the Web

Table 1. Summary of data characteristics collected in the data gathering phase

Several ongoing local projects concern the digital acquisition of land use plans, the automation of document management, and the development of various thematic databases and Web sites. What is required is a unified common schema for the regional databases. Moreover, the need of a uniform data publishing methodology and policy over Internet Web sites emerges clearly from the participants and involves almost all data categories.

## U S A G E M O D E S

The data gathering phase has detailed also a set of user profiles which specify how different access demands can be supported by the system towards a variety of user groups. The design and implementation of profiles have obviously a deep impact on the usability of the system. Hence, a careful analysis of user typologies and profiles has been performed during the specification phase, while an enrichment of profiles and access modes has been planned in a post-implementation follow up. The purpose is to have the system start with a set of pre-configured access typologies, and then tune the access modes and user profiles against the most typical uses observed for a fixed period on the DEAFIN pilot implementation. The first broad difference is between the profile of public administrations and that of private users, due to different data needs. Moreover, two basic interaction modes must be provided: browsing (using thematic areas and other refined search parameters) and querying (using simple and user-friendly interfaces). The system can be regarded as: a *passive tool*, when responding to user questions; an *active tool*, when used as a decision support system, or when used as a *standard application* allowing new European partners to join the DEAFIN consortium.

As a passive tool, the contents of the federated database can show the advantages and disadvantages of an area: the information provided can be related to the specific search needs of a specific user. These searches are different according to user types, but apart from the function of locating (or re-locating) business activities, the motivation probably exists to retrieve general information about a region.

### **3 . S Y S T E M A R C H I T E C T U R E**

In the literature, several approaches and tools for handling heterogeneous data sources have been developed, and standards for distributed information systems have been defined (Wiederhold 1992, Mylopoulos & Papazoglou 1997). For these systems, the use of multiple layer, mediator based architectures, and of a common data model have been employed (GarciaMolina et Al. 1997). Wrapper/extractor and mediator tools (see Terms and Definitions Section) are proposed to obtain a uniform data representation (abstracting from the formats of the original data sources) and to facilitate federated access. Following this direction, in order to handle heterogeneity at each regional site, an architecture has been defined where extractors translate data coming from local heterogeneous data sources to a common reference format defined at the mediator level (Schreiber et Al. 2003).

The aspect of data publishing over Internet Web sites has emerged clearly as a need from the partner regions. The system is thus conceived as a Web-based federated information system, apt to manage and provide access to all the regional relevant information in an integrated and complete fashion, thus satisfying the Web availability need.

Generally, data made available by a given partner region can be both spatial and alphanumeric, requiring each regional component system to be a GIS. Each regional site is in charge of data publishing and certification over the system according to the common federated schema. Each regional Web site contains a specific interface and stores regional data in a database by means of a Web-GIS technology, and operates as a central server that contains the common interface and some aggregate data. The global system is composed of Web-interconnected sites: the central Web site is the reference site for aggregate data and for requirements about new regional sites to be added in the federation or new applications to be developed in the component systems.

#### **F E D E R A T E D D A T A B A S E C O N C E P T U A L D E S I G N**

The federated database conceptual schema provides an integrated high level representation of all data to be handled by the FIS. In order to achieve the highest readability and the widest possible diffusion, a common set of terms is needed. We adopted the dictionary of terms as defined by the EuroDicAutom (EuroDicAutom 2001) of the EU: International System of measurements, and commonly agreed data categories were adopted.

The database conceptual design produced a set of conceptual views, represented in terms of ER diagrams (see Terms and Definitions section), that provide schematic requirement representations, each related to a specific information set. The following views were defined: Vacant Site Description view, Land Use Plan view, Administration view, Eco/Reclamation view, Transportation view, Service view, Document view, Procedure And Task view; the global conceptual data schema is a merge of these different views.

The design identified 36 entities (including entities with spatial representation that describes their shape, extension and location on the earth surface, marked as

*Geo-referenced*), with a total number of about 200 attributes. The XML (eXtensible Markup Language) language was chosen for logical specification, providing a common reference format to facilitate data customization, translation and mediation.

The FIS should enable the user to dynamically query the databases and in particular the geographical data they contain; these requirements, as we discussed above, is fulfilled by a Web-GIS system able to perform advanced operations guided by the client side.

## **4 . T H E W E B - G I S**

The interface of the central Web site allows the users to access data in two search modes: *direct search*, which leads directly to the vacant area of interest, including also information about the surroundings; *navigation*, which presents a list of vacant sites to be filtered progressively, according to various criteria. Indeed, a static benchmarking table presents a list of site features, along with the corresponding values for each partner: from this table the user can choose a partner and obtain pages that describe the partner's area highlighting its most attractive features. Some aggregate data about the vacant sites of each region are also displayed for subsequent browsing operations, where the user can execute queries in order to filter the vacant sites on the basis of preferences or specific criteria such as: the type of usage, area size, costs, surrounding services, accessibility. Finally, links to the regional Web sites are provided.

