

Free and Open Source WebGIS Solutions for the PO-BASyN Project

Glauco MANTEGARI¹ – Marco DE SALVO²

Department of Informatics, Systems and Communication, University of Milano Bicocca, Milan, Italy

¹glauco.mantegari@disco.unimib.it

²mdesalvo@alice.it

Abstract

The paper describes the use of Free and Open Source WebGIS technologies in the context of the PO-BASyN project which studies the Bronze Age settlement dynamics in the Po River Valley (northern Italy). The storage, search and retrieval of spatial information by means of efficient tools and techniques are central elements for this research, and thus GIS applications play a key role. The development of the PO-BASyN WebGIS created several issues related to georeferenced data integration, access and sharing on the web. In order to face these issues the system was conceived as an open and evolutive architecture since the beginning. It was developed using exclusively Free and Open Source technologies. The description of the system's requirements, architecture and functionalities shows the effectiveness of this approach and the possibility of continuously improving the system without the constraints of proprietary technologies.

Keywords

Web-based GIS, Open Source technologies, Open formats, Bronze Age Archaeology, Italy

1. Introduction

The PO-BASyN¹ project studies the Bronze Age settlement dynamics in the Po River Valley (northern Italy). This area is characterized by geographical and environmental uniformity as well as by its vast extent. The regional scale considered by the project involves the Po Valley in itself as well as its surrounding areas, up to the southern Alpine chain, the northern Apennines and the Adriatic Sea.

The Bronze Age in this area has been the object of intensive research (see de Marinis 1997; 1999; Bernabò Brea *et al.* 1997): a large number of excavations were conducted during the last decades and old investigations were checked and revised, thus producing rich datasets and literature. Moreover, a large number of research institutions are working on this subject every year; as a consequence, the datasets and literature are still growing exponentially. In this scenario it is crucial to manage and share heterogeneous information – which ranges from structured and semi-structured data to documents, images and cartographies – with the scientific community.

In particular, the storage, search and retrieval of spatial information are central elements: Geographical Information Systems (GIS) thus play a key role in the project. Moreover, recent developments in Web-based technologies offer the possibility to define new and effective ways of managing and sharing geospatial data by means of Web-based Geographical Information Systems.

In this vision and as already described in a previous paper (Mantegari *et al.* 2007), we began to develop a WebGIS system with simple query capabilities, which was based on three main functional requirements: to collect georeferenced data concerning sites and ceramic findings; to visualize data in an interactive map; to perform queries based on textual attributes. In the middle of 2007 the system was revised in order to make it more flexible and modular and to go beyond the limitations of the first version of the system. In fact, Web-based GIS technologies have consistently evolved in the last few years, thanks to the diffusion of applications such as Google Maps², Google Earth³, Microsoft Virtual Earth⁴, etc., which contributed to the definition of the new scenario of

¹ The acronym stands for “PO Valley and Bronze Age Settlement Dynamics”. The project, started in 2005, involves the University of Milan (Chair of Prehistory and Protohistory), the University of Bologna (Department of Archaeology) and the University of Milan Bicocca (Department of Informatics, Systems and Communication).

² <http://www.maps.google.com>

³ <http://www.earth.google.com>

⁴ <http://www.microsoft.com/virtualearth>

the “geospatial Web” (GeoWeb). GeoWeb, defined as a “global geospatial network that does for geodata what the World Wide Web originally did for textual data, that is make it shareable, searchable and ubiquitous” (Gerlek and Fleagle 2007), gave us new ideas and technologies to re-define the system’s functionalities and architecture. In particular, the main improvements were a more efficient support for sharing, using geospatial web services, such as WMS (de la Beaujardiere 2006) and WFS (Vretanos 2005); the application of *de facto* standard data formats, such as KML⁵; the development and testing of some basic geospatial data analysis functionalities.

Since the beginning the system was developed using Free and Open Source technologies, which have shown over the years to be mature enough to compete with commercial and proprietary solutions and most of all have let us experiment and re-define the system without the high costs of traditional software licenses. The directions in which geospatial web technologies have evolved have furthermore proved the validity of this approach, because of the strict link between the GeoWeb and Free and Open Source technologies and Open formats.

