PETRA 3.0 and the Crusader border
New features of the PETRA archaeological DBMS

Franco Niccolucci[1] and Marco Crescioli[2]

[1]University of Florence, Italy
niccolucci@unifi.it
[2]Unirel srl, Florence, Italy
marco@unirel.it

Abstract. This paper reports on the development of the PETRA archaeological database management system describing the (1999-2000) version 3.0, the latest in use, originally presented at the CAA Conference held in Dublin, and briefly describes the subsequent evolution of the system through the development of a separate on-line GIS module and an XML-based text document module that are now going to be integrated into a new system. A more comprehensive version is due to appear in Summer 2003, integrating all these different modules in one application. The system philosophy is based on openness – we use only Open Source software and make ours available under GPL; portability – the system is scalable from workstations to handheld PCs; simplicity of use; on-line access; platform-independence.
Keywords: Database management, Open Source software.

1 The PETRA Archaeological DBMS

In a previous paper (Crescioli and Niccolucci 1999), delivered at the 1998 CAA Conference in Barcelona, the authors presented a DBMS (Data Base Management System) to manage the archaeological data produced during a multi-year investigation carried out by an Italian team in the Petra Valley (Jordan), directed by Professor Guido Vannini of Florence University; the computer sector has been directed by Franco Niccolucci since 1997.

The aim of the archaeological project is to study the original features of the Crusader settlement in Trans-Jordan investigating the castles of the Petra Valley, Wu'ayra, Jabal Atuff and al-Habis, as well as the other castles of the system, Bayda and Shawbak.

Stratigraphical analysis has documented the sequence of site occupation phases and, in particular, the solutions adopted by the Crusaders to fortify and control the complex rocky configuration of Wu'ayra. This fortress was protected by a double row of defensive walls, extending to the edge of the surrounding wadi, with square towers; the inner citadel and the service area in the centre formed a monumental area.

Sondages on the parts evidenced as most significant during a preliminary survey concern in particular the access system to the castle (the gatehouse, the bridge across the wadi and the ramp strongly defended by watch towers), the inner communication system, the complex of the fortified church and the cemetery area, with 20 tombs, all pertaining to the Crusader occupation period.

Using 'light' archaeological methods, that is non-invasive methods, such as standing structure archaeology and landscape archaeology, a Byzantine phase has also been recognized, which was later incorporated into the Crusader settlement phases.

As in any large investigation, the quantity of data is huge, their nature inhomogeneous, and the data were acquired over a period of several years with different software and computers. So, the first problem faced by us consisted in adopting uniform procedures and forms and creating a data management system that could be accessed easily, regardless of the computer or the software used or known by the user. We chose to use an Internet browser such as Netscape or Internet Explorer to access data and to plan a system that could work on a stand-alone computer as well as on a network, local or Internet based. This led to the choice of Java, which provided also an interesting opportunity to test the efficiency and portability of this language for archaeological data management.

The first version of the system, the one presented at Barcelona, aimed at verifying the acceptance by the users and was limited to generate statically the HTML pages to be displayed. Some programs, written in Java, transformed the data into off-line HTML pages thus allowing hypertext access to the system by means of a browser. This first version, named PETRA 1.0, had several advantages, as it achieved part of the project goals: it had a hypertext interface with the database, it could work on any computer system requiring no other software and it allowed managing together data of different natures, such as structured data, images and drawings.

2 The Development of New Versions

The main disadvantage of PETRA 1.0 however was the off-line creation of pages, which required re-generating the whole set whenever new data were added. Moreover, storing the thousands of pages created by the process wasted memory and was decidedly inefficient. Therefore a new version was devised in 1999-2000, named PETRA 2.0, to create pages on demand and serve them in the same way as static pages.

This version introduced a completely different way of managing data: if static pages can be viewed using a feature of the browser protocol, that is opening and

