A proposal for the digital storage and sharing of remotely sensed archaeological data

Stefano Campana,[1] Barbara Frezza[2]

[1] University of Siena, Department of Archaeology and History of Arts, Landscape Archaeology, (LAP&T)
campana@unisi.it
[2] University of Siena, Department of Archaeology and History of Arts, (LAP&T)
frezza@lapetlab.it

Introduction

The Department of Medieval Archaeology at the University of Siena has been actively engaged in programmes of landscape archaeology and remote sensing for over twenty-five years. This research activity has produced a large collection of data and in particular a vast number of photographs with related information. For instance twenty years of aerial reconnaissance has resulted in a total of about six thousand vertical air photograph, while in the last six years our own aerial survey programme has produced an archive of about 30000 oblique photographs. It is obvious that the management and communication of this quantity of data can only be handled through the use of digital databases. We started our own experiments with multimedia databases in 1999, using Canto Cumulus software, but this product soon showed many problems of inflexibility and difficulty in the sharing of data.

Digital technology, of course, in addition to offering powerful tools for the cataloguing, storage, and analysis of archaeological data, has become a common and speedy platform for the communication of information. Some years after our first experiments we had the opportunity to move to a database that could share information through the Internet. A first attempt was made in 2004 and 2005 to build a web-based database (Donati 2005). The result was excellent for interrogating and managing the data but unfortunately it was not possible to use the system for the everyday storage and use of the data. (See http://www.paesaggimedievali.it/atlante/index.html). This prototype system required us first to catalogue the data in Canto Cumulus and then to export the whole database each time we wished to update the web-based system. In a new development, under design more recently, the revised system has been required to fulfil the following tasks:

- Real time update.
- Quick and easy communication of data within the laboratory as well as with the general public, private individuals and other institutions.
- First steps towards a web-based GIS that could manage the data integration of graphical, geographical and alphanumeric data

It is the development of this system which is discussed in the remainder of this article.

Stefano Campana

Database management systems

Today the Internet represents one of the more effective instruments for data exchange and it will undoubtedly remain one of the most effective channels for the communication of information in the future. Through agreed communication standards its architecture guarantees compatibility of interchange between the different sites that make up the net, facilitating the sharing of experience, services and information sources despite the physical separation of the users.

Our appreciation of the enormous potential of the Internet has led us to put in train a project which aims at the creation of an Internet-enabled information system based on a Data Base Management System (DBMS). What we envisage is a distributed system that allows for the storage, up-dating and recovery of information from a database of aerial photographs of Tuscany, managed through an interactive graphical interface.

A DBMS is in essence an instrument that treats the data as the shared resource of an organisation and of its various users, thereby meeting the ever more obvious need for the distribution and sharing of information and knowledge. The essential properties of a DBMS are:

- The recording, organisation and management of the data according to criteria that answer directly to the needs of the context in which system has been developed.
- The easy communication and sharing of information – the users should be able to access the database by interrogating it in a variety of ways.
- Reliability, in the sense of the capacity to conserve the content of the DBMS through rescue functions and restoration procedures.
- Privacy, guaranteed through the use of authorisation protocols and control procedures.

In addition, the system must be efficient, in that it is able to operate within the resources of space and time available to the users. It must also be effective in helping to make the user's work more productive.

A DBMS is in essence a layer of software interposed between the user and the information stored within the system. This intermediate layer prevents the user and the applications from interacting directly with the data, which can therefore be stored, physically, in the most effective way possible without preventing the user from having access to it. This permits a high degree of independence between the applications and the physical storage of the data. These characteristics ensure that any application that is developed can draw on a data source that is secure, reliable and economical in scale. Such properties are ideal for applications that use the Internet in their infrastructure and which therefore have an obvious need for security and economy of scale.

The decision to place the DBMS on the net also greatly widens the system's capacity for increased effectiveness in the communication and sharing of information. The net is capable of making available to the public, virtually instantaneously, a wide variety of data, thereby allowing them to consult the stored information and to take part in critical appraisal and re-interpretation of its content. In the archaeological context, for instance, this could offers to conservation agencies a simple and immediate means of checking information about sites and other features with which they might be concerned. Internet access also makes information available to users in all parts of the country, and to people who are not necessarily specialists in the field of archaeology. It also permits the database to be enhanced or up-dated in real time from mobile locations or from a variety of different work stations.

The research design

The development now in hand envisages the use of an MSQL database connected to the net through a series of PHP (PreProcessor Hypertext) routines. Access to the database is achieved through the use of the codified language SQL (Structured Query Language). This requires every query to be expressed in terms of commands, or phrases, written in conformity with strictly codified rules. These in turn trigger the execution of operations for the up-dating or interrogation of the data.
The DBMS acts at three levels:

- Storage and management of the data.
- Communication and exchange of the information
- Making the information available to non-specialists within the community.

The structure of the DBMS will use a simple bi-dimensional database, of the kind often described as 'flat-file' in that it is composed of a single table. The table in our case is composed of a field for the individual photo and thirteen related fields – year of photography, type of film or sensor, region, province, district, community, place-name, type of evidence (cropmark, soilmark etc), documentary references, interpretation, reliability, site type, chronology – plus an identification reference for each record.

