AN OPEN SOURCE SYSTEM FOR P.I.C.A.
A PROJECT FOR DIFFUSION AND VALORIZATION OF CULTURAL HERITAGE

E. Agosto[(*)], D. Demarchi[(**)], G. Di Gangi[(***)], G. Ponza[(**)]

[(*)] DITAG, Politecnico di Torino, 10129 Torino, Italy - eros.agosto@polito.it
[(**)]I.S.I.Line Srl – Via Torino 73, 12037 Saluzzo (CN) - d.demarchi@isiline.it, g.ponza@isiline.it
[(***)]CeST Marcovaldo, Università di Torino - cest@marcovaldo.it

KEYWORDS : GIS, Internet/Web, Information System, Mapping, Patrimony

ABSTRACT

The presented work is the result of an integration of different informatics and cartographic instruments in a project named P.I.C.A. (Cultural Informatics Portal of the Western Alps - Interreg IIIA Alcotra) - a CeST-Marcovaldo project, in collaboration with Turin University and the Politecnico di Torino, which has begun in year 2000 - concerning the diffusion and valorization of the cultural heritage. Computer technologies were used to obtain a complete tool for browsing and researching multidisciplinary data, ranging from historic monuments and geological morphologies to anthropological studies. All these types of data are closely connected to the territory and had previously been structured in a GIS; the project itself was integrated in a transalpine scenario, to stimulate the sharing of data between transalpine researchers and institutions, and to link them to their presence on a common territory. The main technical goal was the design and setting up of a WebGIS site that, starting from thematic maps, allows the user to access the data of interest in an easy and efficient way. With these intentions, and considering the need of obtaining an open product that can be used not only by researchers but also by an heterogenic public, OpenSource technologies were considered to be the best solution to reach the different goals of the project. The cartographic data were stored in a Postgres-PostGIS DataBase Server: in this way, both geometric and alphanumeric information were stored in a unique tool. Data were mapped on the Web and an application that was based on University of Minnesota MapServer was developed. The Website was implemented on Apache Server, and a high level of integration was obtained between the data and geographic information. For search engine purposes, a tool named SDX was chosen. This is an OpenSource XML based application specifically developed for the management of documents. In P.I.C.A. the data of interest can be accessed by the user using these tools, through a search of thematic maps that mix chronological and topographic information, or through a more classical search engine based on keywords.

1.BACKGROUND OF THE PROJECT

The project is the continuation of a research project called "*The Marchesato di Saluzzo: an integrated study of the territory*", funded by M.U.R.S.T. (Ministry of University, Scientific Research and Technology), and which had the goal of creating an instrument not only for historical and archaeological research, but also for the conservation and the valorization of the Po Valley[1] landscape and property. This instrument was built to be used by researchers, local administrators who have to preserve the local cultural heritage, and by ordinary citizens who wish to use the land for tourism. In the last five years a great deal of data coming from different fields, such as archeology, geology, botany, and cultural anthropology, have been collected by experts in different and specific fields, and organized in a database. This great research effort required tools that could improve the understanding of the data and deepen the investigation potentiality of such a quantity of data. The preparation of a stand alone GIS made it possible to put together, to compare and to interpret archived data, material data and data from aerial photography interpretation. It is in fact belived that interdisciplinary is necessary to understand land transformations and the role humans play in this. In the activities performed before P.I.C.A. project was started, the researchers worked on many cultural aspects and the management of data was based on specific applications, developed for dedicated workstations and with offline usage. The extension of the system to new Internet technologies, with the developing of a dedicated Web site with the well known facilities obtained by a shared and networked repository of data, was a natural step. After the WebGis tests that were performed in 2003 it was finally decided to extend the application to a more solid architecture and more complete functionalities (in this Proceedings, see G. Di Gangi, C.M. Lebole, D. Demarchi, L. Nejrotti: *P.I.C.A. - Portale Informatico Culturale delle Alpi occidentali: a multidisciplinary and integrated project for cultural heritage*). The goal of the project was not only to share data, but also to exploit the powerful opportunities of Internet network to reach a larger number of users, that can be not only experts or researchers, but also people interested in cultural information for tourist purposes. With these premises, the site specifications were first focused on thinking an easy to use system that could make the user access detailed or less precise, but more general and addressed to cultural diffusion. For these reasons it was chosen to implement a platform that could satisfy both needs, and allow the user to deepen the quantity and quality of the information depending on the navigation choices in the Web site. One of the main specifications was that the user could access cultural information starting from a dynamic map on which entities of interest in different discipline fields were indicated, and where the discipline and the dates on which to work could be chosen. For this reason it was primary need to publish the GIS on the Internet. Basing the project conception on these requests, the OpenSource tools have been selected as to be the best solutions to achieve project goals. In fact with these technologies is possible to build complete and flexible solutions, using the continuous update of software packets and applications and being sure about the tools efficiency. Moreover, in a shared project as P.I.C.A., an open realization is an opportunity to improve functionalities and site contents thank to open cooperation with other partners.