The remainder of the paper is organized as follows: section 2 describes the system architecture, focusing briefly on the characteristics of each software component; section 3 discusses the functionalities that are currently developed with particular attention given to the distinctive features of our application. Section 4 draws some conclusions and intentions for future work.

2. System architecture

PO-BASyN WebGIS is a Web-based archaeological informa-

tion system built on a completely Open Source infrastructure. The system core uses the LAPP stack, which indicates a technological design pattern based on the Linux Operating System, the Apache Web Server, the PostgreSQL Database Management System and the PHP scripting language.

The foundation of our platform is Ubuntu⁶, a Debian based Linux distribution which was chosen for its ease of use, its excellent software repository system and its community support which is strongly present worldwide.

The system architecture is a classical “three-tier model” implementation, which allows a flexible separation of the domain logic from the data source and the data presentation logics (Fig. 1).

The following sections briefly describe the main characteristics of each level.

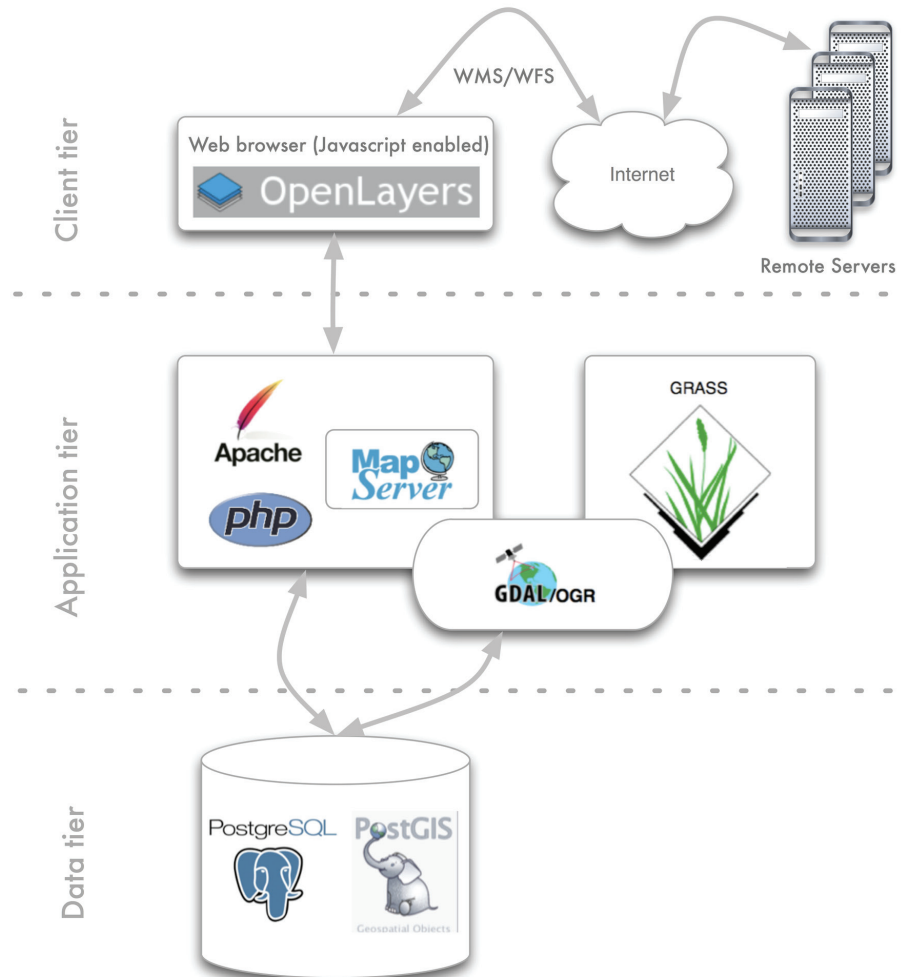


Fig. 1. General scheme of the system architecture.

⁵ KML has been approved as Open Standard by the Open Geospatial Consortium in 2008 (www.opengeospatial.org/standards/kml/).

⁶ <http://www.ubuntu.com>

2.1. Data tier

The data level is based on the PostgreSQL DBMS technology⁷ and the PostGIS spatial extension⁸. PostgreSQL is one of the most powerful existing solutions which offers a good balance between the number of functionalities provided, the efficiency in storing data and the preservation of the dataset integrity. Its most interesting feature is the integration of the PostGIS spatial extensions, which provides a solid WebGIS backend capable of performing even complex spatial analysis functions.