The Web interfaces of regional sites support more specific functions: the exploration of the partner area by choosing a specific theme as land planning, business and market activities, mobility system, services, demographic and macroeconomic data and development strategies; the access to business information as contact offices and particular business opportunities that are currently available for some vacant sites of the area; the access to aggregated data about vacant sites of the area and to a benchmarking table where the user can apply some selection conditions based on the available comparison parameters; the search of vacant sites based on forms that guide the user in the specification of the query, including geography related conditions (SQL with geographic extensions) like "search the vacant sites with a railway station within a radius of <parameter value> km"; the selection of specific of the chosen vacant site, e.g. buildings, eco quality, themes presenting the characteristics technological networks, land planning, mobility system, territorial services, procedures and tasks and documents.

## **5 . F U T U R E T R E N D S**

Web-GIS applications are becoming more and more important to a growing number of activities. Geography, geology, environmental studies, business marketing, and other disciplines have gained benefits from GIS tools and methods. Continual improvements in GIS hardware and software will lead to a much wider

application of this technology throughout government, business, and industry. In particular, integrated geodata services based on data format standardization will increasingly facilitate the exchange of information among users of different systems by allowing data sharing and improving cooperation and communication among the organizations involved in environmental protection, planning, and resource management.

## 6 . C O N C L U S I O N S

The development of the DEAFIN Federated Information System started from an analysis of the requirements and of the locally available information, performed through a data gathering phase. The federated database was modeled in the form of a set of conceptual views both as ER diagrams and XML specifications. The architecture of the overall system is in the form of a federated Web-GIS: at the main Web site a high level view over the entire federated system is allowed, while local specializations are allowed at the level of every regional Web site adhering to the recommendations for the entire database.

The proposed system offers a marketing and information platform in a European context based on advanced Web functions integrating Geographic Information Systems (GIS), decision support tools, and database federation features. Analogous projects in the European area have been developed mainly in the field of tourism (Pühretmair & Wöß 2001) where integration of data sources and access to GIS in a graphical context are needed.

Further details about the project can be found in (Schreiber et Al. 2003).

## R E F E R E N C E S

EuroDicAutom (2001) Automatic Translation of Terms, European Community.  
<http://eurodic.ip.lu:8086/cgi-bin/edicbin/EuroDicWWW.pl?page=expert>

Garcia Molina H., Papakonstantinou Y., Quass D., Rajaraman A., Sagiv Y., Ullman J., Vassolos V., & Widom J. (1997). The TSIMMIS approach to mediation: data models and languages. *Journal of Intelligent Information Systems*, 8, 117-132.

Mylopoulos J. & Papazoglou M. (Ed.). (1997). *Cooperative Information System* [Special issue]. *IEEE Expert*, 12(5).

OpenGIS Consortium. (2001). *Geography Markup Language (GML)*, 20 February 2001, OGC Document Number: 01-029.

OpenGIS Consortium. (2001). *Web Map Service Implementation Specification*, 21 June 2001, OGC Document Number: 01-047r2.

Pühretmair F., & Wöß W. (2001). XML-Based Integration of GIS and Heterogeneous Tourism Information. In Dittrich K., Geppert A., Norrie M. (Eds.), *Advanced Information Systems Engineering* (pp. 346-358), Berlin: LNCS Springer Verlag.

Schreiber F.A., Belussi A., De Antonellis V., Fugini M.G., Pozzi G., Tanca L., Tosi L. (2003). The design of the DEAFIN web-geographical information system:

an experience in the integration of territorial reclamation support services. In Ajantha Dahanayake and Waltraud Gerhardt (Eds.), *Web-enabled systems integration: practice and challenges* (pp 142-168). Hershey, PA: Idea Group Publishing.

The World-Wide Web Consortium, <http://www.w3.org>

Wiederhold G. (1992). Mediators in the Architecture of Future Information Systems. *IEEE Computer*, 25, 38-49.

Worboys M.F. & Duckham M. (2004). *GIS: A Computing Perspective*, (2nd ed.). Boca Raton: CRC Press

Zhong Ren P., & Ming Hsiang T. (2003). *Internet GIS*, New York: John Wiley and Sons Inc.

## Terms and Definitions

**Conceptual Schema of a Database:** a semi-formal high level description of the database, independent of its implementation.

**ER (Entity-Relationship) Diagrams:** the most widely used model to express the database conceptual schema.

**eXtensible Markup Language (XML):** markup language proposed by the World Wide Web Consortium [W3C] for data and documents interchange.

**Federated Information System (FIS):** an Information System is named federated when it supports interoperation among several autonomous and possibly heterogeneous Information Systems, by means of a shared global data schema.

**Geographical Information System (GIS):** Information System storing geographical data along with alphanumeric and spatial components. GIS systems also provide the data structures and algorithms to represent and query efficiently a collection of geographical data.

**Mediator:** a software component providing a uniform integrated interface to process and execute queries over data stored in multiple, heterogeneous data sources.

**Web-GIS:** a GIS system empowered with a Web-based interface.

**Wrapper:** a software tool to extract content from data sources and perform data format translation.