---

[1] The Po Valley is one of the main valleys in Marchesato di Saluzzo, an area in the south of the Piedmont region, in the north-west of Italy. In the Middle Ages it underwent a period of prosperity, before becoming part of the property of the Savoy royal family.

2. WEBSITE AND SOFTWARE SYSTEM

2.1 The open source tools

In the wide software scenario of OpenSource solutions, we selected the applications that had to fundamentally solve three problems:
1. Web presentation
2. Map and data management
3. Powerful Search Engine

There are plenty solutions available on Internet, but we chose the most diffused and stable ones in a way to be always up to date and to reach best performances.

As Web Server was selected *Apache Server* (http://httpd.apache.org) that is the most used and powerful on Internet network. As indicated in the official Website, the Apache Project is a collaborative software development effort aimed at creating a robust, commercial-grade, featureful, and freely-available source code implementation of an HTTP (Web) server.

The site is based on Java technologies; *Apache Jakarta Tomcat* (http://jakarta.apache.org/tomcat) and the *Application Server Jboss* (http://www.jboss.org) have been used to manage Java code and pages. The Java 2 Enterprise Edition (J2EE) is a multitiered architecture for implementing enterprise-class applications and web based applications. This technology supports a variety of application types from large scale Web applications to small client server applications. The main aim of J2EE technology is to create a simple development model for enterprise applications using component based application model. In this model such components use services provided by the container, which would otherwise typically need to be incorporated in application code. J2EE applications are made up of different components. A J2EE component is a self-contained functional software unit that is assembled into a J2EE application with its helper classes and files and that communicates with other components in the application. The J2EE specification defines the following J2EE components:

- Application clients and applets are components that run on the client.
- Java Servlet and JavaServer Pages technology components are Web components that run on the web server.
- Enterprise JavaBeans components (enterprise beans) are business components that run on the application server.

The servlet container selected for the project is Jakarta Tomcat that is used in the official Reference Implementation for the Java Servlet and JavaServer Pages technologies. Jboss was selected as Application Server and it is the application server that implements the complete Java 2 Enterprise Edition (J2EE) stack, including Java Server Pages (JSP), servlets, and Enterprise JavaBeans (EJB).

MapServer (http://mapserver.gis.umn.edu), which is an Open Source development environment for building Web mapping applications, was the adopted solution to manage the cartographic system. In MapServer it is possible, through a CGI-based application, to deliver on-line, interactive GIS and image processing information quickly and reliably. This system description will be deepened in Section 0 below.

It was decided to adopt an XML description of documents based on the SDX system to optimize the search engine. SDX (System for Documents in XML) is an Open Source software which allows to retrieve and display XML documents. This Java platform is implemented on a servlet engine, either after compilation or as a deployable Web application.

Finally, the heart of the system, the DataBase, is the *Postgres SQL Server* (http://www.postgresql.org) that, as written in his official website, is a highly scalable, SQL compliant, open source object-relational database management system. With more than 15 years of development history, it is quickly becoming the de facto database for enterprise level open source solutions.
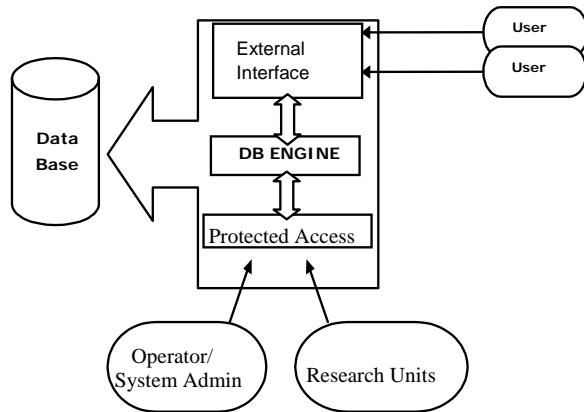
2.2 System architecture

Figure 1. System Architecture

The system architecture has the classical structure of a dynamic system, where are integrated public accesses for users Web pages navigation and protected access to allow administrators and operators to manage the data contained in the system.