2.2. Application tier

The domain logic tier is based on the University of Minnesota's MapServer⁹ and the GRASS¹⁰ Geographical Information System.

UMN MapServer is particularly suited for the development of a Web-mapping environment: in fact, it is able to produce interactive and highly customizable maps and can also support a wide range of vector and raster data source formats. Above all, Map Server is used as an Open Geospatial Consortium¹¹ WMS server, which can publish the project's geodata sources on the web. Moreover, one interesting feature of this product we take advantage of is its rich thematic mapping capabilities that give a precise control on the users' web mapping experience.

In addition to MapServer, GRASS GIS – the leading OS solution for geospatial data management and analysis, image processing, spatial modelling and visualization (Neteler and Mitasova 2008) – was experimentally integrated and tested as a powerful engine to perform geostatistical computations on the data which is progressively collected in the system (see section 3.3).

2.3. Presentation tier

The presentation tier is based on the OpenLayers¹² technology. OpenLayers is an Open Source Web mapping framework which heavily relies on the AJAX communication paradigm to give the users a highly interactive and easy-to-use WebGIS environment. The choice of OpenLayers came after a technological survey which we conducted in order to evaluate the OS products that offered rich WebGIS interaction via AJAX and/or Flash and some experimental prototypes.

The survey took these main requirements into consideration:

- the openness of the source code, in order to be able to fully access the functionalities of the tool and to make *ad hoc* configurations;
- the capabilities of each application in terms of the overall user experience and the interface usability;
- the efficiency of the graphic user interface.

Beyond OpenLayers and several other applications, which proved to be unsatisfactory for our purposes, two applications need to be cited. The first one, named "msCross"¹³ is an AJAX lightweight interface for UMN MapServer, which proved to be particularly suited in the case of the quick integration of an interactive WebGIS interface. However, this solution is completely dependant on an application server, such as UMN MapServer, GeoServer¹⁴, etc.

The second solution, called "Flamingo MapComponents"¹⁵, is a Flash Web-mapping framework, that offers all the advantages of the Flash technology in terms of the development of an eye-catching, highly interactive and open standards compliant interface. However this solution proved to be too difficult to program, in particular the integration of a JavaScript engine with the Flash engine.

⁷ <http://www.postgresql.org>

⁸ <http://www.postgis.org>

⁹ <http://mapserver.gis.umn.edu>

¹⁰ <http://grass.itc.it>

¹¹ <http://www.opengeospatial.org>

¹² <http://www.openlayers.org>

¹³ http://datacrossing.crs4.it/en_Documentation_mscross.html

¹⁴ <http://geoserver.org/display/GEOS/Welcome>

¹⁵ <http://www.flamingo-mc.org>

3. System functionalities

The system is organized into four main areas, which allow a user to access three groups of functionalities:

- Navigation and query
- Management
- Geospatial data analysis

The following sections briefly describe the main characteristics of each area.

3.1. Navigation and query

Navigation and query can be performed in several different ways. The main requirement behind the design of the navigation paths was to allow easy and intuitive access to data, in order to make the system friendly for different categories of users. The same concept was applied to the design of the query modalities; they are based on the definition of the data categories that are considered to be the most relevant by scholars.

A typical navigation scenario is proposed in *Fig. 2*: the user queries the databases of the sites and obtains a results list. The sites can be viewed in the WebGIS map. Then, the user can consult the file containing the details about one or more sites of interest and see their geographical position in a small map on the left section of the file. The detailed file of the site allows access to a text-based list or to a thumbnails gallery concerning all the ceramic findings that pertain to the site. It is of course possible to view the details of each finding in a separate file. Thanks to hyperlinks several other navigation paths can be followed.

An additional feature that is available in the system is the possibility to export the entire dataset and the set of results of a specific query. In most occasions, scholars may need to work and/or consult data locally, rather than in a Web-based environment. For this reason an export facility was developed that converts the desired recordset into Google Earth's KML file format, which is one of the *de facto* standards for geospatial data representation on the web and is based on XML grammar, thus allowing for a completely open data representation.

