ARCHEOGIS: AN INTEROPERABLE MODEL FOR ARCHAEOLOGICAL DATA

Maria Antonia BROVELLI, Andrea MAURINO

Politecnico Di Milano - Facoltà di Ingegneria - Campus Como - ITALY
maria@ipmtf4.topo.polimi.it
maurino@morgana.elet.polimi.it

KEY WORDS: data models, GIS, interoperability, archaeological metadata

ABSTRACT
There are some open questions when an archaeologist wants to create a GIS application. First of all the choice of the hw/sw platform: each GIS has its own format to store data and metadata and the different packages use different algorithms to process the information. Secondly, the user should be so expert to be able to create a specific GIS application, because often the package must be customized in order to answer to the user requests. Finally, the last aspect is the availability of an archeological GIS application on the Web in such a way to have the possibility to access archaeological documents and maps from a remote host. Unfortunately, archaeologists are usually not good computer literate and the software complexity doesn't urge them to use GIS.
This paper presents ARCHEOGIS, an interoperable XML-based model allowing to describe archaeological data. This model has been designed in order to help the archaeologists in creating and managing their GIS application and in easily distributing data.

1 INTRODUCTION

GIS is becoming one interesting and useful tool for archaeologists to archive, manage, analyze and display data. However the power of this tool can be exploited only if its functions are correctly understood. An interesting research, in the form of survey, carried out within the Internet, was made by K.Gourad (Gourad, 1999) showing the main characteristics of people working within the archaeological environment and using GIS. From this research it appears that the preference, in choosing the software packages, obviously goes to programs with user-friendly interfaces instead of packages based on line-commands (even if GRASS is used by 20% of the participants to the survey). A second interesting result is that the surveyed people regard as the most problematic operations the collection, conversion and the consistence of the data. Finally the last point we want to outline is the sources of data: manual digitizing takes the first place among the procedures to collect data, while at the second place is the downloading of existing databases from the Internet and the subsequent conversion into a specific GIS readable format.
The conclusion we can derive from this survey and from inquiries we have carried on in archaeological environment is that the archaeological world has became curious and interested in these new technologies. The main problems one has to face are

- the difficulty in using software packages sometimes very unlike and complex;
- the difficulty in integrating varied geodata, in different formats, coming from different sources, in different reference systems,…

The answer to these problems will be the availability of an interoperable geoprocessing technology and open models. The work presented in this paper represents wants to be an effort in this direction.

2 THE LEADING CONCEPTS

2.1 The data model

The first problem we have to face, if we want to design an archeological model, is that the archaeological data, which can be defined as the evidence about the past are, too different from one other to be considered all together (anthropic or natural facts, sites, features, anecdotes, historical texts, …). At the same time and consequently the information about them are not homogeneous (usual/aerial/digital/near infra-red photographs, remote sensing images, drawings, plans, sections, contour surveys, paper/digital maps, CAD layers, GIS databases, field notebooks, paper correspondence, gazettes, dictionaries, books, cultural norms and reference values, digital texts, web pages, paper/electronic journals, e-mail, …) In order to manage all these data, we first of all decided to separate them into different information layers

based on the scale at which the feature is usually represented both from the mapping and geographical point of view. Particularly we took into account the layers: stratigraphic level, quadrant, excavation, site, park.
The layers may be hierarchically represented, according to the scheme shown in Figure 1.
In each layer the geodata model follows the specification proposed by the OpenGIS Guide (K. Buehler and L. McKee, 1998) and the Geographic Information European Prestandards and CEN Reports, (ENV 12160, 1997) (ENV 12656, 1998) (ENV 12657, 1998)
The formal language used to describe the layer features is XML (Bray, 1998) and particularly the SVG specification (Ferraiolo, 1999).

