Creating of a database for prehistoric sites: which are the goals, the strategy and what means to put in place?

Pierre Corboud
Dpt of Anthropology and Ecology
Geneva University
Case postale 511
CH-1211 GENEVE 24, Switzerland
Phone: +41 22 702 69 89 - Fax: +41 22 300 20 46 - E-mail: Pierre.Corboud@anthro.unige.ch

Abstract: Nowadays, computer science allows the creating of archaeological database to enable the gathering, updating, and management of all kinds of information necessary for the research and synthesis in the field of archaeology. However, in view of the past experiment in the seventies, one must be very cautious and be aware that the «archaeological data banks» must be plentiful, maintained, and exist for years before giving any profitable results. We wish to discuss the problems of the data pertaining to archaeological sites, especially those from prehistoric time. The archaeological site constitutes the ideal unit of description and study to deal with theoretical and methodological aspect of the creation of a relational database. The relationship between the geographical chronological and documentary entities imposes the use of very precise options in function of the chosen goals. We intend to show what could be the line of reasoning behind those options with the help of some real world examples and an ongoing project.

Introduction

The database from archaeological sites, and particularly those of prehistoric sites require the organization and implementation of a specific type of archaeological database. Indeed, the constitution of a database of archeological sites can be justified according to extremely varying goals and concerns. It is, for instance, the process of gathering and laying out of information relating to sites from the same geographical area, belonging to the same cultural phase, in order to study the diffusion of a culture. It could also be the assembly of all the data relating to the sites from one county or from one province in order to manage and protect the saved remains. Thus, to simplify, one can distinguish between the goals relating to safeguard of the patrimony and those directed rather towards the scientific research. Sometimes, these two categories are more or less the same, but in other cases, the options chosen in the creation of such a database are relatively divergent, especially when it comes to choosing and organizing the various fields. We won't go on developing this opposition, but rather we will attempt a systematization of the theoretical and strategic steps, which will lead to a management or research tool adapted to the needs of the users. Moreover, it is essential that this tool guarantee a certain perennity of the data recorded in the base. Indeed, the lifetime of a database is a very important criteria when it comes to its creation out of the necessity to facilitate the unavoidable upgrades and to allow and easy evolution towards other database management systems. It is not useful, here, to mention the software packages and applications relating to managing such databases. Currently, the technical evolution of the hardware and the software is so fast and unforeseeable that it is not desirable, at least in the first phase of elaboration, to create a database according to a precise computer science environment. It is, however, preferable to seek to ensure the structural solidity of the data model, according to the environment of the research or the management of the objects to be described in this system. This last remark is certainly obvious to all of you, but it is not useless to point out it ...

As always in archaeology, like in the majority of the scientific fields, the success of a project depends mainly on the adequacy between the objectives, the characteristics of the available data and the environment of the research in the broad sense.

Accordingly, the principal goal for which the database of the sites will be created is essential in the definition of the various documentary modules and the relations to be established between them, especially when it comes to the input modules and output products. By documentary module we mean all the fields or headings of a data table, which share the answer to the same type of question.

The basic model

Independently of the choice of the goals and the definition of the environment of the research, the structure of the various documentary modules can be summarized in the definition of the basic model (fig. 1). This basic model is intended to answer the three following questions :
1) Where is the site?
2) When was it occupied?
3) What are its intrinsic characteristics?

This basic model thus gathers, in the shape of three modules of information, the essential data or interpretation, which one can expect to find in the description of an archeological site. On the other hand, this information does not give any details on the source of the data used for the description of the site. It is, in fact, more a synopsis of the main fields and proposed interpretations.

Such a model would present very little interest within the framework of a data management in the short or long term. Indeed,

each update would require a return to the primary data, which are not contained in the table but in external documents, and which, is not always referenced.

This defect, fatal for the long-term survival of such a system, must absolutely be corrected by the implementation of an extended model, which gives reference to the sources, or, still better, which contains the sources of the information needed for the descriptions or the interpretations relating to the archeological sites.

The extensive model (fig. 2)

Indeed, one cannot consider that an archaeological data (for example a layer of habitat belongs to a specific culture) is an ⌐absolute fact⌐whose validity cannot be questioned. Each interpretation or conclusion is the result of an assessment based on a set of information and documents. This assessment could be modified several times by the appearance of a new data or simply the intervention of a more qualified expert could modify. There too, it shows evidence, which is too seldom taken into account in the most widespread documentary applications.