As indicated in Figure 1 the Java Engine is shared by different accesses and is the interface with the DataBase.

Besides administrative operations, the site must be updated by researchers of different disciplines, to allow input and editing of the information and data contained in the system. This access is also protected by password and each researcher will have his dedicated password, to monitor activities and to associate data to researcher or association that input them to system, protecting their property.

Analyzing the system more in detail, it can be described as in Figure 2, where are indicated the logical connection between different software components
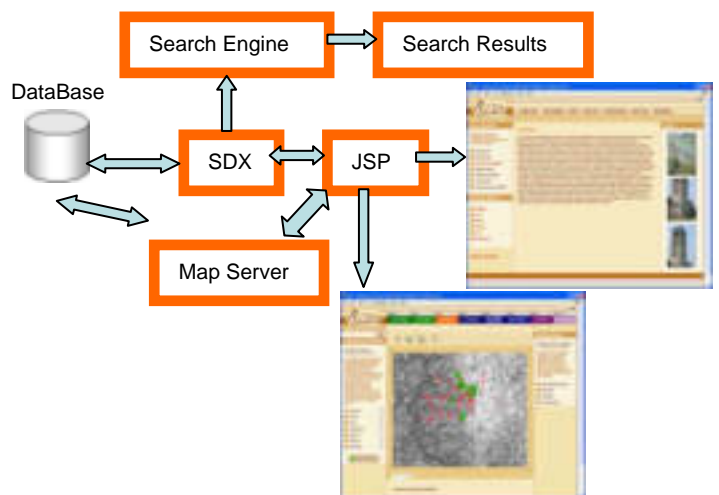

Figure 2. Software System

From DataBase, the data are extracted to be showed in the two main methods: the cartographic visualization and the card

describing the entity selected, as for example an historic monument.

As mentioned in paragraph 2.1, the management of data on the dynamic map are passed through MapServer system and then elaborated by Java Engine to be prepared for Web visualization as HTML page. Descriptive cards are generated by Java Engine also, but the repository of reference is based on SDX system, that is connected to Postgres Database. From SDX are then directly generated the pages that are result of searches introduced in Web site. In this case the functionality is similar to a classical textual search engine.

2.3 Web presentation and Site Flow

The navigation flow of a Website is fundamental for interest that can be received from users and an easy structure of the site is one of the most important requests from Internet users.

For this purpose, the project of Website structure was discussed and decided with the researchers of cultural part, with the goal to find the optimal specifications for navigation flow inside the Website.

To improve interest in the data contained in P.I.C.A. it was decided to develop a didactic section, accessible from any web page. As mentioned, the project is in fact addressed to general users also, not necessary experts of the disciplines contained in P.I.C.A.. For this reason several didactic pages have been implemented to explain terminology, keywords and the parts of the site that were selected to be more useful for schools or for simple tourist usage. Specific tourist information have been integrated to the description of the entities contained in P.I.C.A. (accessibility of cultural sites, opening times of monuments, …) and these information has been completed with simple explanations of P.I.C.A. contents.

Figure 3. P.I.C.A. Homepage

From the homepage (Figure 3) it is possible to choose the region where to work and to have access to the dynamic map where to select the entity of interest; from the homepage it is also possible to insert a specific research and to directly use P.I.C.A. search engine.

On the map it is possible to use all the powerful facilities of MapServer that will be described in section 0 below and to choose the entities on which to have specific information.

The card that describes for example a monument (Figure 4) has been projected to have basic information in the first page that is visualized to user, but it is possible from this first page to access more specific data connected to entity. These information can be both tourist and technical for expert people (researchers) usage. In this manner is reached one of the main goals of

P.I.C.A., the possibility to have in the same time and with the same access different types of information.

Figure 4. Entity Description Page

In upper left part of the page it is possible to choose to enter in these more specific information and the sections are named for example *Documents, Photo Gallery, Specific Analysis, Tourist Information, ...*

In the lower left section of monument card it is possible to access to didactic pages (Figure 5) that explain arguments connected to the entity described in the card. An interesting characteristic of these pages is that have been implemented using a simple language and a large use of illustrations has been done, in particular using picture made by hand with drawings technique, obtaining an appealing presentation of arguments.

Figure 5. Didactic Page

3.THE CARTOGRAPHIC DATA

3.1 The mapping engine: Mapserver

The chosen software to deliver cartographic information to the web is the University of Minnesota Mapserver.