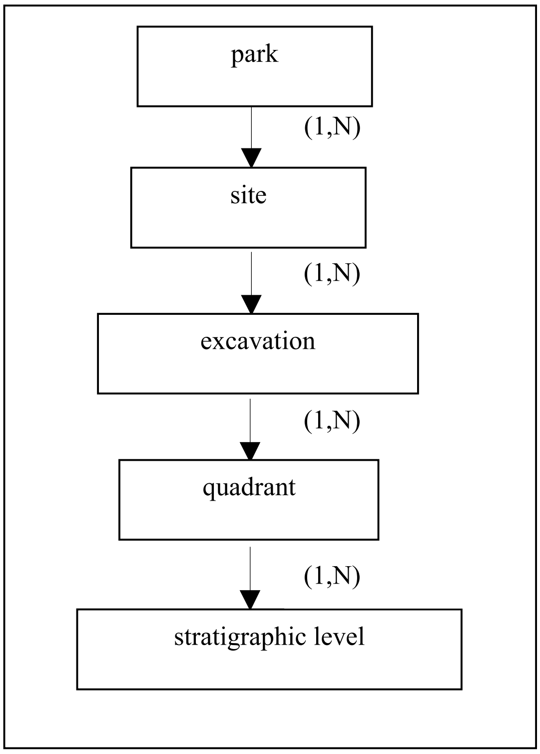

Figure 1. The hierarchical structure.

2.2 The metadata

The term 'metadata' means 'data about data', that is all the information we need to sensibly describe the data we manage or we want to. The information supplied with the data can be distinguished in:
Information about the nature of the feature;
Information about the location of the feature;
Information about similar features or features which can be (for some reason) grouped or linked to the feature considered.
The richness and completeness of metadata is the best strategy to manage large quantities of available distributed information: sometimes the unavailability of information about the available data corresponds to the fact that they become unusable. In this sense, it is evident the interest among the scientific community in defining standards of metadata.
But, as in other application fields, the adoption of standards of metadata in archaeology is not simple. The first problem is the absence of an ISO metadata standard for the geographical information. The second problem is about deciding if it is better in the archaeological world to agree upon a standard general metadata system, always extremely detailed, such as the Federal Geographic Data Committee's (FGDC, 1994) Content Guidelines for Digital Geospatial Metadata or the European Prestandard ENV 12657:1998 for the Metadata of the Geographic Information or to use a simpler and more specific metadata set.
As these questions are yet not solved and the archaeological community has so far not adopted a single set of standard, we propose (at least for the Italian case) to merge the metadata proposed by the CEN 287 and the information usually required in the card catalogue used by the Italian Central Catalogue and Documentation Institute (ICCD - Istituto Centrale per il Catalogo e la Documentazione - Ministero dei Beni Culturali e Ambientali - Soprintendenza Archeologica di Roma). (F.Parise Badoni, M.Ruggeri Giove, 1984)
Following this approach we obtain a set of metadata for each informative layer. As an example we show in the follows tables part of the metadata set proposed for the stratigraphic level (Table 1: Identification; Table 2: Overview, Table 3: Dataset Quality Elements)

METADATA	**CEN 287**	**ICCD**	**DESCRIPTION**
IDENTIFICATION			
Dataset Title	x		Name of the stratigraphic level
General Catalogue Number		x	Number used to uniquely identify the level. The number must be assigned by the expert office (Local Archaeological Service)
International Catalogue Number		x	Number used internationally to uniquely identified the resource.
Number		x	Number given to the stratigraphic level by the excavation director.
Origin		x	A first hypothesis about the origin (natural or anthropic) of the level.

Table 1. Identification of the Archaeological Geodata Stratigraphic level

METADATA	CEN 287	ICCD	DESCRIPTION
OVERVIEW	x	x	
Summary:	x	x	A textual description of the stratigraphic level. The item must follow the descriptive scheme proposed by the ICCD.
Producer Organization Name:	x	x	The name of the organization (Local Archaeological Service office) responsible for the resource.
Spatial Schema Type:	x		User defined scheme: the geometric features available in XML-SVG have been used. The topologic features are not present.
…			…
Intended Application Scale	x		Value of the intended application scale.
Interpretation		x	Description of the function of the level, the relationship to other stratigraphic levels both in spatial and in temporal scale.
Related Dataset	x		Title, owner organization, brief description of other datasets of possible interest for the user.
Related Datasets: maps (scale 1:10 - 1:20)		x	Catalogue number of the analogic maps representing the level.
…			…
Related Datasets: photos		x	Catalogue number of the photos (aerial, terrestrial, …) representing part or the whole level.
…			…

