

# P.E.T.R.A.-Data: an Integrated Environment for Archaeological Data Processing.

**Marco Crescioli**

E-mail:marco@unirel.it

**Franco Niccolucci**

Università di Firenze.

Via Pisana 56. 50143. Firenze. Italy.

E-mail: niccolucci@cesit1.unifi.it

## **The archaeological project: aims, methodology and results**

An archaeological project has been in progress since 1986, carried out by an Italian team, in order to investigate the features of the presence of the Crusaders in the Holy Land, by means of the analysis of patterns in the Western settlement system in Transjordan, during the 12<sup>th</sup> century; these patterns are better preserved and are easier to identify, from a stratigraphic point of view, as all the fortifications were definitely abandoned, by the Crusaders, after their defeat by Salah al-Din, in the battle of Hattin (1187). Results of the investigation have been published in VANNINI - VANNI DESIDERI, 1995 and VANNINI - TONGHINI, 1997, as well as in papers quoted in these two publications; an overall book is also now in the works.

In the first stage of the research (1986-1989), a wide-range survey of the region, between Syria and the Red Sea, proved the significant role of the valley of Petra, for the Western occupation of Transjordan, after a long period of abandonment (started in the early Middle Ages), with the system of fortresses of al-Habis, Jabal Atuff, al-Wu'ayra, al-Bayda, and al-Shawbak. In the second stage (1992-1996), the Italian archaeological mission investigated thoroughly the castle of al-Wu'ayra, with a systematic survey of the standing structures, and the excavation of trenches, sondages, and open-area sondages. In the most recent stage, (1997), field research is being extended to the site of al-Habis, the second Crusader castle of Petra. The possibility of extending the research to other neighbouring regions, especially to Grater Syria, is under consideration.

The fortification of al-Wu'ayra consists of a double curtain of walls (hugging the steep cliff edge), including a number of squared towers, and of an inner citadel, which was the centre of defense (*cassero*). The *cassero* consists of a monumental area, which was strongly fortified, and of a lower bailey. In the area of the *cassero*, 120 Topographic Units were identified; in most cases, it was possible to organise these Topographic Units into a relative chronological sequence (indicated by the stratigraphy of the walls and the occupation floors), and to demonstrate that the *cassero* had undergone successive alterations.

Excavations concentrated on the only access points to the *cassero*, including a gatehouse, a bridge across the *wadi*, and a ramp, strongly defended by watch towers, the system of

internal communications (pathways, passages, ramps, postern gates, movable bridges), and the fortified church, which was the nucleus of the entire fortification's internal organization.

Eight phases were identified, in the course of the excavations of the sondages. Three of them, with occupation levels in primary deposition, can be ascribed to the Crusader period (two phases can be dated to the mid-12<sup>th</sup> Century). The fourth phase is characterised by a sporadic occupation, after that castle was abandoned, and the last four phases are related to the abandonment of the site, consisting of a clear stratigraphy of collapses. A thorough study of the finds, in particular, of the pottery, provided a detailed classification, and enabled an assessment of major changes in the material productions, for the area considered.

## **The role of computers**

Since the beginning of the investigation, computers have been intensively used, to file the records of the finds; in 1995, a computer expert joined the team, and actually, computer applications played an important role, to support laboratory work and to manage data. During every excavation campaign, at least one of us (in the meantime, the computer group had grown in number to four) reached the field team in Petra, to ensure full functionality of data recording. In this way, a complete integration between different skills and experience evolved.

Recently a web site, [www.unifi.it/project/petra/](http://www.unifi.it/project/petra/) has been dedicated to this project; this synthetic information on the research, is described there, in greater detail.

Due to the spontaneous use of computers, in the early stages of the research, the data recording developed habits, which made data base management quite complicated. One half of the team chose Macintosh and File Maker, while the other chose Windows and Access: so, a complete database could be obtained, only after tedious conversions to and from different data formats. Moreover, data consistency could be obtained, only with a verbal division of the computer work, which was very hard to maintain. Besides this, as the research team was formed by people living in different places, a complex traffic of disks (or e-mail attachments) was needed, in order to keep every personal copy of the database updated. So, we decided to convert all data to a more efficient

database, and to develop a remote, access system relying on the Internet.