MapServer is an Open Source development environment used to build spatially-enabled internet applications; it is not a full-featured GIS system, but excels at rendering spatial data (maps, images, and vector data) suitable for the Web. It was originally developed by the University of Minnesota (UMN) for ForNet project in cooperation with NASA and the Minnesota Department of Natural Resources (MNDNR). Beyond browsing GIS data, it allows the creation of "geographic image maps", in other words maps that can direct users towards the content. For example, the Minnesota DNR Recreation Compass provides users with more than 10,000 web pages, reports and maps via a single application. The same application serves as a "map engine" for other portions of the site, providing spatial context where needed. Its main features include:

- Advanced cartographic output
- Scale dependent feature drawing and application execution
- Feature labeling including label collision mediation
- Fully customizable, template driven output
- TrueType fonts
- Map element automation (scalebar, reference map, and legend)
- Thematic mapping using logical- or regular expression-based classes
- Support for popular scripting and development environments: PHP, Python, Perl, Ruby, Java, and C#
- Cross-platform support: Linux, Windows, Mac OS X, Solaris, (and more)
- Map projection support * On-the-fly map projection with thousands of projections through the Proj.4 library.

Many factors were taken into account when choosing the Web mapping software: the first was the possibility of using preconstituted useful functions; the second one was the possibility of personalizing the software, to solve problems connected to the project. What is usually called an API (Application Programming Interface) was needed in order to interface with the software. The last factor was the economic aspect of the project. It was necessary to reduce the construction costs because they can be an important factor, in particular in cultural conservation where sponsors are not so common, or funds are only available for limited periods of time or for specific projects. Choosing a commercial product also makes you dependant on the chosen solution: the terms and conditions of the software producer have to be accepted, and they do not always meet the users' requirements. The users are tied to the expensive software updating and new releases; everybody can surely recall the decision made by ESRI to stop supporting Avenue language, forcing thousands of users to rewrite their applications in Visual Basic for the new version of the GIS software.

The Politecnico di Torino Geomatics group has already used this software in different research projects and have had the chance to appreciate its reliability; being an Open Source product, it also matched the P.I.C.A. philosophy and requirements: low developing and maintenance costs, many available functionalities, flexibility and freedom to integrate it in a customized application.

The wide range of supported data formats was also highly appreciated; on the raster side almost all the used formats can be handled via GDAL libraries, whereas on the vector data side, both traditional formats such as ESRI shapefile both spatially enabled database tables, such as Postgres-Postgis Server or ESRI ArcSDE, or others, can be achieved via OGR.

3.2 The unique reference system problem

The researches group needed cartographic products to map and study the culturally interesting territorial sites of the P.I.C.A. area. National Cartographic systems maps were obviously initially used to georeference the information they had. In the early steps of the project the Italian working team started to develop a Gauss-Boaga based GIS project, whereas on the French side, the team developed a NTF Lambert II etendue one. The cooperation between the research groups to plan the documentation database structure had also to face the problem of different cartographic bases.

As both National Cartographic systems have reliable conversion tools toward WGS84, and considering recent INTERREG II choices in the production of border excursion maps, and that GPS surveys would have been widely used in the subsequent project phases, the cartographic team suggested choosing WGS84 as the P.I.C.A. reference system.

On the Italian side the IGM[2] Verto2 software was used in the conversion process, whereas on the French side, commercial GIS tools were being used.
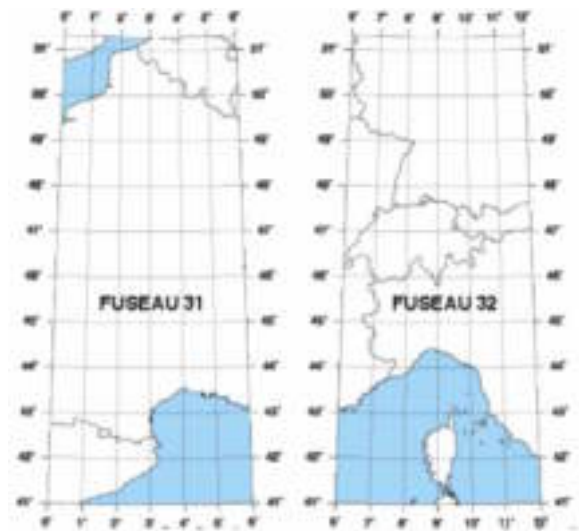

Figure 6. The UTM zones involved in the P.I.C.A. project

A dummy projection system was used to display cartographic maps on the Web: WGS84 geographic coordinates were projected using a fictitious UTM zone (between UTM zone 31 – 32) with a 6° longitude central meridian. All these transformations were made using software procedures based on Hirvonen formulas on the Geometry column of the Postgres - Postgis Database. This was possible as this spatially enabled Database uses the Well-Known Text (WKT) formats to store geometric information.