Table 2. Overview of the Archaeological Geodata Stratigraphic Level

METADATA	CEN 287	ICCD	DESCRIPTION
DATASET QUALITY ELEMENTS	x	x	
LINEAGE	x	x	
Producer Organization Name	x	x	The Organization(s) responsible for the creation of the original resource (i.e.:responsible for the excavation).
Process Method (Sampling)	x	x	Description of the sampling methodology.
Process Organization Name	x		…
Process Organization Date	x		…
Process Method (Flotation)	x	x	…
Process Organization Name	x		…
Process Organization Date	x		…
Process Method (Sifting)	x	x	…
Process Organization Name	x		…
Process Organization Date	x		…
Process Method (Digitizing)	x	x	…
Process Organization Name	x		…
Process Organization Date	x		…
Process Method (Dataset XML creation)	x	x	…
Process Organization Name	x		…
Process Organization Date	x		…
…			…
Quality Text	x	x	Textual description of the dataset quality.
…			…

Table 3. Dataset Quality Elements of the Archaeological Geodata Stratigraphic Level

2.3 The use of XML

Several question must be answered when an archaeologist wants to create a GIS application. First of all the choice of the hw/sw platform: this happens because each GIS has its own format to store data and metadata and the different packages use different algorithms to process the information. Secondly, users must be so expert to be able to create a specific GIS application, because the package must be often customized in order to answer to the user requests. Finally, the last aspect is the availability of an archeological GIS application on the Web in such a way to have the possibility to access archaeological documents and maps from a remote host. Unfortunately archaeologists are usually not good computer literate and the software complexity doesn't urge them to use GIS.
ARCHEOGIS is an interoperable model for the description of archaeological data. With this model it is possible to describe, by always using the same syntax, the geographical and contextual data, the metadata and the query over data. The syntax that we generated is XML-based.
XML is a subset of the Standard Generalized Markup Language (SGML); it is defined in ISO standard 8879:1986 and it is designed to make the interchange of structured documents into the Internet easier. XML files always clearly mark where the start and the end of each logical part (called element) of an interchanged document occurs. By defining the role of each text element in a formal model, known as a Document Type Definition (DTD), the XML users can check if each document component occurs in a valid place within the interchanged data stream. An XML DTD allows the computer to check, for example, if users do not accidentally enter a third-level heading without having first entered a second-level heading (this cannot be checked by using the HyperText Markup Language (HTML) which was previously used as the only language to code the documents that are part of the World Wide Web (WWW) documents accessible through the Internet).
We made a set of DTD documents to model a general archeological site by using, as suggested in the researches and applications pointing to the interoperability and to the Open GIS criteria, collections of simple features (described by the three basic elements: geometry, semantic properties and metadata). In this way archaeologists can write an XML file to describe the specific site they are interested in by using the DTD general document.
We used SVG (a WWW consortium XML-based standard) to represent vector data. In this way context data and geographical data are describe in the same languages. Again, a lot of companies are going to create a plug-in to show SVG files on the web; so data modeled by using this language will be put on Internet in a very simple way.
The choice of an XML language to describe our model is justified by two more important advantages: the first one is that a lot of tools that let you manage XML files and DTD using graphical user interface can be found. The second one is that there are a lot of proposals to create query languages for XML files (Marchiori, 1998).
All these features make an XML-based model a good way to increase the interoperability at design level.
After the ARCHEOGIS instance generation, in a semiautomatic way, the XML file can be translated on a particular hardware/software platform. This methodology increases interoperability because archaeologists can port their data over multiple platforms without changing anything.