While our colleagues were preparing the final report on the Wu'ayra excavations, a hypertext approach was suggested, to be useful.

We planned to produce a system which would allow us to:

- convert data from the file manager programs, customarily used for input (which were, in any case, best for the portable computers employed in field work);
- produce on-the-fly HTML pages, derived from data;
- produce stand-alone HTML pages, from the records: they could easily be distributed to researchers, periodically updated (the database was static for long periods), and consulted, by means of ordinary Internet browsers, even if the Internet was not available. This solution could also be used to publish the results of the research, avoiding costly print reproductions of photos, and giving access to all the data.

The first goal was easily obtained and did not deserve much attention: it consisted of using simple batch programs, to download data from the file managers used for input, and other programs, to upload data to the RDBMS in use.

The second and the third goals were similar: in the second case, we needed a server-side approach, to generate, upon request, the desired page. Existing utilities did not allow us to format the page as desired (in particular, to generate the Harris Matrix), so we had to rewrite them.

For the last goal, instead, all pages were automatically generated, at the same time, and saved on disk as single files, which could easily be copied and distributed as CD-ROM.

#### Data formats

Three forms were in use: two, the Topographical Unit (TU) form and the Stratigraphic Unit (SU) form, contained general information, pertaining to the topography and stratigraphy of the site, and were usually grouped according to sondage. The last form contained information on finds, belonging to one of the stratigraphic units, or found on the surface of a topographic unit. So, relationships between the corresponding, three archives followed the simple, following pattern.

Cross-references were based upon key fields: while sondage and TU had a unique number, SU units are numbered progressively for each sondage, so a unique key is formed joining sondage number to SU code. On the contrary, finds had a unique inventory number, but they could be related to the proper stratigraphic unit, through its key (sondage+SU code number), which was stored in the find form.

As noted before, at input data were stored in the proprietary formats, used by the programs and employed for this purpose. After a first check by researchers, and when data were sufficiently stable, usually immediately after every field campaign, they were extracted from the original files and converted to a common text format. For this, we chose DBF format, due to its availability on any data management

program and its simplicity, being little more than ASCII text format.

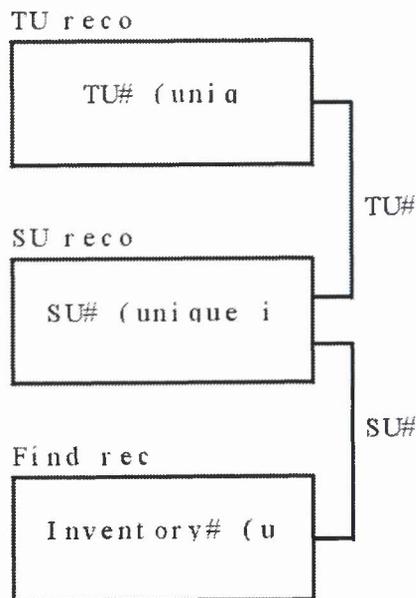


Fig.1. Database relations.

DBF files were then fed into a parser/checker, written in Java, which performed the following tasks:

- check of duplications, in key fields;
- formal validation of content;
- check and standardization of separators, between multiple instances;
- decomposition and replication of multiple references (for instance, in the stratigraphic sequence).

An error log was then produced and submitted to the form compiler, who could easily correct mistakes and produce final archives, which underwent the conversion process again.

At the end, a combined DBF file was produced for each of the archives, and these files were the starting point, for the relational database generation or the HTML pages production.

Importing DBF files, into any of the many available RDBMS was a straightforward matter; however, the HTML pages production required more processing.