3.2. System management

System management is ruled by a user access control based on a login name and a password. Four categories of users are currently defined:

1. Non-registered users: can freely consult the WebGIS as well as the database entries;
2. Editors (normally the students involved in data entry): can add new data, both by georeferencing new sites, by filling the database forms concerning sites and artifacts, and by modifying previously entered data. The modifications will be marked as “pending” to non-registered users until they are reviewed and approved by an administrator;
3. Administrators (normally scholars and professors): can delete records and approve the

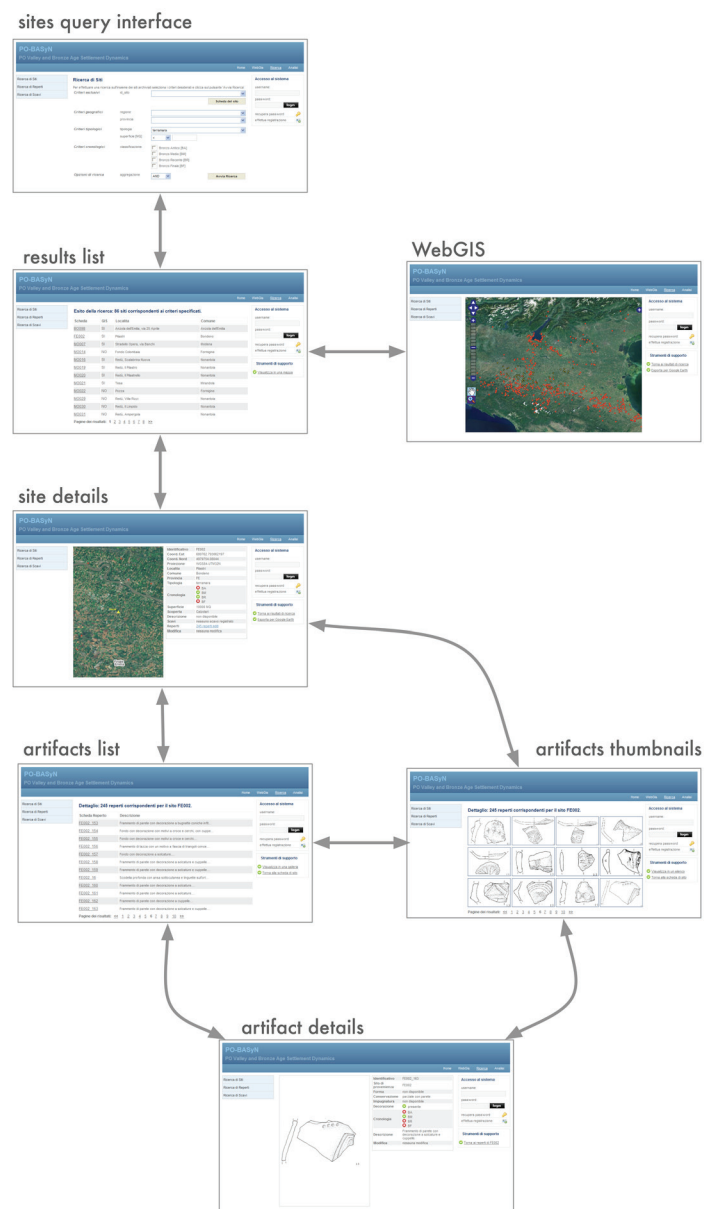


Fig. 2. A typical navigation scenario.

modifications made by the editors, thus granting the validity of newly entered data.

An important issue concerning data management regards georeferencing. Due to the extent of the area and the different geospatial data sources – which range from 19th-century maps to very precise modern cartographies – the precision in the localization of the sites varies a lot. Moreover, errors and imprecision are due to the different procedures with which each research group scans and georeferences the paper-based maps and from the different base cartographies that are available for each sub-area of the Po Valley. Integrating geospatial data from different sources may result in imprecise overlay and this, in turn, may negatively condition the analysis of the settlements' distributions over the area considered by the project.

A shared platform for data georeferencing was therefore considered fundamental in order to increase the precision of the positioning of the sites.

The choice was the Google Earth application, because of its ease of use and detailed coverage available for our area of interest. Some tests were done to verify the precision of the georeferenced cartographies available in Google Earth, by using known locations over the entire area of the project. The results were surprisingly good, thus convincing us of the reliability of a georeferencing procedure based on this application.