The fields selected for inclusion in the DBMS are those which we feel will most readily supplement the user's understanding of the individual images. It was necessary, therefore, to make judgements about which of many available kinds of information would best help the user to form a more complete picture and to understand the ways in which the information is organised and managed.

In planning the database an aspect which has been given particular attention has been standardisation of the language used, especially that relating to the fields which involve one form or another of interpretation. Ordinary spoken or written language has a vast capacity for expressing shades of meaning, depending on the context, the tone of communication or the subject under discussion. The computer is, by its nature, incapable of coping with such subtleties and it becomes necessary to formulate strict standards for the language used in the compilation and interrogation of the DBMS. This is done by establishing a series of fixed terms for use in fields where there might be several different ways of expressing the ideas or factual information involved.

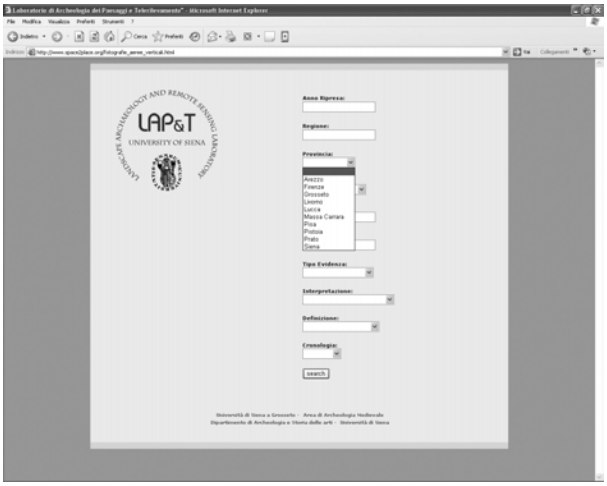
The operations involved in the compilation and interrogation of information within the table, and more generally the proper operation of the database as a whole, can be assured only if these basic rules of planning are respected. This formal clarity of expression has a critical impact on accessibility and on the correct use of the database, both for enquirers and for those responsible for up-dating its content. The best solution lies in the creation of restricted lists of terms in the entry and search screens of the DBMS, chosen as required from pull-down menus. In creating these lists of terms it has been necessary to consider all the possible variations that might present themselves in up-dating or interrogation of the archive, taking account of clearly defined criteria and of the degree of detail that one intends to add to or extract from the system.

The descriptive fields which require the use of fixed terms are those that deal with type of evidence, interpretation, reliability, site type and chronology. The routines for operations involving the addition or extraction of information from the record force users to employ only the pre-determined formulae, allowing no scope for variations or additions. Only the administrator of the database is allowed, in particular circumstances, to add to the lists, modify the definitions or insert new terms.

All the graphical documentation in our own system, covering approximately 35000 between vertical and oblique air photographs of Tuscany, is stored in a folder outside the data bank. In the photo field there is recorded only the path through which the image can be recovered from this folder. In this way the size of the DBMS itself is kept more economical. Recovery of the images from their original positions, in response to each enquiry, also increases the speed with which the system can obtain and deliver the images to the user.

Another element that directly influences the effectiveness of any DBMS with a user interface is the need for it to have a clear and linear architecture, so that the user can benefit to the full from the potentiality of the data. The interface which we have been developing meets this requirement and offers the user various ways of securing access to the data. The interface has been designed so that users can include in their interrogation almost all of the fields present in the table, either by keeping to the lists of fixed terms or by digitising the text directly when these are not pre-determined. The only fields excluded from enquiries are those concerned with the reliability of the information and the presence or otherwise of bibliographical references, plus of course the photo field itself. Once users have made codified entries in the fields which they wish to use in their enquiry, they simply click on the search button to start the interrogation process. The structure of the query routine provides for the interrogation on several fields simultaneously. Setting the system up in this way increases the speed with which users can carry out enquiries on specific problems.

The same interface screen is used for searching and for up-dating the DBMS.



Entry of new data takes place directly on the net, in real time, from any computer connected to the Internet. The added text information is transferred automatically to the DBMS, creating new records, while the photographs, through an up-loading routine, are transferred to the image folder. Access to the entry page for the creation of new records is controlled by a system of authorisation and protection, using passwords. Control protocols are also under development to guarantee secrecy and to prevent manipulation of the data bank's contents by unauthorised users. Operations on the database itself can only be carried out by operators who hold authorisation for the particular operations or kinds of information concerned.

The development of the DBMS has also involved work on design of the pages through which the results are presented on the user's screen in response to enquiries. Once a successful search has been made, the results are presented on-screen. It is possible, of course, that the records extracted from the database will be too numerous for presentation on a single web page. To ensure clarity of presentation in such cases it is necessary to 'format' the results through the SQL code.

