From a Relational Database to an Integrated System: a Milan University Project

Glauco Mantegari and Tommaso Quirino

Università degli Studi di Milano, Dipartimento di Scienze dell'Antichità, Italy
mantegla@yahoo.it, qui.tom@libero.it

Abstract. This paper describes the experience carried out during the last years in the creation of an integrated digital solution for the management of the entire documentation collected during the archaeological excavations organized by the Chair of Prehistory and Protohistory of the University of Milan. This system is composed of three main modules (an alphanumeric database, a GIS software and a multimedia database) interconnected by simple interfaces developed using Visual Basic programming language. The choice of personally managing the creation of the system gave us the opportunity to answer the needs raised by the researches on the field, laying the foundations for the development of a solution that balances the methodological issues linked to data management with the opportunities offered by new technologies.
Keywords: database, GIS, multimedia, Visual Basic programming

1. Introduction

The experiments of informatics applied to archaeology, in the research of the Chair of Prehistory and Protohistory of the University of Milan[1], started in 1999. In this first phase, due to the lack of a solid tradition for this type of applications, the use of informatics took place at a fairly superficial level. The PC was used almost exclusively as instrument for data representation, the database for the management of the excavation documentation was of linear type, the GIS software was mainly used for the vectoring and georeference of the excavation plans and lacked an organic model of the data (de Marinis, Baratti ,Longhi and Mangani 2002, de Marinis, Sidoli and Rapi 2002, Mangani and Longhi 2003). However, since 2001 the increasingly constant use of the informatics instrument has forced us to fully exploit the possibilities offered by information technology to produce a complex and complete instrument for the collection, management and interpretation of archaeological data. The main aim of our research is therefore the development and testing of a digital system for the management of the entire documentation collected on the field (alphanumeric, graphic and multimedia data), considering the close relationship existing between the methodological matters and the choice of the most suitable technologies to represent the data and consequently combining the functions of the different types of software which best meet our needs. The model on which our system is inspired is the "GIS solution" produced by LIAAM (Laboratorio di Informatica Applicata all'Archeologia Medievale – Laboratory of Informatics Applied to Medieval Archaeology) of the University of Siena (Fronza 2000, Valenti 2000, Fronza, Nardini, Salzotti and Valenti 2001).

2. The Modules of the System

The system is composed of three main modules which have been developed independently and then interconnected to permit a multidirectional reference among the different data categories. The alphanumeric data are organized in a relational database, the graphic data are managed through a GIS software and the images or any other multimedia files are collected in a multimedia database.

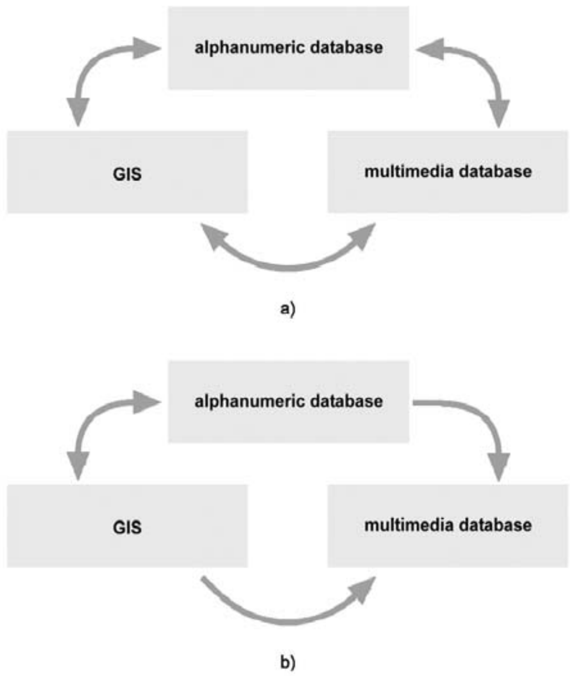

Fig. 1. The conceptual framework of the system (a) and the current development (b).