Thus, the extensive data model should integrate the categories of the following information to allow for a easy return to the sources :

A) References to the articles, works, reports of excavations, maps, graphical documentation, etc. from which the 'archaeological facts' were extracted ;

B) The essential information textually extracted from these sources, or more precisely the transcription of the quoting of articles or works which provide significant data, whose interpretation could be ambiguous or which represent useful information on the history of the research and interpretations of the site. In this category we will retain for example the citations of texts who mention the locations of formerly spotted sites, but whose localization is vague or unknown.

Generally, it is always preferable to transcribe an original citation rather than to interpret it and translate it according to our system of today's reference. The interpretation in a homogeneous system, for example in a typo-chronological system of common use, must intervene later on and in another documentary area accompanied by a reference to its author. We shall come back later about this rule which could be named 'traceability principle of data and interpretation.'

Historic of knowledge

This last remark brings us in another field of documentary analysis which is historical knowledge on an archeological site, as much as to the physical operations realized on this site, the results obtained or as to subsequent interpretations. In this type of module, we find the information relating to discovery of the site, to the studies carried out (survey, excavations, analyses of artifacts, etc.) from which it is possible to identify the

results and interpretations and, finally, old interpretations resulting from these various interventions.

Thus, the incorporation of these new documentary modules in a extended data model should allow the update of such a database permanently whatever the type of objective one considers for the database of sites, without requiring a new document search, but simply by supplementing the references to the sources and by correcting out-of-date interpretations and conclusions.

Importance of the relation to the bibliographical references and to the documents

With the exception of the ground, all the articles, reports and works that mention the observations, constitute the principal documentary source of an archeological site and the interpretations gathered on this site since its discovery. One of the essential characteristics of the archaeological research, in opposition to certain exact sciences (such physics and biology) is to not be able to neglect the writings of the former authors. Indeed, the archaeological observation – in particular the excavation – must destroy its object of study to describe it and understand it. This remark is even more relevant for the sites not excavated, described in the past, but partially or completely destroyed by the action of nature or the man. In this situation, the mention of the bibliography of reference related to a site described is an essential and mandatory element in any database of archeological sites. Moreover, only the relation between each work of this bibliography, the page or the figure, will make it possible to retrieve the source of information which made it possible to answer the questions included in the basic database model.

The articulation of the site database with a bibliographical reference table, which includes also the gray literature, is a central element of this type of applications. This goes as well for all the non-bibliographical documentation (maps, drawings, collections of archaeological artifacts etc,) the mention of which is necessary as well as the indication of their location.

Relation with the collected archaeological artifacts

Independently of the bibliographical references to the archaeological artifacts collected or observed on the site, the mention of the artifacts preserved in the museums is essential to discuss the dating of the various levels and successive occupations. In these applications, the goal of the relation artifacts–site is not to carry out a museographic inventory of all the artifacts collected and attributed to each of the sites of the database. Such an objective is not only utopian, but also especially inadequate in this context. It is rather preferable to mention the sets of artifacts preserved in each depository, sorted by their categories and their typo-chronological dating.

Before treating relations between the various modules and their logical articulation, it is necessary to open a parenthesis to define the significance, which we give to an archaeological site, which is the unit of elementary description.

Definition of the elementary unit of description

In the applications which are used as example here, the unit of elementary description corresponds to the definition of the archeological site as physical, topographic and logical entities, whose geographical and chronological limits can be traced with more or less precision. It is for example about a surface of ground on which the existence of a Neolithic village is known. There, later, a necropolis of the Bronze Age was built. In this case, the unit of description is the topographic layout, which include all of the vestiges of these two periods and these two types of sites (fig. 3). The limits of this definition are obvious. They prohibit to easily distinguish the characteristics from each identified archaeological level and to describe for each occupation the archaeological artifacts discovered and the structures observed. In this case, this limitation is desired, because the objective is to treat a whole of sites and not a whole of occupations belonging to the same site. Moreover, the attribution of several occupations to an archaeological site is an interpretation and not a fact. Thus, the various chronological horizons recognized as well as the characteristics of those horizons should be described in a secondary table belonging to each site's record.

It should be noted that the conservation of a site is not necessary to its description according to such a definition. A destroyed or displaced site, for which we have enough precise information to locate the old position of it, the dating and the characteristics will have as much its place in our database that a site well preserved, studied or not.

Some principles of relations between the documentary modules and the fields

Before approaching the various logical structures between the units of information of the database of the sites, it is useful to propose and explain some principles, which govern the construction of these relations (fig. 4).

The first principle could be named 'recognition of the data sources'. It is a question of allowing, in all the cases, the access to the primary data, be it bibliographical, graphic or constituted by collections of artifacts. This principle is rather easy to apply as long as the documents and artifacts concerned are catalogued in known systems of inventories (libraries, documentary archives, museum's catalogues, or photographic records.)