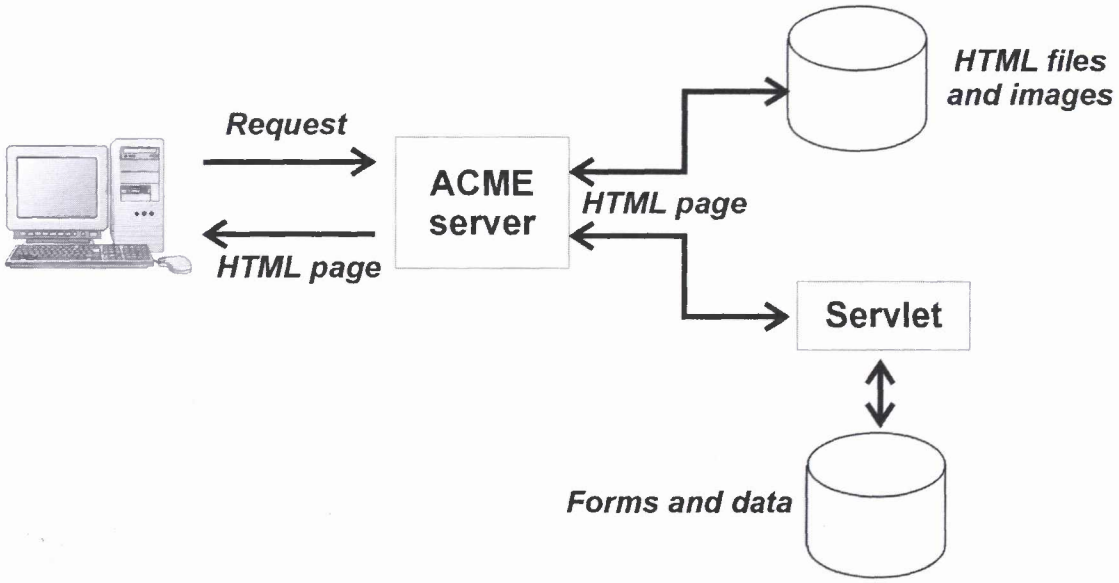

Fig. 1. Generation of HTML pages in PETRA 2 DBMS

viewing a local file, dynamic pages, like the ones we wanted to use, require a server and consequently a more complex structure, possibly contradictory to our simplicity and universality assumptions.

The problem was solved by using a public domain mini-HTTP server named ACME, developed and distributed by Jef Poskanzer, which can work on any computer with very little demands on memory and processing power. This server may be downloaded from www.acme.com. It allows the use of servlets, which are small applications that add functionalities to the server itself. So a servlet was developed to prepare HTML pages, picking data and inserting them into preprepared forms, the same as used in PETRA 1.0.

For the user there is no apparent difference between ordinary HTML pages, as the introductory ones, and HTML pages prepared by the server, neither in aspect nor in functionality. The mechanism underlying the system is the following.

When the user accesses a page clicking a link, the request is sent to the ACME server. If the required page is a normal HTML page, the server looks for the corresponding file(s) in the archives and sends them to the browser, as in normal Internet use. If, on the other hand, the request contains a call to a servlet, the latter is invoked and fills an empty form with the required data, compiling on the spot a new HTML page which is sent, as before, to the browser and usually contains also links which allow further navigation in the database. However, when it is sent, it contains only plain HTML code and requires no processing by the browser.
The complete process is shown in figure 1.

While preparing the system, we noticed that even if the database contained several thousands of records, the complex of all the data required much less than one megabyte, so it could easily be stored in RAM memory. Therefore the system was designed to load all the data at startup and subsequent searches are carried out directly in the memory. Since the system is conceived as a searching system, and is not intended for input, this choice gives only advantages with no apparent drawbacks.

As in the previous version, records are organized in a hierarchical way. There is a form for topographical records, one for stratigraphic units, then a form for materials. All contain links to drawings and photos, when available in digital format.

Accessing hierarchically, the database proceeds from topographic units to the list of stratigraphic units. Choosing one of these, the corresponding detailed form is visualized: it includes also links to related units (for instance, the ones above and below) or to materials. Clicking on the latter shows an inventory list, with summary data of the finds, giving access to more detailed forms.

Hence this way of accessing data proceeds from general to particular and allows also to move, as mentioned above, between stratigraphically related units. There is also a graphical interface (created off-line by a program already present in PETRA 1.0) based on the Harris matrix of every sondage, in which the symbols of stratigraphic units are buttons linking to the corresponding form; they also display, when passed over with the mouse, a short description of the unit. This Harris matrix is automatically generated based on the stratigraphic relationships stored in the database.

Another way of accessing the database uses a query mechanism introduced in the latest version of PETRA. The user fills out a form with the requirements on materials and obtains a list of all the finds satisfying these constraints. The list shows a short description of each item and has links to the find form or the stratigraphic unit form, so that from this list the user can start to navigate the database.
The access procedure is shown in figure 2. It should be mentioned that there is no difference in the use of the system whether the user connects to a local database or to a remote one.