3.3 The data

In the project, the geometric data are fundamental key, P.I.C.A. project couldn't exist without; there was a considerable amount and different kinds of spatial data and these will increase as the project itself develops; they are made up of:

---

[2] Istituto Geografico Militare, one of the main Italian Cartographic Institutes

- Punctual, linear and areal layers, mainly representing interesting site locations; these information comes from the field and archive investigation activities of each specific disciplinary research unit. They were initially created as shapefiles in a desktop GIS project, and eventually dumped into Postgres - Postgis tables.
- Raster cartographic maps, both RGB (IGN3 French 1:25.000 maps), and B/W ones (Piemonte CTR 1:10.000); these data come from national or local cartographic institutes. They are used as the project mapping background, and also in the georeferentiation process.
- Vector cartographic maps (Piemonte 1:10000 CTRN), used to display map information at lower visualization scales, to avoid long waiting times for web users due to the loading of several raster files.

It was decided to store the vector data in a GIS enabled Database; the Postgres-Postgis database server was chosen, as it is greatly appreciated by the users. This tool proved to be very useful and efficient in the development of the project. It was possible to store all the data in a unique tool, and to use it in the dummy UTM zone projection operation; the integration with Mapserver was very successful and easily obtained. One of the main reason for choosing the DB was connected to the possibility of having spatial indexes available in view of the foreseen increase in data during the future P.I.C.A. activities. This choice made it possible to manage the web user choices in a simple way, exploiting SQL queries potentiality and demo file configuration parameter changes via URL.

Figure 7. A Piedmont CTR and PICA data view

Raster data were chosen to store as Tiff files: this means high amounts of disk space, but JPEG2000 usage tests gave very poor performance results in web displaying time. Visualization scale limits and tile indexing tests were made to improve the display performance of heavy raster data. It was decided to use Mapserver in the normal CGI mode and to develop a custom application to feed it with proper URLs. The user interface is based on the JBOX tool available on the UOM Mapserver website, in order to offer the user advanced rubber band functionalities together with the common punctual pan, zoom query ones.

---

[3] IGN, National French Cartographic Institute

Figure 8. A 1:25000 IGN map view

4. CONCLUSION

P.I.C.A. project has been the result of different expertises that produced a complete Web tool for simply usage of a GIS system. Actually the technical system is setup and the future works will be directed to improvement of interfaces, but more to the input of new data to system. In this first version some disciplines (archaeology, history, anthropology) have been considered, but in project prosecution the aim is to add new ones (botany and geology); the necessary tools to fit the system to the requests of new disciplines will be implemented.

REFERENCES

Agosto E., Cosenza A., Rinaudo F., The "Marchesato di Saluzzo" project: an OpenSource based Web/GIS for historical and cultural research, conservation and valorization, Proceedings of CIPA XIX International Symposium, Commission V, WG5, Antalya, 2003.

Boccardo P., Dal Molin M., Rinaudo F., Spanò A., Il sistema informativo interdisciplinare per l'analisi del contesto storico territoriale del Marchesato di Saluzzo, in Di Gangi, G., Lebole C.M., ed., 2003, *Leggere il territorio. Metodi d'indagine e finalità a confronto*, Atti del Colloquio Nazionale, Saluzzo 15-16 novembre 2002, Caraglio, pp. 157-180

Ardissone P., Rinaudo F., Geomatic applications in archaeological research: the middle age in the valle po landscape, Proceedings of CIPA XIX International Symposium, Commission V, WG5, Antalya, 2003.

Di Gangi, G., Lebole C.M., ed., 2001, *La gestione del Territorio: memoria, partecipazione, sviluppo della ricerca,* Atti del Convegno Nazionale, Saluzzo 11-12 novembre 2000, Dronero.

References from websites

http://httpd.apache.org
http://jakarta.apache.org/tomcat
http://www.jboss.org
http://www.ign.fr
http://www.igmi.org
http://mapserver.gis.umn.edu
http://ms.gis.umn.edu
http://refractions.net