The second principle is that of the 'traceability of interpretations'. Its strict application is more difficult, because, in this view, each term in each field should mention its author, the date and the data used for this interpretation. Practically, this principle can be simply solved by mentioning the successive dates and responsibilities for the data entries and the corrections and by admitting that the last expert has read again the whole record for the site and takes the responsibility for its contents.

The third principle, which is corollary of the first two, is to guarantee the logical and graphic distinction, in the screens of the system of database management, between the modules and the fields which contain raw data resulting from the documents and observations, and those which contain interpretations or expertise. Such a distinction is essential as much for the people who enter the data than for those which question them. One could name this principle 'separation of the raw data and interpretations'.

From the various data modules quoted and of the three principles stated here, it is possible to build a data model adapted to the goals that we fixed ourselves, by limiting to the maximum the number of tables.

Logical structure of the various documentary modules, models simplified

This model takes into account only the primary data to be integrated in an archeological site's database (fig. 5). Moreover, this model is in conformity with the recommendations of the CIDOC (CIDOC 1995) One can thus distinguish various modules :
- The topographic module, in the table of sites, which includes the fields, which provide the localization administrative and geographical of the site, with the mention of the sources from which the co-ordinates were established.
- The bibliographical module, which gathers the extent of the references to the texts and documents relating to the site, as well as the quotations chosen in these texts.
- The historical module, which resume the succession of observations and studies listed for this site.
- The module of dating, with all information, not interpreted which makes it possible to attribute the occupations to a culture or a given period.
- Lastly, the module of interpretation, which presents a synthesis of all the significant data and proposes a chronological attribution of the occupation of the site.

A sample application : the ARCHEODAT database

At this point we are through with the theoretical considerations on the archeological site's database, in order to briefly present a concrete application, which started in Western Switzerland twenty years. At its conception, the goal of this project was to gather all the informative data resulting from an extensive archaeological survey on the littoral sites of the Lake of Geneva, Morat and Neuchâtel, and to integrate the results of this operation. It is now a management tool on archaeological knowledge about a set of sites, as well as a referential database, allowing the access to the original data, which made it possible to elaborate this knowledge.

The archeological sites that we consider here were occupied between the middle of Neolithic period and the end of the Bronze Age, between 4000 and 800 years before Christ. None was excavated completely, but all were the subject of observations or small surveys, which made it possible to gather a certain number of information on their situation, their dating and their nature. One of the characteristics of this project was that the

majority of the studied sites had been discovered between 1854 and 1900, particularly for the south bank of Lake of Neuchâtel and the Lake of Morat. For the majority they were not the objects of studies or observations during this period. As well as the recent observations carried out on the ground by our team, the old data thus occupy a significant place.

With regard to its structure and its interface, the ARCHEODAT database underwent several transformations, in order to regularly adapt it to the new management tools of computer sciences, but also to the evolution of the questions and goals of our project. Currently, after having developed the last application with the MS-Access software, we are in the process of transferring all the data in an Oracle database, which permits an access by the Web, far beyond the limits of the limited local network of the University of Geneva.

The simplified structure of the data of ARCHEODAT is as follows, with the articulation between four principal tables, which are the table of the sites, the table of the bibliographical references, the table of the 14C dating and that of the dendrochronological dating. In detail, the fields are distributed in various modules, which are :
- The topographic module, with the topographic co-ordinates and the references to the source of these co-ordinates.
- The bibliographical module, which is in fact the table of the bibliographical system of management of our department.
- The module of the ancient citations.
- The module of the historic of the studies, which is also a reminder of the main results of recent observations.
- The module of the 14C dating, which proposes the standard dates, but also the calibrated dates.
- The module of the dendrochronological dating, essential to treat sites preserved in wetland.
- The module of synthesis and interpretation of the data,

which summarizes it.

Lastly, as a conclusion with this presentation, homage should be paid to a Swiss prehistorian, who was perhaps despite himself, the inventor of the database of archeological site. Frederic Troyon was, indeed, the first to circulate in the Kanton Hof Vaud a questionnaire about the antiquities and archaeological richness, in order to collect the data necessary to the protection and the study of the sites of his land. This endeavor has enabled the creation of the first archaeological map of the Kanton Hof Vaud, published in 1874. This is the prototype of our current archaeological and GIS database (Troyon 1858 and 1860, Bonstetten 1874)

Also, during a stay in Sweden between 1845 and 1846 F. Troyon had the merit of proposing this questionnaire to the Royal Prince of Sweden, to be used as model for the archaeological inventory of the archeological sites of the Kingdom of Sweden. This marked an historical and initial point of the management and the study of the archaeological patrimony of Sweden.