Images had a special database, which is now under revision; it contained references to the reproduced object (TU, SU, find, or anything else), and the inventory number of the slide, which would be used also as a reference to the file, containing the digital image, which we planned to store, together with other data. A similar organization was planned for drawings, which, unfortunately, up to this point, had only been done on paper.

#### HTML pages

Particular attention was paid to the graphic appearance of the HTML pages, as they were intended, not only for research use, but also for publishing its results. Commercially available programs, which automatically generate HTML

code from database, were not suitable for this purpose, as they produced standardized pages and, moreover, did not allow hypertext linking, within fields.

The web site consisted of the following kinds of pages:

- 1) a few general pages, with maps, explanations, credits, and so on, produced manually;
- 2) a list of TU, grouped by area and sondage;
- 3) a list of stratigraphic units pertaining to each sondage and each TU, containing a short description of each SU, its stratigraphical relationship with other SU's, and links to drawings and/or photos;
- 4) a page for each SU with all its data, reproducing the original form;
- 5) a list for each SU of finds with a concise description, derived from coded fields, and references to drawings/photos;

6) the full original form for each find;

7) the stratigraphic matrix for each sondage and topographic unit.

Every page was divided into three parts: a menu on the left, a title on top, and data in the body (the central part of it). Obviously, the latter differed from page to page, but also menus and titles were context-sensitive, as they referred or were linked to different pages, according to content. In the body of the page, any reference was a hypertextual link: for instance, in the lists, every cell (except the description) allowed navigation to the referred item.

The stratigraphic matrix could be reached from any page.

The following picture shows a map of the site, illustrating its hierarchical structure.

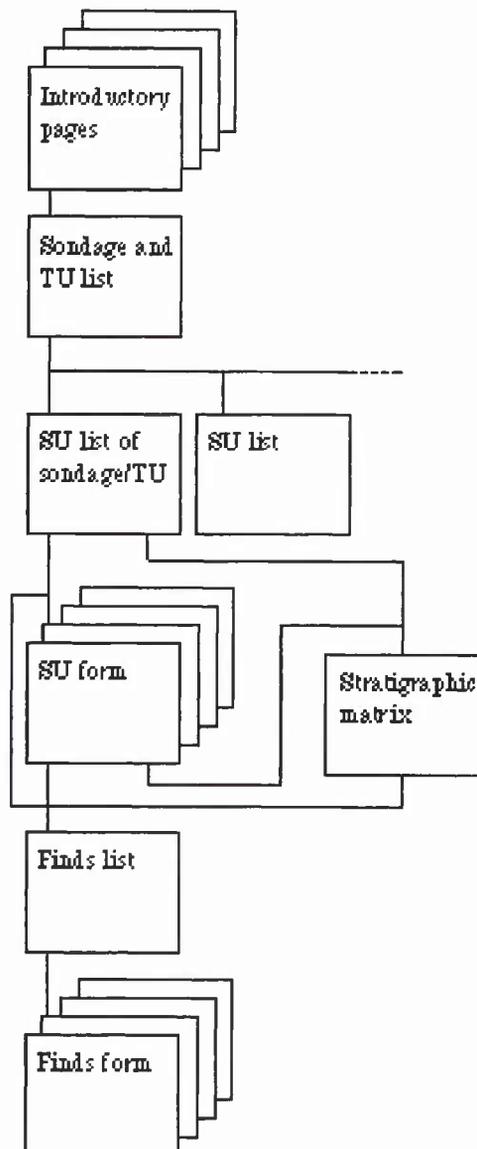


Fig. 2. Map of the www site.

### HTML pages generation

The burden of generating HTML code was given to a Java program, which loaded the structure of the (empty) model page, loaded the data from the DBF file and then compiled and saved to disk, the combination of the two: in this way, many thousands of pages were generated, grouped into

directories to make browsing easier. Similar programs were used for the different kinds of pages.

In this process, the HTML generation program had to fill in the gaps, of the model page, with the correct HTML expressions, derived from data, formatting them as required and creating, if necessary, links to other parts of the site.