2.1 Alphanumeric Database

Special care was paid, from the start, to the implementation of a relational database for the management of the excavation data[2]. The alphanumeric database represents a necessary start for the construction of a functional GIS system. In fact, the presence of well structured and organized information determines the possibility of associating attributes to the graphic elements of the GIS and to the images of the multimedia database.

A long and deep design has led to the creation of the conceptual diagram (figure 2), which explicitly refers to that produced in the University of Siena. It is characterized by a hierarchical tree constructed using the relational model which is at present the most diffused and consolidated form for database construction. At the top of the tree is the "site" entity functional for the management, inside the same database, of the data pertaining to research carried out in different places. Starting from the "site" entity the tree structure is subdivided into three main "branches", which group entities of common topic.

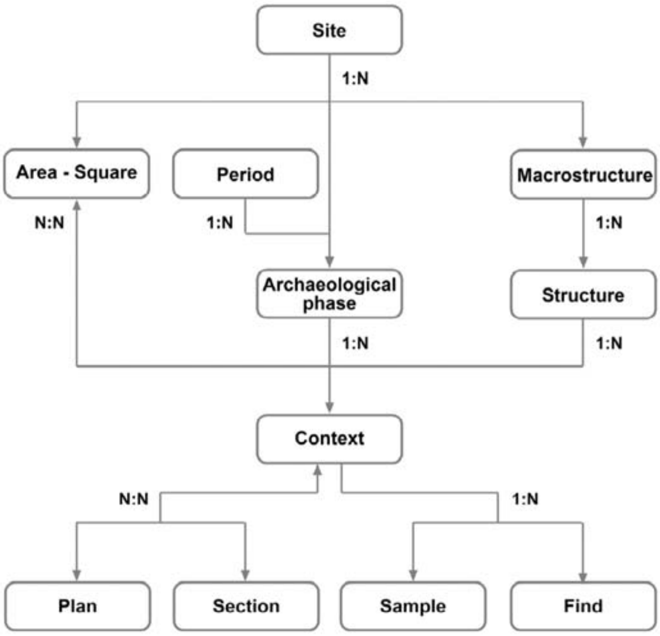

Fig. 2. The structure of the alphanumeric database with the main logical relations.

The first branch represents the information of topographic nature, i.e. that concerning the space organization of the excavation area in "sectors" and "squares". The second branch represents the information of chronological nature: the "period" entity collects the data concerning the absolute chronology and the "archaeological phase" entity those concerning the relative chronology of the deposit. Finally, the third branch represents the information of structural nature, i.e. those concerning the groupings of stratigraphical units which identify a "structure" and the groupings of structures which identify a "macrostructure", arranged to produce a hierarchy of assemblies and logical subassemblies. At the next level, the three branches of the diagram meet in the "context" entity, which constitutes the fundamental element of the survey of an archeological deposit and consequently represents a key point of the entire database. Finally, a further five entities are connected with the "context". The first two collect the alphanumeric data connected with the survey documents ("plans" and "sections"); the following ones regard the significant components coming from the single units of the deposit, i.e. "finds" and "samples".
As regards software, Microsoft Access 2000 was chosen, mainly due to the availability of PC's based on the Microsoft Windows operative system and of licences of the Office 2000 package in the University.