Users can add placemarks for sites in Google Earth and then export a KML (or even a KMZ) file. Placemarks must be named with the code assigned for each site of the system, in order to guarantee data integrity. In the Management area a wizard guides the user through importing the KML/KMZ file which is parsed for the extraction of well-formed placemarks. If a site does not yet exist in the WebGIS database, it will be added and a new file will be automatically created in order to allow data entry concerning the details of the site itself. If a site is already present in the database, its georeferencing will be automatically updated. This system proved to be very effective because it is extremely easy to use; moreover, it is also efficient, because it was shown to be able to manage even dozens of georeferenced sites in a quick and precise way.

3.3. Towards geospatial data analysis

Geospatial data analysis constitutes one of the most important and distinguishing features of GIS. The

analytical possibilities of a GIS mainly depend on the specific software suite being adopted, which can provide a more or less articulated set of tools. An interesting improvement of the PO-BASyN WebGIS is the development of a simple set of tools for geospatial data analyses on the web.

This way, scholars would be able to perform some of the most common analyses regarding the distribution of ancient settlements on a constantly updated dataset. Thanks to the possibility of exporting data in standard formats (e.g. KML export, described in section 3.2) it would be then possible to perform more sophisticated geospatial data analyses in a desktop GIS environment.

In fact, several constraints emerge when trying to offer geospatial data analysis functionalities in a Web-based system. At the moment these constraints regard, for example, the hardware and software resources of the server, which may be insufficient to compute complex algorithms efficiently and to offer the results in the users' browsers, especially when a large number of clients require the same service on different data. Because of this we tested some very simple analysis functionalities by using different approaches, as will be explained later. In particular, the system currently offers (*Fig. 3*):

- Gaussian kernel distribution
- Thiessen polygons.

Moreover, it is possible to perform thematic mapping on the basis of the query results on the database of the findings, with different symbols representing the different occurrences of any particular ceramic type.

In order to develop these tools, different approaches were investigated and implemented, using MapServer and PostGIS functionalities, as well as integrating GRASS GIS into the system.

In particular, MapServer offers the possibility to perform basic types of analyses related to thematic mapping. It is possible to specify the layer to be categorized, as well as to filter the interesting items and their values and to represent the results of this selection by using different symbols. This is the case of the display of ceramic types' occurrence shown in *Fig. 3*.

PostGIS offers SQL-based constructs that can build complex spatial elaborations, ranging from distance calculations to buffers and to the use of topological predicates. These may represent non-trivial analyses for a WebGIS and, in our case, they

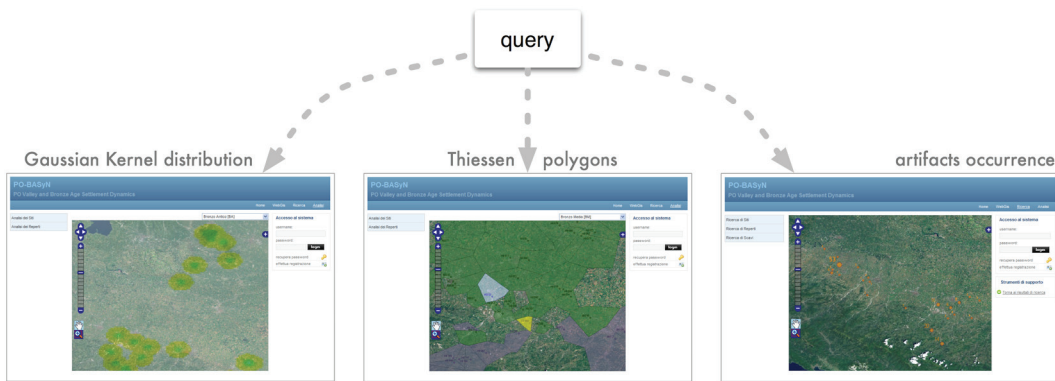


Fig. 3. Three examples of the analysis and thematic mapping capabilities of the system.

were able to build spatial queries over databases of the sites and artifacts.