Every page was then saved as a separate file, with a name, following a simple rule: for instance, the materials form #1796, belonging to SU16 of TU86, in Sondage 2, became the page, stored in file /UT86/S2US16/M1796.html (the inversion of initials was due to the Italian order of adjectivation), i.e. the file M1796 in subdirectory S2US16 of directory UT86; the corresponding SU form is file US.html in the same directory, and so on. To be more precise, the three parts (i.e., menu, title, and body), were generated and stored in separate files (all sharing the same name, suffixed respectively with -menu, -body, and -title) while a fourth file contained the frame description of the page.

In the process a final validation was performed: relationships among SU were checked, to ensure that every relation had a corresponding inverse (e.g., if SU24 is over SU30, SU30 must be under SU24). In case of errors, a warning was issued.

During processing, a partial order relationship among SU was created, reducing stratigraphic relationships to the basic three: "equals", "precedes", and "follows". This partial order was stored in a (numerical) symmetric array, to be used later to generate the stratigraphical diagram.

At the end of the pages generation, loops in the order relation were resolved (and warnings generated); then, a simple algorithm created the graph, starting from the order array. In the first step, the "ancestor" (ancestors) of the arborescence could be easily found, as the element(s) having no predecessors, and were thus marked; its "sons" were then, the unmarked elements, which had no unmarked "parents". This procedure was then repeated until every element of the array was marked, that is, every element of the array was assigned to its relative position.

For the graphic output, an HTML page was then generated, filling in the cells of a table, either with vertical/horizontal lines, to represent element connections/order, using previously created GIFs, or, with the code number of an SU, which was also a link, to its page.

The Harris matrix, thus obtained, was not the final diagram, but only a first approximation of it. The researcher had to eliminate redundant information and to reposition cells; but, the tool we made available to him/her was very useful as a starting point, and also could be used to check quickly, using hypertext links to data, the correctness of his/her deductions. After completing HTML generation, more than 4,000 pages were generated, for a total of about 15,000 files (counting frames, menus, titles, and bodies), including forms, lists and diagrams.

The entire process of generation required 10-15 minutes, on a Pentium 200, running Linux; but, as Java is platform independent, the generation may be performed, under Windows 95, Windows NT, or Macintosh OS. Windows 3.1 should have no drawbacks, but it would require a reformulation of the naming rules.

The web site has a storage occupation of about 100 Mbytes, depending on the operating system. Access to pages is quick, due to the absence of heavy graphic elements, and any browser, even if not Java-enabled, gets satisfactory results. A sample of the web site can be seen at the following URL:

[www.unifi.it/project/petra/](http://www.unifi.it/project/petra/), which is the home page of our project, following the links to "Computers in Archaeological Research".

## Conclusions

This work proves that the Internet may be used, as a research tool, and not only for the diffusion of knowledge.

With this approach, the full body of data may be shared, not only by all the researchers involved in the investigation (which was our primary concern), but also by the whole scientific community, since access, (restricted when data are not ready for publication), may later be opened to the public.

The hypertext access to data may be helpful during archaeological research, when conclusions are being drawn. It may also be useful also after (electronic) publication, as links in the text can point directly to relevant data.

This encourages us to proceed to the second step of our work, in which HTML pages will be created on demand, dramatically reducing storage requirements.

## References

- VANNINI G. and VANNI DESIDERI A. (1995), "Archaeological research on Medieval Petra: a preliminary report", *ADAJ*, 39, pp. 509-540.
- VANNINI G. and TONGHINI C. (1997), "Mediaeval Petra. The stratigraphic evidence from recent archaeological excavations at al-Wu'ayra", *The Sixth International Conference on the History and Archaeology of Jordan*. 1997, pp. 371-384.

## Acknowledgements

We want to thank Prof. Guido Vannini and Dr. Cristina Tonghini for the help given.

This research was funded by a grant from the Italian M.U.R.S.T.