The deep analysis of the database requisites has considerably facilitated the program implementation phase, making it possible to devote more attention to the problem of language standardization and the creation of a personalized user interface.
The standardization problem was faced during the phase of fields choice and analysis of their properties (contents and restraints). The insertion of information expressed in homogeneous form is, in fact, indispensable to be able to perform effective and reliable research in the database. We have therefore been tried to standardize the possible language variants used in the compilation of a field creating a series of vocabularies containing the most significant items for the description of the subjects to be filed. In this way, on one hand an excessive reduction of the complexity of the aspects of the subject and on the other a useless redundancy of definitions were avoided.
The final moment of program implementation was the creation of the graphic interface, fundamental to make the application usable by any type of user, even with a low level of EDP knowledge. The interface layout is based on navigation criteria similar to those of a hypertext: through linear navigation inside the same file, or through the relational one between different files, the user is enabled to reach any mask from any other mask in simple and intuitive manner. The interface was largely produced through the Access editing environment and further personalized with push-buttons created through raster graphic programs. To optimize data display and reference, macro and specific routines Visual Basic were abundantly used.
As regards the control of file access, three different work areas, corresponding to different user levels: data reference, data input and files maintenance, were provided. The first level is dedicated to any user who must have access to the data for research purposes: all the fields are blocked, any type of interrogation may be made but without modifying, adding or eliminating data. The input area is on the contrary reserved to authorized users who may adjust the data, to add new records or update the items of the vocabularies. Finally, the maintenance area, destined solely to the system administrators, makes it possible to have access to the DBMS structure and make modifications or additions.

2.2 GIS

In the GIS software[3] all the graphic data concerning the excavation are filed. These have been subdivided into two main levels: the macroscale, or the whole excavation area, and the microscale, or the single stratigraphical trench. The first level is represented by the topographic survey of the area affected by the research, which contains contour lines, road system, present buildings, irrigation drains and areas under survey and excavation. The second level contains the polygons which identify the stratigraphical units, with their surfaces, characterization and elevation points.
The data entry consists in the digitizing of the excavation plans, scanned traditionally or obtained from photomosaics. After this first phase of filing of the documentation, a single shapefile, which contains the polygons of all the

stratigraphical units of the deposit with which the context number, together with other alphanumeric data, is obviously associated, is obtained from the shapefiles of each plan.

2.3 Multimedia Database

The multimedia documentation (photographs of contexts and finds, films, QTVR) form a fundamental part of the documentation of an archaeological excavation which however is often ignored. Already in the past the filing of hundreds of slides and traditional photographs raised many difficulties which inevitably caused considerable lacks in the subsequent reference. With the arrival of the digital photograph different problems still exist; for example extension, dimension and name to assign to the files, how to file them, how to make fast queries.

To manage this type of documentation and make the important information which it contains accessible and referable, in useful way, a specific software, Canto Cumulus 5.5, has been adopted. It is a database dedicated to the filing and organization of multimedia files which makes it possible to associate each subject with different categories and keywords on which it is possible to carry out research.

In our case, since the alphanumeric data associated with the digital photographs are already inserted in the relational database, the file has been structured in very simple manner: the records have been indexed exclusively for context number and find number. Although these data are very limited to carry out research in the single software, however, being key attributes, they make it possible, similarly to the GIS, to make the connection with the alphanumeric database and therefore to carry out research on all the data which it contains.

3. Links

Re-using the Visual Basic language it was possible to structure some data interchange mechanisms among the different applications, obtaining simple interfaces which automatically permit the cross-reference of the data as illustrated in figure 1b. In fact, both ArcGIS and Cumulus make available a series of libraries which make it possible to manipulate the components of the application developing personalized functions.

Inside the GIS platform a preliminary work on the so-called "tables of the attributes", which were deprived of the fields not strictly necessary for the space display and processing of the GIS, was indispensable to lighten its structure and prevent the presence of data redundant to those already present in the alphanumeric database. The union then proceeded of the single files concerning the surveys of the context in a new single shapefile, which is configured as a large compound plan of the entire surveyed stratigraphy on which to automatically make the selections.

As shown by figure 3, after making a research in the alphanumeric database using the mask dedicated to it (a) it is sufficient to press one of the two push-buttons present in the small pop-up mask named "Connections" (b) to obtain the display of the surveys of the units selected in the GIS (c), or of the multimedia documents filed in Cumulus (d). The mechanism of automatic selection of the multimedia documents from the alphanumeric database was also extended to the management of the photographs of the finds.