3 SOFTWARE ARCHITECTURE

In order to support the development process we define the software architecture. This architecture is shown in the figure 2

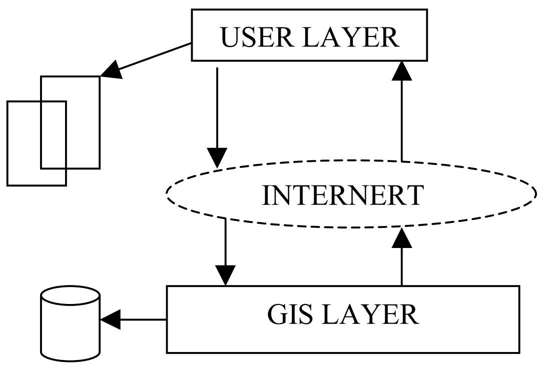

Figure 2. The software architecture

User Layer is devoted to define the user interface, show data, write data and operations in XML-format and send data and operations to GIS Layer.

The GIS Layer hides a particular GIS engine to show User Layer a unique point-of-view. GIS layer has many purposes: translating AVD textual part into a particular DBMS, translating AVD graphics part into a specific format used by GIS engine, translating AVD operation into a particular script program language.
Other purposes are capturing output form GIS engine and translating it into AVD format. In the following section details will be given on the User layers

3.1 The User Layer

The User Layer shown in figure 3

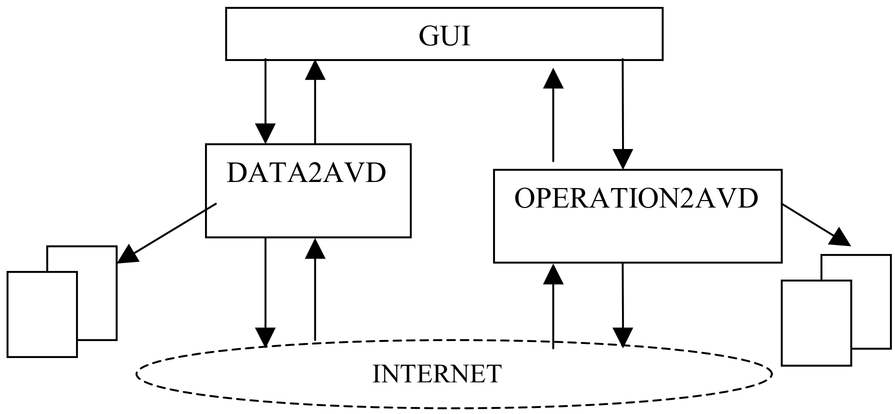

Figure 3. The User Layer

There are many blocks, which have different purposes. The GUI block permits to show a friendly user interface in which the user can insert data and operation and where he/she can see the output generated by the GIS engine.
The DATA2AVD module permits to translate (graphical and textual) data into XML-format. This module saves those data into the user's file system. The Operation2AVD module has a similar functionality.
The GIS engine can be changed by changing the GIS Layer. But since data are stored into the user's file system, too, they can be moved anywhere without problems
When GIS Layer sends new data, as a result of some operations, these data are translated from AVD into graphical way, they are stored into a file system and are shown by the GUI interface.

3.2 GIS Layer

The most important elements of the GIS layer is shown in Figure 4.

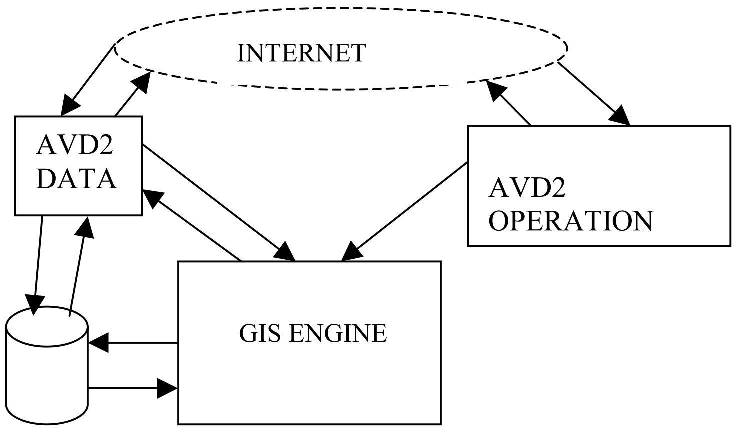

Figure 4. The GIS layer

The AVD2DATA block translates the AVD textual part into a specific DBMS, while the graphical part is translated into the specific format used by the GIS engine.