References

Bonstetten (G. de). 1874. *Carte archéologique du canton de Vaud.* Toulon : C. Mihière.

CIDOC-ICOM. 1995. *Projet de norme documentaire internationale pour les sites archéologiques. Draft international core data standard for archaeological sites and monuments.* Paris : ICOM.

Troyon (F.). 1858. Statistique des Antiquités de la Suisse Occidentale. *Indicateur d'antiquités suisses*, 2, juin, 25-29.

Troyon (F.). 1860. *Habitations lacustres des temps anciens et modernes.* Lausanne : G. Bridel. (Soc. d'histoire de la Suisse romande, Mémoires et documents ; 17).
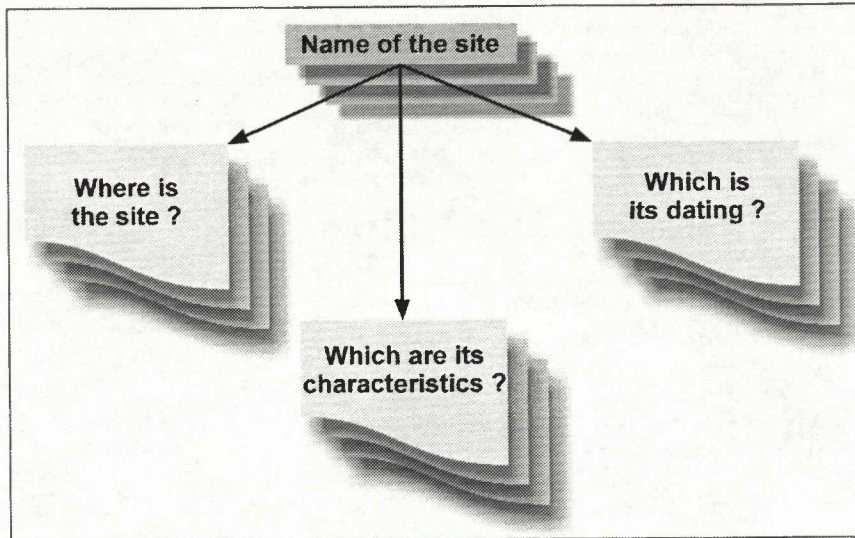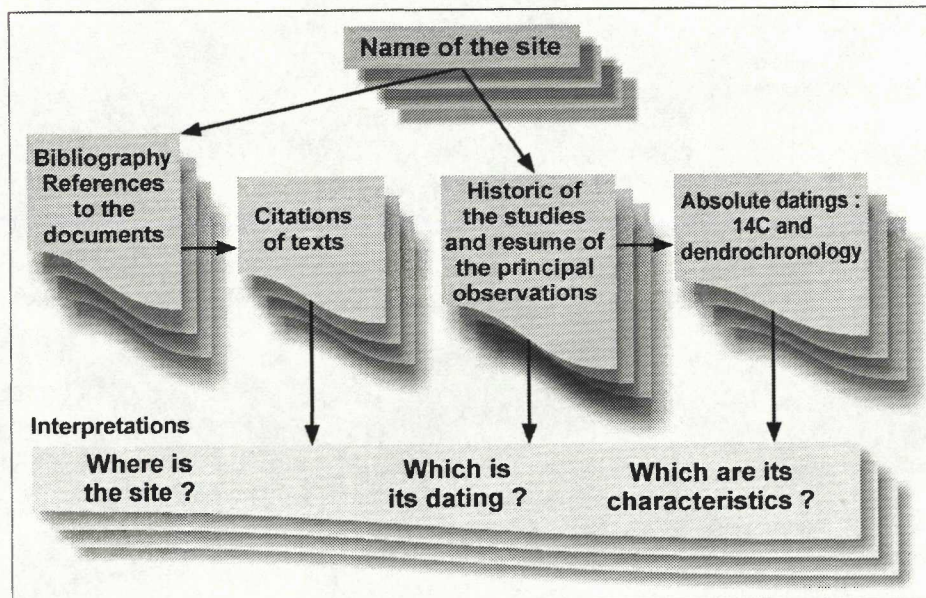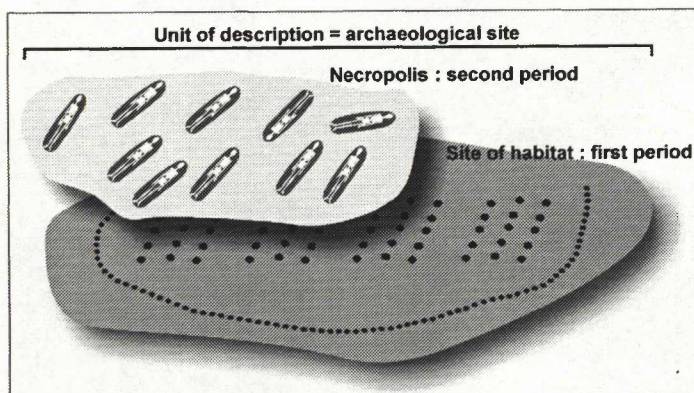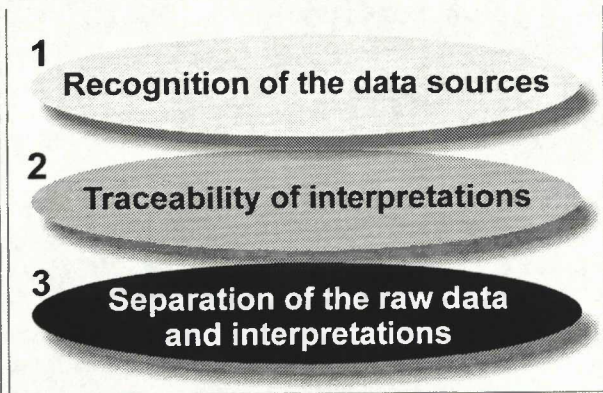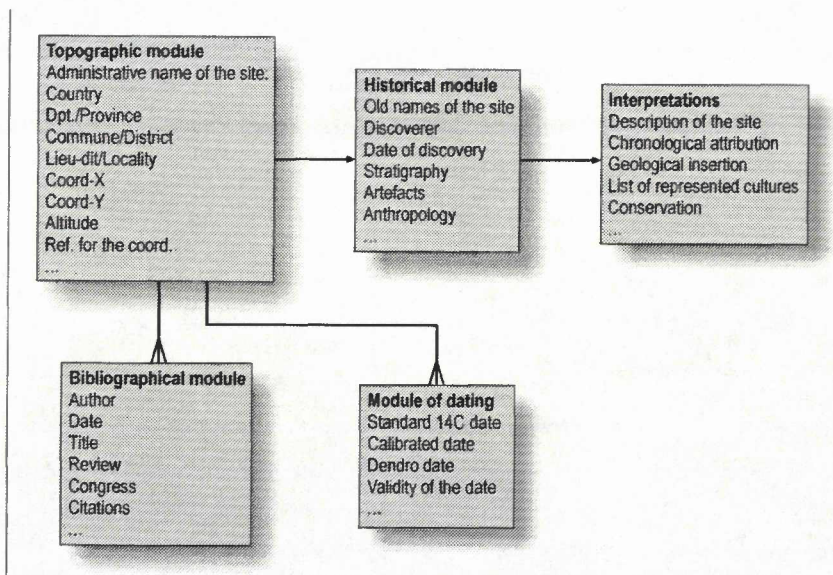
Figures

