Archaeological databases: what are they and what do they mean?

Robert Schlader
Idaho State University
Idaho Museum of Natural History
344 north 15th Ave Apt. 11B
Pocatello, Idaho, 83201
Phone: +1-208-478-2642 - E-mail: schlrobe@isu.edu

Abstract: In order to understand what an archaeological database is and what it means we have to do a few things first. We need to discuss what a database really is. Once done with that, we need to discuss what makes a "good" database; specifically what we need to do to ensure that the information we distill from the data contained therein is both good and useful. Following this we take a brief look at the ways in which databases are currently being used in archaeology. Then we will look at a fairly new approach to archaeological databases, which is attaching an archaeological database to a Geographic Information System (GIS) for a better method of creating and displaying our information. A problem exists with this however; there is a problem with META-DATA. We will look into this problem and then dissolve it. When all this is complete, we will be at a spot where we can begin to (finally) talk about the theories driving the use of databases in archaeology. When this is complete we will have a closer fit for what databases are, and what they mean for archaeology.

Key words: Databases, GIS, METADATA, DBMS, American Archaeology

Introduction

Archaeologists require a method to organize and store their data. One method commonly used is to record everything into a series of notebooks and draw maps. Another method is to create and fill an electronic database. Both can be used to access the data stored in them, but only one can re-organize that data on the fly. This is only one benefit of using a database. Currently most archaeologists spend little to no time actually thinking about their databases. Most archaeologists will do no more than consider what few pieces of information they feel they need to look at to answer their research question (personal communication with Richard Holmer, November 2000). This lack of thought about proper data design has left archaeology lagging behind the rest of the academic world (Carroll, 2000). This lack of consideration about proper data design has been the cause of a major difficulty among groups who are attempting to share information, especially over the internet (Eiteljorg, 1997). When we consider this, it begins to look strange that archaeologists really don't understand what databases are. What archaeologists do understand are the theories of materialism, functionalism, processualism, post-processualism, and other "-isms." What we will find is that, even though not aware of it, archaeologists utilize these theories in their databases, and that they tend to design their databases to exploit whichever theory they happen to be using to ask their research questions.

Welcome to databases

A database is a shared, integrated computer structure that houses collections of end user data (raw facts that are of interest to the end user) and Metadata ("data about data" through which the data are integrated). In a sense, a database resembles a very well organized electronic filing cabinet in which powerful software, known as a Database Management System (DBMS), helps manage the cabinet's contents (Coronel, 1997, p. 4). Information is the key to our modern world; even the world of archaeology, and information is constructed by the proper manipulation of raw data. Databases not only store the raw data that we collect, they allow us to manipulate that data in very complicated ways (Codd, 1974, 1990). We can re-configure the way our data is displayed on the screen, the way it is printed out, and even the way it is stored on the computer. Through these manipulations we create the information that we need to answer our research questions, as well as appease those that hold the purse strings of our budgets (Date, 1995).

Making a database work correctly

The currently popular method for creating a good database is to create a database using a Relational DBMS (Coronel, 1997). What allows relational databases work are a few simple concepts. These concepts are: keys, controlled redundancy, entity integrity, and referential integrity (Coronel, 1997). The way that the data is manipulated in the database is through the use of "keys." Primary keys uniquely identify some piece of data (which is known as entity integrity)(Coronel, 1997), and foreign keys link one primary key to another by placing a copy of one table's primary key into another table (which is known as referential integrity)(Coronel, 1997). This "copied key" allows the two entries in two different tables to refer to one another, and thus a relationship is created between the two tables (and hence the name relational database)(Coronel, 1997). By only allowing small parts of the tables to be copied the redundancy amidst the data is reduced to a controllable level (strikingly enough this is called controlled redundancy)(Coronel, 1997).

And now for something completely different: how archaeologists use databases

In archaeology there are two main reasons for having a database. One is for the long-term storage of data, this is better known as a data warehouse or an electronic archive (Eiteljorg, 1997). The other is for the express use of researchers and other active users of the data, better known as a "user group." It is a fact that most researchers maintain independently designed, widely varying databases for the express purpose of answering their specific research question (Eiteljorg, 1997).

The differences between the types of databases are not necessarily a problem until that database is finished being actively used. When the data is no longer in use, where does it go? What happens to it? Most researchers simply save a copy of their data on whatever medium they find handy, send that copy to a museum (if in fact they do anything with the copy), and that's the end of that (Robinson, 2000).