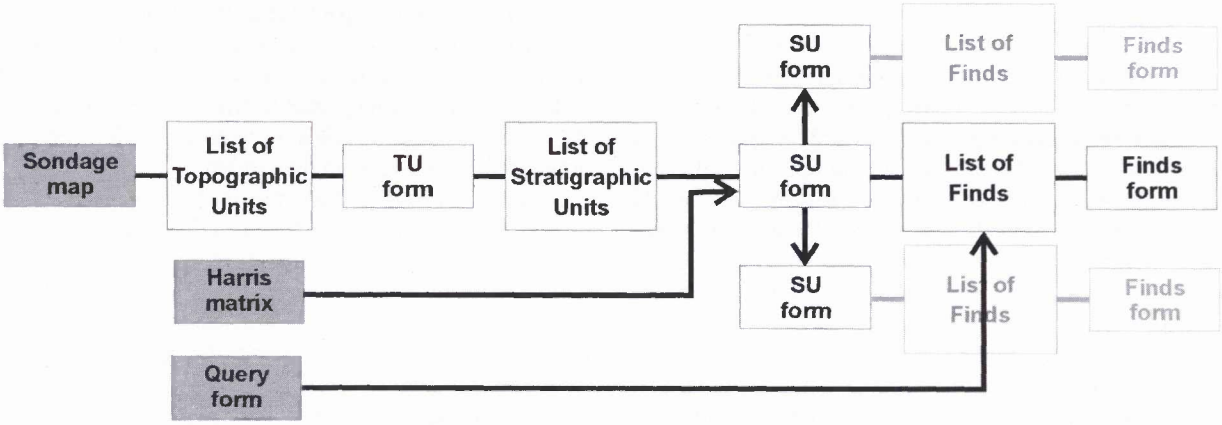

Fig. 2. Access to data

Data may reside on the same PC or be stored in a remote server, connected via a LAN or the Internet. However, the aim of the system was to use the database when Internet access is unavailable, as often happens in the field. Our idea was to store data on a central computer, used as an Internet server, and to access data in normal conditions using an Internet connection. Before going in the field, data could be automatically converted to the format required by PETRA 3.0, downloaded to a PC and then accessed using Petra 3.0, to have at hand a 'light' but complete version of the database.

We experimented once in early 2001 with the feasibility of transferring the system to a handheld computer, the Compaq (now HP) IPAQ PDA. We had to redesign the interface page for the different screen sizes, and since this was just a test of portability, we did not develop a general system adapting automatically to different screen sizes. However, the system worked reliably and when presented at a meeting it was considered with interest and curiosity by archaeologists.

3 The Impact of the Data Management System on Archaeological Activity

One of the main objectives of the DBMS was to verify the acceptance of a hypertext access to data in the synthesis stage of the investigation. In our opinion, the work carried out during the interpretation step could greatly benefit from the possibility of freely navigating the archive, following the inspiration of the moment, and consulting images when required. The system is therefore designed to answer immediately questions such as "What is there above/under this layer? Which materials have been found here? Where were similar materials found?"

These are perhaps some of the questions that could arise during interpretation, and also the graphic access to stratigraphy and stratigraphic units using the Harris hypertext matrix should greatly aid the archaeologist. Another advantage is the easy 'learning curve' of the system: the wide diffusion of the Internet and knowledge of the very simple mechanism of browsing will help in becoming quickly familiar with the data structure and will not require the use of less elementary tools like SQL queries. The query form, which uses a query-by-example metaphor, is only slightly more difficult to use but is absolutely not indispensable for browsing the database.

Finally, the system is platform independent, meaning that it will work on PCs, Macintoshes and Linux systems. This feature is particularly important in those countries where there is a wide diffusion of Macintosh computers (as in Italy or France). For instance, Microsoft Access based databases can be transferred to a Macintosh only after a long and cumbersome conversion process to another DBMS. Other DBMS, available on both platforms, do not seem to have the necessary features for a professional use or admit (as FileMaker does) incorrect practices such as the use of 'multiple fields', forbidden under the rules governing relational databases, in which "all attribute values [must be] atomic" (Date 1986: 243), a mistake that sometimes can still be seen in excavation databases.

4 Further Developments

The situation sketched above corresponds to the development stage at some time between 1999 and 2000. The system proved its usefulness but also showed some limitations, and suggested the necessity of revising the archaeological records in view of the fact that they were collected over a period of some ten years. The team therefore decided to proceed separately to investigate further technological developments, and to revise the archaeological data, which involved converting the older ones to digital format from the paper on which they had been recorded at the end of the 1980s. A lecture at the First Italian Workshop on Computational Archaeology and the publication of that lecture (Crescioli *et al.* 2000) formed a milestone as they marked

the end of the first stage of development and the beginning of the second.

In our opinion, the major limitations of the package lie in the unavailability of a GIS component to deal with site topography and add metric relations to the topology of stratigraphic relations; in the modest treatment of vector graphics; and finally in the impossibility of relating the database records to texts, a feature particularly important for our investigation. Such features are especially significant when the archaeological research is extended to a wider area to include the famous fortress Shawbak, as eventually happened in 2002. Such an enlargement means more spatial relations to deal with, more architectural structures to document with digital drawings (the castle is very well maintained) and more historical texts to relate to the database.