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Topographic module
Administrative name of the site.
Country
Dpt./Province
Commune/District
Lieu-dit/Locality
Coord-X
Coord-Y
Altitude
Ref. for the coord.
...

Historical module
Old names of the site
Discoverer
Date of discovery
Stratigraphy
Artefacts
Anthropology
...

Interpretations
Description of the site
Chronological attribution
Geological insertion
List of represented cultures
Conservation
...

Bibliographical module
Author
Date
Title
Review
Congress
Citations
...

Module of dating
Standard 14C date
Calibrated date
Dendro date
Validity of the date
...

Figure 5.

ICONODAT   **BIBANT**   **ARCHEODAT**

ICONODAT
Image database

ARCHEODAT
Topographic module

ARCHEODAT
Module of dendro
dating

BIBANT
Bibliographical
references

ARCHEODAT
Module of 14C
dating

ARCHEODAT
Historical module

ARCHEODAT
Module of ancient
citations

ARCHEODAT
Interpretations

Figure 6.

	ICONODAT	BIBANT	ARCHEODAT
Public	ICONODAT - PUBLIC Image database	BIBANT - PUBLIC Bibliographical references	ARCHEODAT - PUBLIC Archaeological sites "In progress"
Producers	ICONODAT Image database "In progress"	BIBANT Bibliographical references	ARCHEODAT Archaeological sites
Administrators	ICONODAT - ADMIN Image database	BIBANT - ADMIN Bibliographical references	ARCHEODAT - ADMIN Archaeological sites

Figure 7.