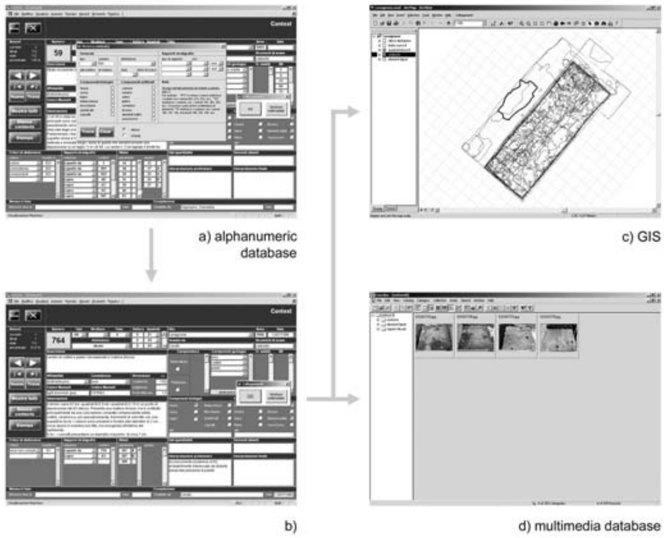

Fig. 3. Data exchange among the alphanumeric database (a, b), the GIS software (c) and the multimedia database (d). Similarly a pop-up mask similar to that described (figure 4) was then created in the GIS. It makes it possible to obtain the automatic display of the records in Access (b), and in Cumuls (c) starting from a selection made on the excavation platform (a).

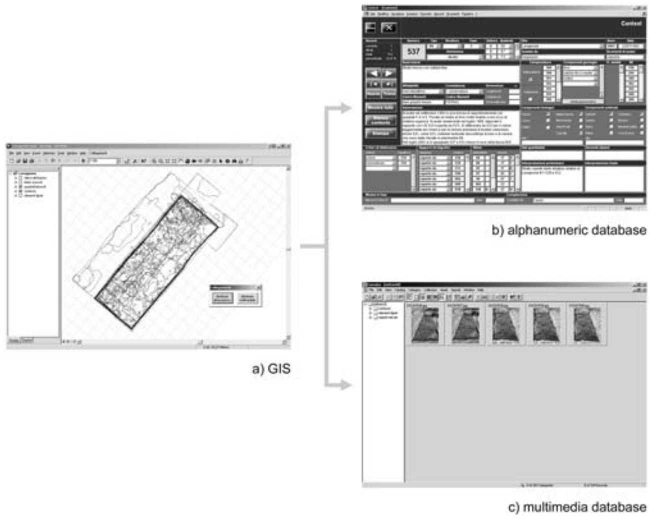

Fig. 4. Data exchange among the GIS software (a), the alphanumeric database (b) and the multimedia database (c)

4. Conclusions

The test performed up to this moment has been marked by the prevalence of the design and constructive aspects compared with those strictly applicative.

The insertion in the system of a quantitatively significant group of data has made it possible to check the effectiveness of the solution so far prepared. Thanks to the automatic data research, selection and removal mechanisms it was to make an accurate and complete revision of the documentation collected on the sites in question during over ten years of excavations. In particular, it was possible to punctually identify and correct a series of errors and incongruities mainly caused by the management difficulties traditionally raised by the documents produced exclusively in paper form. This revision led to a more precise definition of the stratigraphical sequence of the two sites and consequently to a better understanding of the subsequent installation phenomena.

The completion of the data input is therefore the main target which it is intended to reach in a short-term period. In fact, the availability of an exhaustive and completely revised digital information base is indispensable to be able to reach analysis levels which are really deep and productive.

Parallel to this activity it is intended to continue the construction of the system, in particular completing the data interchange mechanisms among the applications in accordance with the conceptual architecture of reference (figure 1a) already mentioned. In this connection it should be noted that the use of Visual Basic has so far supplied more than satisfactory results thanks to the easy use of this language and the availability of numerous libraries in each software package.

