FUSION of CULTURES

Abstracts of the XXXVIII Annual Conference on Computer Applications and Quantitative Methods in Archaeology, CAA 2010

F. Javier Melero · Pedro Cano · Jorge Revelles

Editors

FUSION OF CULTURES

Abstracts of the XXXVIII Conference on Computer Applications and Quantitative Methods in Archaeology

Fco. Javier Melero, Pedro Cano & Jorge Revelles (Editors)

Granada, Spain

April 6-9, 2010

Spatial Data Infrastructures and Archaeological Excavation Data: *SILEX*, the SDI of the Neolithic Flint Mine of *Casa Montero* (Madrid, Spain)

Fraguas, A., Menchero, A.,[1] Uriarte, A., Vicent, J., Consuegra, S., Díaz-del-Río, P., Castañeda, N., Criado, C., Capdevila, E, Capote, M.

Research Group "Prehistoria Social y Económica", CCHS-CSIC, Spain
{alfonso.fraguas, antonio.uriarte, juan.vicent, susana.consuegra, pedro.diazdelrio, nuria.castanyeda, cristina.criado, enrique.capdevila, marta.capote}@cchs.csic.es, [1]antonio.menchero@gmail.com

1. Introduction: the flint mine

SILEX (*Sistema de Información Locacional en XML*, in English *Locational Information System in XML*) is an information system, under the form of a Spatial Data Infrastructure (SDI), designed for the systematic management and distribution of data recovered through the excavation of a particular archaeological site. SILEX has been developed by the Research Group *Prehistoria Social y Económica* (Instituto de Historia, CCHS - CSIC), one of whose lines of research is the application of Information and Communication Technologies (ICTs) in Archaeology.

Casa Montero (DÍAZ-DEL-RÍO et al. 2006, CAPOTE et al. 2008) is an Early Neolithic flint mine (c. 5300 B.C.) located in the eastern side of the municipality of Madrid (Spain). It was found during the archaeological monitoring works preceding the construction of the M-50, the third highway belt surrounding Madrid. Due to its uniqueness and relevance, the historical heritage service of the regional government of Madrid forced the modification of the route, preserving part of the site and excavating the one that would be finally affected by the highway.

The archaeological excavation was carried out in three campaigns between September 2003 and August 2006 and produced a complex and large data set. In a 4 hectares surface, almost 4000 mining shafts were discovered, 324 of which were sampled and excavated. These shafts contained 3375 different archaeological deposits.

Most of recovered archaeological items are flint artifacts that are the waste products of the extraction and first knapping of nodules. Their total weight amounts to 65.885 kg. 173.585 items have been sampled and analyzed. The archaeological analysis of these artifacts has allowed the reconstruction of the operational chain of flint exploitation by Neolithic communities. Among other finds are pottery, bone industry and quartzite maces.

The abundance and complexity of the archaeological data suggested the need for a precise and systematic methodology, as well as an efficient information management system.

2. The database

The archaeological data contained in SILEX has a double dimension: thematic and spatial.

An Entity-Relationship (E-R) data model (CHEN 1976) has been used for thematic data. Information is stored in a native XML database. XML has been the preferred language code, because it is widely accepted as a data exchange standard between different systems. As usual in native XML databases, information is structured in collections, subcollections and documents.

SILEX has a complex data structure, considering around 40 different kinds of archaeological entity types. The heart of this network is articulated by three main entity types:

1) Shafts: They are the result of digging the soil for finding and extracting flint seams. According to Harris Matrix terminology (HARRIS 1989), they are *negative stratigraphic units*, that is, the result of removing sediment.

2) Deposits: They are the result of throwing the soil

and waste products of flint knapping and other materials back into the shafts (Figure 1). They are *positive stratigraphic units*, that is, the result of adding sediment.
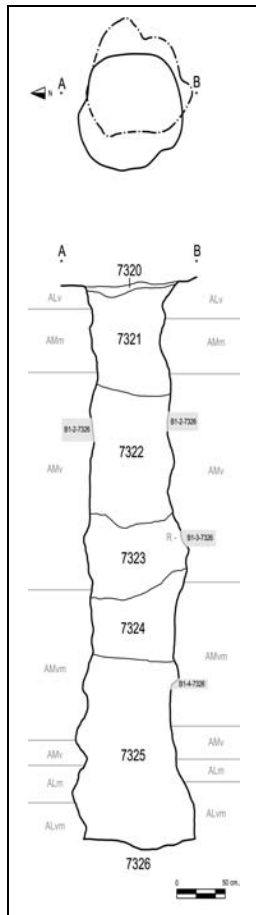

Figure 1: *Archaeological section of a shaft and its deposits*.

3) Archaeological items: They are items recovered from deposits. Most of them are flint, for whose description, based on the logical analytical system (MORA et al. 1992), an important part of the database is devoted.