The GIS layer is intended to keep in a consistence state the server Database with the client file system archive. AVD2OPERATION block translates all operations into a particular GIS engine by writing scripts. For each GIS available we must write both blocks present in Figure 4.

4 THE AVD MODEL

The AVD model (ARCHEOGIS Vector & Data) was designed trying to define an abstract model to represent a typical archaeological park. It is very difficult to define a model with a high level of abstraction, which contains all aspects of archaeology, because this field covers thousand of years. So we considered AVD as a first step towards this goal, while we work on this problem to tune it in a better way.
The AVD model is split into three main parts

- **Cartography.** In this part we describe the cartographic part of an archaeological GIS. Here metadata, layers and all graphical information needed to show entities are defined. These informations are described by SVG element.
- **Textual.** Here the textual part of AVD is defined splitting in several information layers, each layer containing different kinds of objects present into a generically archaeological park.
- **Operation.** The last part of AVD allows to define all operations supported by AVD

4.1 AVD: Cartography

The cartographic part of the AVD model is composed by 3 parts: metadata, layers and entities.

4.1.1 **Element <metadata>.** For this part refer to section 2.2

4.1.2 **Element <layer>.** This tag describes the characteristics of a GIS layer. A layer element has a rectangular area that includes all layer extensions. A layer cannot include another layer, while it could be mapped onto many information layers.

In the following an XML-syntax for layer element is described :

```
<!ELEMENT layer        EMPTY>
<!ATTLIST layer
        id                  ID          #REQUIRED
        name                CDATA       #IMPLIED
        color               CDATA       #IMPLIED
        linetypename        CDATA       #IMPLIED
        isfrozen        ("0"|"1")"0"
        innewviewports      ("0"|"1")"0"
        islocked        ("0"|"1")"0"
        isxrefdep           ("y"|"n")"n"
        >
```

4.1.3 **Entities.** Entities are graphical entities that are shown on the monitor. The entities are line, point, circle, trace and text. These elements are described using the SVG syntax for the graphical part while we added new attributes for the AVD goals.

In the following an example of XML-syntax for a circle object is described :

```
<!ELEMENT circle EMPTY>
<!ATTLIST circle
        id      ID          #REQUIRED
        layer   IDREF       #REQUIRED
        cx      CDATA       "0"
        cy      CDATA       "0"
        r       CDATA       "0"
>
```

4.2 AVD Textual.

The textual part of AVD describes how the textual information of AVD is organised. According to a hierarchical structure we define five information layers.
Each level contains objects that represent objects of the real world. An element described in a upper level refers to more elements described in a lower lever. For example a building is described at excavation layer while its walls are described at quadrant layer
All objects are mapped into one or more graphical entities. The information layer is mapped into one or more layers.

4.2.1 **Stratigraphical level.** It is the smallest information layer available in the AVD model. Each graphical layer is represented at the scale 1:10 and describes a single stratigraphical level. Objects present in this layer represent mobile objects found (like pottery, iron, glass …).
4.2.2 **Quadrant.** In this layer a set of stratigraphical levels is described. Objects present here are wall, simple structures and so on
4.2.3 **Excavation.** It is the amount of more quadrants mapped between them. In this layer buildings, roads and so on are described.
4.2.4 **Site.** It describes an area containing one or more excavations. It represents a part of soil containing homogeneous archaeological data. Here the characteristics of the site like stratigraphical sequences, methodologies interpretation and so on are described. Examples of Sites are a necropolis area or the district of Pompeii.
4.2.5 **Park.** It describes one or more excavations even if they are not close. In this layer all administrative and legal information about the park are described. Examples of parks are the city of Pompeii or archaeological park of Como (Spina Verde).