However, the further development will probably need reflection and the consequent adoption of more powerful programming instruments which make it possible to construct a more flexible and stronger application. Specific reference is made to the object-oriented languages, in particular in Java, and to XML, which has now become the de facto standard for data interchange. In fact, the last target is to construct a real application at system level which on one side substitutes the present programming routines which have been developed in each single program, and on the other makes it possible to obtain a greater flexibility of the system thanks to the possibility of exportation in standard formats. XML could also be used in the coding of all the information of conversational or narrative nature (for example the excavation diaries) which are generally not found in the stiff structure of a relational database but which often form a documentary source of primary importance.

There is still a lot of work to be done and the choice of personally managing the creation of the system definitely involves an initial extension of the times connected with the need to learn new languages and technologies. However, this choice makes it possible to always have the complete control on the developments of the system and surpass the simple test phase to progressively reach a solution which, reflecting the real needs of the research, is really productive.

Notes

[1] Excavations in the Etruscan trade center of Forcello di Bagnolo S.Vito (MN) and in the bronze age pile-dwelling settlement of Lavagnone di Desenzano del Garda (BS).

[2] Subject of the graduation thesis of the authors: Mantegari 2003, Quirino 2003.

[3] The software used is ESRI ArcGIS 8.1.

References

Fronza, V., 2000. Il sistema di gestione degli archivi nello scavo di Poggio Imperiale a Poggibonsi. Una soluzione all'interno della "soluzione GIS". *Archeologia e Calcolatori* 11, 125–137.

Fronza, V., Nardini, A., Salzotti, F. and Valenti, M., 2001. A GIS Solution for excavations: experience of the Siena University LIAAM. In Stančič, Z. and Veljanovski, T. (eds), *Computer Archaeology for Understanding the Past CAA2000, Proceedings of the 28th Conference, Ljubljana, April 2000*, Oxford, ArchaeoPress. BAR International Series 931, 173–177.

Longhi, C. and Mangani, C., 2003. GIS spatial analisis in the etruscan trading centre of Bagnolo S.Vito (Mantova, Italy). In Fischer Ausserer, K., Börner, W., Goriany, M. and Karlhuber-Vöckl, L. (eds), *Enter the Past CAA2003, Proceedings of the 31th Conference, Vienna, April 2003*, Oxford, ArchaeoPress. BAR International Series 1127.

Mantegari, G., 2003. La programmazione di un database relazionale come strumento per gestire ed interpretare I dati di scavo. Il settore B del Lavagnone di Desenzano del Garda (scavi 1991–2002): un esempio applicativo. Tesi di Laurea, A. A. 2002–2003.

Mantegari, G., in press. Verso la costruzione di una "soluzione GIS" per il Lavagnone: la sperimentazione sul settore B. *Notizie Archeologiche Bergomensi*.

Mantegari, G. and Quirino, T., in press. Un database relazionale per la gestione dei dati di scavo. Analisi, programmazione e sperimentazione nell'ambito delle ricerche della Cattedra di Preistoria e Protostoria dell'Università degli Studi di Milano. *Notizie Archeologiche Bergomensi*.

de Marinis, R. C., Baratti, G., Mangani, C. and Longhi, C., 2002. L'insediamento etrusco del Forcello di Bagnolo S.Vito (Mantova). In Peretto, C. (ed.), *Analisi informatizzata e trattamento dati delle strutture di abitato di età preistorica e protostorica in Italia*, 303–318.

de Marinis, R. C., Sidoli, C. and Rapi, M., 2002. Il Lavagnone di Desenzano del Garda (Brescia). In Peretto, C. (eds), *Analisi informatizzata e trattamento dati delle strutture di abitato di età preistorica e protostorica in Italia*, 243–256.

Quirino, T., 2003. La programmazione di un database relazionale per la gestione della documentazione di scavo: analisi, revisione ed elaborazione dei dati del Forcello di Bagnolo S.Vito (MN) – Settore R18, Scavi 1988–2002. Tesi di Laurea, A. A. 2002–2003.

Valenti, M., 2000. La piattaforma GIS dello scavo. Filosofia di lavoro e provocazioni, modello dei dati e "soluzione GIS". *Archeologia e Calcolatori* 11, 93–109.