Finally, we tested the integration of GRASS GIS in the system in order to verify the possibility of performing complex geostatistical computation. In this case we used a “CRON job”, i.e. the execution of a time-based service. Two small programs were written, which respectively elaborate the Gaussian Kernel density and the Thiessen polygons for each phase of the Bronze Age on the sites that are classified as “settlements” (Fig. 3). A CRON daemon each day at a specified hour triggers the two jobs, which are executed to produce a fresh elaboration. Then, in this case, the elaborations are not “on-the-fly”, but this can be a good workaround for the current situation, where data updating is not dramatic because it involves a few number of sites each day.

4. Conclusions and future developments

The paper presented the recent improvements in the development of an Open Source-based WebGIS system for the PO-BASyN project. Our experience confirms the effectiveness of this software development methodology with regards to some characteristics which we retain to be fundamental: the flexibility and modularity of the system and the possibility to fully customize and integrate different solutions into an evolving architecture. The system currently offers all that is necessary for our purposes, in particular with regards to the collaborative acquisition and to the sharing of archaeological data in the context of the Bronze Age in Northern Italy.

Moreover, beyond the considerations of strictly technological aspects, it is important to stress how the native support that Open Source software generally has with regards to open standards constitutes a crucial and distinctive characteristic. In fact, the easy management of data coded with open standard

formats is a central element in the scenario into which the PO-BASyN system finds its collocation, i.e. that of the geospatial web.

Many improvements can be made in geospatial data analysis, by refining the analysis tools that were experimented with and which offered satisfactory results both in terms of the system performances and the usefulness of the functionalities themselves. It is our purpose to define more precisely the kind of analyses that are considered valuable by scholars, in order to increase the effectiveness of the system.

Finally, research has already started for the integration of semantic aspects in the representation and retrieval of information. In fact, as the system grows and other institutions join the project and share their datasets, the problem of information integration becomes crucial. Each institution, or even each scholar, may use a different typology for artefacts and sites, which in turn means that there may be significant differences both in the vocabulary and in the conceptual schemas. An information representation that is more expressive and richer than the one currently offered by the entity-relationship model may help in integrating these different datasets. Regarding this, ontologies and Semantic Web technologies seem to be the ideal solution. Moreover, we aim to investigate the possibility of performing semantic retrieval on the dataset, in order to further increase the effectiveness of the system and to offer truly innovative tool for the study of the Bronze Age in Northern Italy.

References

de la Baujardiere, Jeff (ed.) (2006). OpenGIS Web Map Server Implementation Specification. Open GIS Consortium Inc. Online publication accessible at: http://portal.opengeospatial.org/files/?artifact_id=14416

- Bernabò Brea, Maria, Andrea Cardarelli and Mauro Cremaschi (eds.) (1997). *Le Terramare: la più antica civiltà padana*. Milan: Electa.
- De Marinis, Raffaele C. (1999). Towards a relative and absolute chronology of the Bronze Age in Northern Italy. *Notizie Archeologiche Bergomensi* 9, 23–100.
- De Marinis, Raffaele C. (1997). L'età del Bronzo nella regione benacense e nella pianura padana a nord del Po. In: Maria Bernabò Brea, Andrea Cardarelli and Mauro Cremaschi (eds.) *Le Terramare: la più antica civiltà padana*. Milan: Electa, 405–419.
- Gerlek, Michael and Matt Fleagle (2007). Imaging on the Geospatial Web Using JPEG 2000. In: Arno Sharl and Klaus Tochtermann (eds.) *The Geospatial Web*. London: Springer, 27–38.
- Mantegari, Glauco, Maurizio Cattani, Raffaele de Marinis and Giuseppe Vizzari (2007). Towards a Web-based Environment for Italian Prehistory and Protohistory. In: Jeffrey Clarke and Emily Hagemester (eds.) *Digital Discovery. Exploring New Frontiers in Human Heritage. CAA2006. Computer Applications and Quantitative Methods in Archaeology Proceedings*. Budapest: Archaeolingua, 374-381.
- Neteler, Markus and Helena Mitasova (2008). *Open Source GIS: A GRASS GIS Approach*. Third Edition, Springer: New York.
- Vretanos, Panagiotis (ed.) (2005). *Web Feature Service Implementation Specification*. Open GIS Consortium Inc. Online publication accessible at: http://portal.opengeospatial.org/files/?artifact_id=7176