4.3 AVD Operation

In the spatial database world there exists no well-defined standard to manage data, so each GIS Engine can perform different kinds of operations. For this reason we defined a very basic type of operation starting from well-know languages like SQL. For example in AVD you can make queries on the textual part only. For example you can write "find all objects like 'pottery' present at stratigraphical level". The result of this query is a new graphical layer.
You can define queries over textual and graphical part. For example you can write "find all wall founds inside a circumferences of given radius R centred over object Y". Finally, you can define overlay of one or more maps

5 CONCLUSIONS

In this paper we presented ARCHEOGIS, an XML-based interoperable model allowing to describe archaeological data. The leading concept is the data model, in particular we defined a hierarchical structure to split the complexity of the problem. Concerning metadata, we merged different local and international standards to help archaeologists to understand better the data present into the GIS. The choice of XML-syntax like modeling language was due to its simplicity and to the fact that it is a well-known standard, with data entry tools available. To help the development process we defined modular software architecture layer-based. We showed a preliminary version of the AVD model that represents first step towards a unified model in the archaeological world.
The next goal will be to create a data entry tool to simplify the writing of XML files by using an ad hoc graphics user interface and a translator tool from an XML file to a specified selected platform in (semi) automatic way. We will improve the AVD model adding more operations. As a result and an application, we will use the ARCHEOGIS technology to create a GIS for the Archaeological Park of Como

REFERENCES

Brandon J., Kludt T., Neteler M., 1999 Archaeology and GIS -- The Linux Way
http://www.linuxjournal.com/current/2983.htm (08, November 1999)

Bray, Paoli, 1998 - Extensible markup language,
http://www.w3.org/TR/1998/REC-xml-19980210, (10 February 1998)

Buelher, K and McKee L., 1996. The OpenGIS™ Guide, Introduction to Interoperable Geoprocessing, Open GIS Consortium.

Burrough, P., McDonnel, A., 1998. Principles of Geographical Information Systems. Oxford University Press, Oxford

European Prestandard ENV 12160:1997 Geographic Information - Data Description - Spatial Schema

European Prestandard ENV 12656:1998 Geographic Information - Data Description - Quality

European Prestandard ENV 12657:1998 Geographic Information - Data Description - Metadata

Ferraiolo J, 1999, Scalable vector graphics
http://www.w3.org/1999/07/30/WD-SVG-19990730/index.html, 30 July 1999

FGDC,1994. Content Standards for Digital Geospatial Metadata. Federal geographic Comitee.
http://geochance.er.usgs.gov/pub/tools/metadata/standard/standard/metadata.html (13 March 2000)

K. Gourad 1999, GIS IN Archaelogy: a Survey. http://research.hunter.cuny.edu/arch/master.htm (7 August 1999)

Kvamme K. L.,1996. A view from across the water: the North American Experience in Archaelogical GIS in Aldenderfer M. & H.D.G. Maschner (eds.), Antropology, space and Geographic Information Science, New York, Oxford University Press : 1-14.

Marchiori M, 1998. *The Query Languages Workshop* http://www.w3.org/TandS/QL/QL98/pp.html (14 January 1999)

OpenGIS Consortium Home Page. OpenGIS Consortium.
http://www.opengis.org/homepage.html (13 March 2000)

F.Parise Badoni, M.Ruggeri Giove - Istituto Centrale per il Catalogo e la Documentazione-Soprintendenza Archeologica di Roma, 1984, *Norme per la redazione della scheda del saggio stratigrafico*, Roma, Multigrafica.

Sellis T,(1999) - *Research Issues in Spatio-temporal Database Systems*. Lecture Notes in Computer Science, Vol. 1651, Springer,5-11,

The Dublin Core Metadata Element Set Home Page. http://www.oclc.org:5046/researc/dubin core/ (13 March 2000)