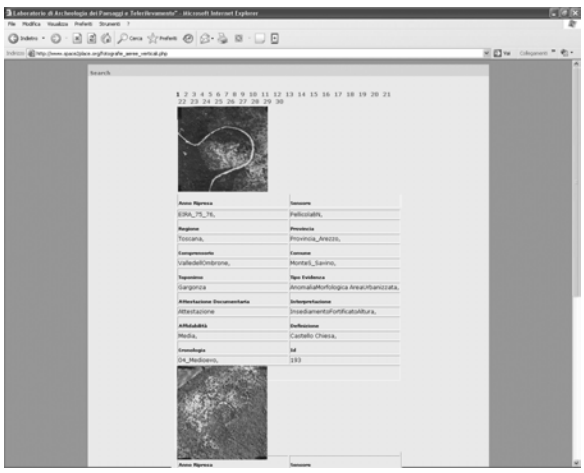
'Formatting' means the subdivision of the query results into smaller parts, pages in this case. Once transmitted to the user these can be presented on-screen through a navigation system which allows the user to move easily from one page of results to the next.

Failure to use this kind of formatting would be risky because it could:

- compromise readability of the information (a long page, rich in data, has less chance of being read to the end);
- lengthen the time taken to load the page;
- increase the quantity of data to be exchanged between server and user for each request;
- take up an excessive amount of server time, especially if more than one query is being made about the same record at any one time.

For these reasons the division into pages achieves an improvement in overall performance of the operation. It is also worth remembering that the query results often follow extended journeys on the net and it is therefore best to limit their number. It was therefore decided to subdivide the data extracted from the database into groups of five pages, and to attach, at the foot of the last page, a link to the previous or next page so as to allow the user to navigate through the results on-screen. In effect the page is always the same but the contents change according to user's choice of page number for scanning, the relevant number being passed on through a QueryString.

The presence of a photo field containing exclusively graphical data has made it necessary to introduce thumbnails, or miniature previews of the images concerned. A thumbnail is simply a copy of a larger image, modified to reduce the size of the file. Through a simple link to the larger image the user can see the image in its original form.



The 5000 or so air photographs in our database are all in JPEG format, with an average a size of about 2 MB, too big to be downloaded and presented on screen relatively quickly. We have therefore made provision for the creation of thumbnails of 200 x 200 pixels and a file-size of about 90 GB, using SQL code to create them as and when needed, without taking up any permanent storage space. In response to each query the database simply collects the image from its original position and creates an instantaneous thumbnail that can be seen by the user on the results screen.

Conclusion

The DBMS described in this short article has proved to be an extremely effective instrument for the transmission and exchange of all types of archaeological data in alphanumeric form. It is not our intention to present this as a finished product, but rather as a prototype to be tested in the future through a suitable number of case-studies. It could represent, however, a first basic model for the organisation of easily exportable information, readily adaptable to all types of data. The system should be seen as in a state of continuous evolution, in that it must necessarily adapt to new products coming onto the software and hardware markets, and to the opportunities presented by future advances in our researches.

Barbara Frezza

Acknowledgements

The authors wish to acknowledge their gratitude to the following:
Giancarlo Macchi (University of Siena) for help and criticism at the beginning of the project.
Enrico Donati for the development of the first prototype web-based database system.
Chris Musson, from Wales, for his invaluable comments and for his help with the English version of the text.

References

ATZENI P., CERI S., PARABOSCHI S.,TORLONE R. 2002, *Basi di dati: modelli e linguaggi di interrogazione*, Milano.

ATZENI P. [et al.] 2003, *Basi di dati: architetture e linee di evoluzione,* Milano.

BUYENS J. 2002, *Sviluppo di database per il Web: passo per passo,* Segrate (Mondatori informatica).

CAMPANA S., FRANCOVICH R. 2003, Landscape archaeology in Tuscany: Cultural resource management, remotely sensed techniques, GIS based data integration and interpretation. In M. Forte, G. Williams (eds.), T*he Reconstruction of Archaeological Landscapes through Digital Technologies, Proceedings of the Ist Italy-United States Workshop, Boston, Massachusetts, 1-3 November 2001*, BAR International Series 1151, pp. 15-28.

D'ANDRIA F. 1997, *Metodologie di catalogazione dei beni archeologici, Lecce-Bari* (Beni Archeologici - Conoscenza e Tecnologie, Quaderno 1.1)

DOERR M., SARRIS A., *CAA 2002 The Digital Heritage of Archaeology, Computer Application and Quantitative Methods in Archaeology. Proceedings of the 30th Conference, Heraklion, Crete, April 2002*.

DONATI, E. 2005, *Un database on-line per le fotografie aeree oblique dell'Università di Siena. Prototipo di contenitore dinamico.* Undergraduate thesis in the Department of Medieval Archaeology at the University of Siena. Unpublished.

GOTTARELLI A. 1997, *Sistemi informativi e reti geografiche in archeologia: Gis-Intenet*, Quaderni del Dipartimento di Archeologia e Storia delle Arti, Sezione archeologica – Università di Siena, VII Ciclo di Lezioni sulla Ricerca applicata in Archeologia, Certosa di Pontignano (SI) 11-17 dicembre 1995.