Spatial information relies mainly on shafts. These are the basic spatial entities, and the other (deposits and archaeological items) are spatially referenced through them. GIS vector model has been used for representing mining shafts, extracting vector spatial entities from the CAD planimetry of the excavation (Figure 3). Vector information is stored in an Object-relational DBMS with spatial module. Additionally, there are several raster layers with geographical information about the surrounding landscape: geology, land use, orthophoto and so on. These are stored as GIS files in the directory structure of the web system server. Finally, metadata relative to spatial information is stored in the same DBMS that vector layers.
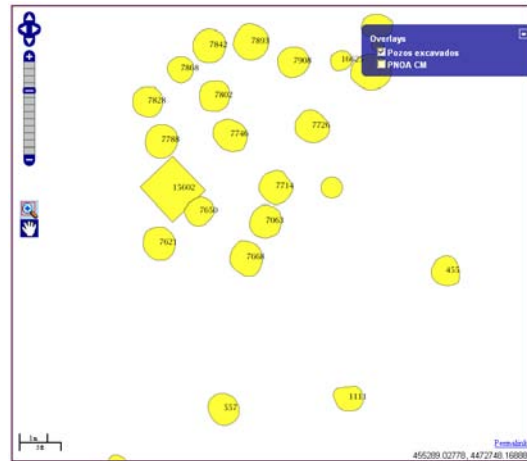

Figure 2: *Visualization of the shafts layer through WMS.*

Archaeological entities, both stratigraphic units and items, have abundant multimedia information linked to them, both textual and graphic, which is stored in the directory structure of the web server.

3. The distributed architecture

As a SDI, SILEX is a distributed information system. The use of web technologies allows the dissemination of information in a friendly way, for it uses Internet architecture itself. So, each resource has its own universal identifier (URI, *Uniform Resource Identifier*) (BERNERS-LEE 1996). Moreover, each spatial resource is accessed by means of calling a remote procedure (RPC, *Remote Procedure Call*), which is located in a URI. The use of URIs as pointers for accessing information makes SILEX a real graph-oriented database, a net of information nodes connected by edges or relationships through a distributed system (Figure 2).
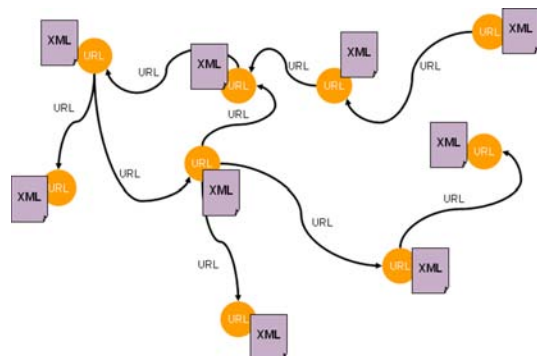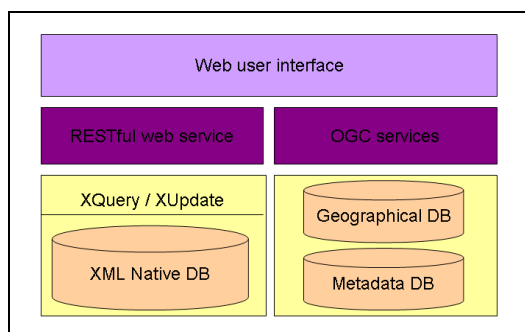

Figure 3: *Graph representing relationships between entities.*

SILEX has a typical multilayer architecture, composed by three levels (Figure 1): data storage or

persistence layer, web service (for accessing data), and web user interface.

Figure 4: *SILEX web architecture*

The web service used for thematic information turns read-and-write requests into XQuery/XUpdate requests to the native XML database via HTTP. In this way, the thematic web service can use any database as long as it supports XQuery 1.0, XML:DB and XUpdate over HTTP.

Standards, both *de facto* and *de iure*, play a crucial role (INSPIRE, REST, OGC and so on). SILEX is organized as a Mashup that combines several web resources. According to previous SDI prototypes (FRAGUAS 2008), we have used a mixed software architecture that uses web services with RESTful design criteria for thematic data and OGC services (CSW, WMS, WFS, WCS) for spatial data.

Thematic data is accessed through services with a Resource-Oriented Architecture (ROA), consisting of a set of guidelines (RICHARDSON et al. 2007) for the implementation of the REST architecture (FIELDING 2000). According to its designers, ROA has four advantages: 1) *addressability*: the web service exposes at least one URI for each information resource; 2) *connectedness*: the requested representation includes URIs from other resources and forms with which one can build another URIs, so that the client can discover new information just following the links; 3) *statelessness*: each request to the web service happens in isolation to others; when the client makes a request, this includes all the information necessary to generate a reply, that is, the server does not require additional information about preceding requests, but if it was necessary this information would be provided again by the client; 4) *uniform interface*: operations that can be executed are the same for any resource.

Regarding spatial information, we have followed the INSPIRE Directive (*Infraestructure for Spatial Information in Europe,* European Union Directive 2007/2/CE, March 14th 2007), technically implemented by means of SOAP services created by the Open Geospatial Consortium (OGC). It is thus a Service-Oriented Architecture (SOA).