Unfortunately no such data was immediately available for the Petra Crusader Castles investigation, so it was unanimously decided to test the technology on other excavations for which geographical, geometrical and textual data were already available. This could also be an interesting benchmark to expand the applicability of the package beyond the borders of medieval archaeology. Fortunately, these data were available for an investigation in Southern Italy of an Etruscan cemetery, carried out by a team from the Istituto Orientale di Napoli (Naples), led by Professor Bruno d'Agostino, who kindly made available his data. Thus we started collaborating on this project with Andrea D'Andrea, the computer specialist responsible of the Naples team. The database development was for the time being set aside as we were not interested at that time in understanding the customization made necessary by the different archaeological framework (medieval for the original dataset, Etruscan for the Naples one), having decided to focus instead on developing the GIS and text modules with the same philosophy of portability, openness (we use and release only GPL Open Source software), scalability from Unix workstations to handheld PCs, modularity and platform-independence.

This research produced the Grass-based, online GIS module (D'Andrea *et al.* 2001) and the XML database-text connection module (Crescioli *et al.* 2002). These papers seem deviations from the initial Petra database philosophy as previously described in this paper, and were usually classified as "Internet applications" by proceedings editors, but in our minds we knew that they were the pillars for the construction of the improved Petra DBMS, apparently distant from each other until the entire edifice will show in its entirety.

Moreover, a small contribution to the vector graphic and 3D treatment came from a paper on virtual reconstruction using X3D, presented at CAA2002 (Niccolucci and Cantone 2003), but we are aware of the fact that much work is still necessary on this issue, for which we envisage an extended use of XML-compliant graphic formats such as SVG or X3D, in accordance with suggestions by other scholars (see for instance Ryan 2001).

Now the time has come to put together the *disiecta membra* of our system and we are starting to create a new, improved Petra DBMS, finally linking all the necessary functions in a package free not only from costs (it will be released under the GPL license) but also from the quirks of commercial software vendors who, for instance, release every year a different data format for the same package (a well-known one, possibly the most popular and unreliable in the world).

We will provide customization tools and all the documentation necessary for installing and using the software. Possibly we will prepare a 'distribution' as well, following the Linux tradition, in the hope that it may be of some use to the scientific community or at least that somebody will copy the good things and throw away the bad ones.

This PETRA 4.0 project has just started and is expected to produce the first results in early Summer 2003, and will be tested in the field during the Autumn 2003 archaeological campaign.

More ambitious projects under preparation aim at linking this and other online available archaeological databases within a digital library, accessible via the Internet by anybody from anywhere in the world, extending and improving the work by Clark *et al.* (2002) in the framework of a collaboration with their team at NSDU.

References

Clark, J.T., Bergstrom, A., Landrum, J.E., Larson, F. and Slator, B., 2002. Digital Archive Network for Anthropology (DANA): three-dimensional modeling and database development for Internet access. In: F. Niccolucci (ed.), Virtual Archaeology, Oxford Archaeopress (BAR International Series 1075).

Crescioli, M., D'Andrea, A., and Niccolucci, F., 2002. XML Encoding of Archaeological Unstructured Data. In: G. Burenhult and J. Arvidsson (eds), Archaeological Informatics: Pushing the Envelop CAA2001 Proceedings of the 29th Conference, Gotland, April 2001, Oxford Archaeopress (BAR International Series 1016).

Crescioli, M. and Niccolucci, F., 1999. PETRA-data: an Integrated Environment for Archaeological Data Processing. In: J.A. Barceló, I. Briz, and A. Vila (eds), New Techniques for Old Times CAA98 Proceedings of the 26th Conference, Barcelona, March 1998. Oxford Archaeopress (BAR International Series 757).

Crescioli, M., Niccolucci, F, Tonghini, C. and Vannini, G., 2000. PETRA: un sistema integrato per la gestione dei dati archeologici, Archeologia e Calcolatori 11.

D'andrea, A., Crescioli, M. and Niccolucci, F., 2001. Web Access to an Archaeological GIS. In: Z. Stančič and T. Veljanovski (eds), Computing Archaeology for Understanding the Past, CAA2000 Proceedings of the 28th Conference, Ljubljana April 2000. Oxford Archaeopress (BAR International Series 931).

Date, C.J., 1986. An Introduction to Database Systems. Reading: Addison-Wesley.

Niccolucci, F. and Cantone, F., 2003. Legend and virtual reconstruction: Porsenna's mausoleum in X3D, to appear in CAA2002 Proceedings of the 30th Conference, Crete, April 2002.

Ryan, N. 2001. Documenting and Validating Virtual Archaeology, Archeologia e Calcolatori 12.