According to INSPIRE recommendations, SILEX allows the integration of detailed archaeological information layers into local, regional and global SDIs. That is, services and resources used by SILEX can be included as a node in a wider SDI.

4. Free Software and Open Source

One of the driving principles of SILEX is the use of Free Software and Open Source: sharing knowledge helps to the development of science and society in an equitable and enrichful way.

The software packages used are the following:

1) Apache Tomcat, as a servlet container for mounting web applications.

2) GeoServer, as a server for providing geographical information layers via OGC standards such as WMS, WFS and WCS.

3) OpenLayers, as a Javascript library for showing dynamic maps in web pages.

4) GeoNetwork, for offering metadata according the OGC's CSW specification.

5) PostgreSQL and PostGIS, for storing and recovering maps and metadata.

6) Apache Cocoon, as a web publishing environment based on XML, for designing web applications, both the web service and the web interface for accessing thematic data.

7) Orbeon XForms, for converting XForms into cross-browser forms based on HTML and Javascript.

8) Saxon, a XSLT transformation motor, for converting XML documents.

9) eXist, a native XML database, for storing thematic data, with a query motor based on XQuery.

5. Conclusions

SILEX defines a data model for describing and organizing the complexity of different types of archaeological entities (deposits, shafts, lithic artifacts an so on) and the relationships between them. It includes two well established archaeological "ontologies", as the Harris' system for describing and relating stratigraphic units and the logical-analytical system for classifying lithic industry. The data model has been implemented by means of a XML based format.

On the other hand, SILEX is a distributed system that offers this information via Internet. The web framework allows to share data, so that archaeologists may publish information produced by themselves or, inversely, retrieve that created by others and analyze it with their own tools and criteria. A web admin

application has been developed to edit, save and publish archaeological data through the web browser; that way archaeologists do not have to know the internals of the publishing framework or XML format and do not need to install proprietary applications. Moreover, a web query application allows archaeologist to query the full data set and retrieve the subset they are interested in. Using free software and open source tools, we have built a Mashup that gives access to geographic information through OGC services, following OGC standards based on EC Directive INSPIRE.

SILEX has been presented in the *2nd International Conference of the UISPP Commission on Flint Mining in Pre- and Protohistoric Times*, celebrated in Madrid in October 14th-17th 2009 and is accessible in the URL *http://www.casamontero.or*g/.

SILEX will be integrated into more general SDIs, such as the *IDEE*, the Spanish SDI for institutional geographical information, and the *IDE-CSIC*, the SDI designed for managing and serving spatial information generated by scientific projects by the *Consejo Superior de Investigaciones Científicas* (CSIC), the Spanish research council.

Although SILEX is devoted to a specific site and designed according its particular features, it can be useful as a SDI model for new archaeological site information systems. Because of its flexibility, it can be developed to incorporate other levels of archaeological information, such as new sites.

References

BERNERS-LEE T., 1996. Universal Resource Identifiers - Axioms of Web Architecture. http://www.w3.org/DesignIssues/Axioms.html

CAPOTE M., CASTAÑEDA N., CONSUEGRA S., CRIADO C., DÍAZ-DEL-RÍO P., 2008. Flint Mining in Early Neolithic Iberia: A Preliminary Report on 'Casa Montero' (Madrid, Spain). In Allard P., Bostyn F., Giligny F., Lech J. (eds.) Flint Mining in Prehistoric Europe. Interpreting the Archaeological Records, pp: 123-137. BAR International Series 1891, Oxford.

CHEN P. 1976: The Entity-Relationship Model - Toward a Unified View of Data. *ACM Transactions on Database Systems* 1 (1), pp: 9-36.

DÍAZ-DEL-RÍO P., CONSUEGRA S., CASTAÑEDA N., CAPOTE M., CRIADO C., BUSTILLO M. A., PÉREZ-JIMÉNEZ J. L., 2006. The Earliest Flint Mine in Iberia. *Antiquity* 80 (307). http://antiquity.ac.uk/ProjGall/diazdelrio/index.html

FIELDING R.T., 2000. Architectural Styles and the Design of Network-based Software Architectures. University of California Press, Irvine (CA).

FRAGUAS A., 2008. The ARANO SDI: A Spatial Data Infrastructure for the Rock Art of Northeast Africa. *Archaeological Computing Newsletter* 68, pp: 1-8.

HARRIS E. C., 1989. Principles of Archaeological Stratigraphy. Academic Press, London & New York.

MORA R., MARTÍNEZ J., TERRADAS X. 1992. Un Proyecto de Análisis: El Sistema Lógico Analítico (SLA). In Mora R., Terradas X., Parpal A., Plana C. (eds.) Tecnología y Cadenas Operativas Líticas, pp: 173-199. Bellaterra.

RICHARDSON L., RUBY S., 2007. RESTful Web Services. O'Reilly, Sebastopol (CA).