There are a number of data warehouses available other than museums. One is the Digital Archiving Pilot Project for Excavation Records, which is better known as DAPPER (Robinson, 2000). Another is Eiteljorg's Archaeological Data Archive Project or simply ADAP (Eiteljorg, 2000). Both of these digital archives will take a user database and maintain it for long-term storage and re-use.

GIS: a new perspective for information dissemination

There are a small number of new technologies that have come into their own recently that are of great interest to archaeologists concerning data. One of these new technologies is the Geographic Information System, or GIS, but what is a GIS?

In essence a GIS can be said to be a database that contains some sort of spatial information that then can be used to create a map for display, manipulation, and presentation (my definition). Where this comes in to play is that most archaeologists are collecting spatial data about the objects they uncover. This spatial information not only (potentially) covers large areas, but also may extend over many layers of the site. Due to this glut of spatial information, GIS becomes a godsend when it comes time to map out object distributions and densities as well as creating a series of overlays to account for site depth. What this does is allow for identification of spatial patterns within a site, and will (hopefully) facilitate future regional analysis and intersite comparisons of archaeological data (Moyes and Awe, 2000).

METADATA: the glue that holds a database together

One key to the way that GIS software functions is through the use of METADATA. In a standalone database the metadata for the database consists of two elements: the data dictionary, and the data definition language.

> The *data definition language* is the part of the DBMS that allows the user to set up a new database, to specify how many attributes there will be what the types and lengths or numerical ranges of each attribute will be and how much editing the user is allowed to do. This establishes the *data dictionary*, a catalog of all of the attributes with their legal values and ranges (Clarke, 1997).

What most GIS software does is to extend this simple set of metadata to include other information as well. This additional data is of a more general nature than that contained in the database's data dictionary (Weibel, 1995). The extended metadata contains information about the nature of the database, the location of the data in the database, the existence of similar information, what data the database contains, how you can view the data in the database, and who it is that holds and maintains the data (Wise and Miller, 1996) among other things.

The extended set of metadata allows us to do three things that the data dictionary does not allow. First it allows us to find out what the data set covers without having to look at the data. Second, it allows us to locate data faster. Third, it allows us to group the data into larger groups based on the similarity of the dataset (Wise and Miller, 1996). What this allows us to do, once the metadata is filled out, is narrow our search to exactly the elements in the database that we need to look at without actually having to sort through the data.

The true benefits of using metadata in this way only become evident when one is attempting to get information out of databases from different locations (specifically online databases), or different databases (such as an excavation database versus a data warehouse). In essence, the more abstract the metadata the faster we can narrow our searches and find the information we are looking for.

So, if metadata is used to speed up the process of narrowing down a search for information, what kinds of problem(s) can there be with that? There are really two problems with metadata. First there is the problem of a consistent set of fields, i.e. each database has the exact same set of attributes (date entered, type of object, name of person entering the data, etc.) in their metadata (Wise and Miller, 1996). Second there are the allowable entries within each field in the metadata, i.e. whether the "type of object" field will allow "Desert Side-notched", "DSN", the number "3" which codes for Desert side-notched, or any other of many different names for this kind of object (Wise and Miller, 1996).

The first of our two problems with metadata is a real problem that can only be solved by establishing an agreed upon structure. There is a lot of work being done currently to solve this problem (Wise and Miller, 1996; Cathro, 1997; Weibel, 1995). The second of our two problems is in fact a very old problem that probably will never be solved. This old problem is the problem of typology that has plagued archaeology for centuries. With the current state of DBMSs all we need to do is simply hardcode a lookup table that contains all the possible entries for any typological conflict that may occur for any object classification.

Unfortunately the proponents of each individual typology refuse to come together in agreement for what to call these objects for a variety of different, and always-personal reasons. Once we come to terms that most of the fighting concerning a standard structure for archaeological metadata is in fact a battle over typology we can then, hopefully, come to a final agreement on the much needed metadata standard. Once this standard is agreed upon and implemented there will no longer be a problem with metadata.

What databases really mean: Archaeological theory through databases.

Now that we know what databases are, why they are the way they are, what neat things we are beginning to do with them, and the major problems we are having with them, we can finally get to what are the driving theories behind the use of databases in archaeology.

Most archaeologists truly don't think much about what theory is driving their use of databases; they just use them as one of a suite of tools to answer their main research question. What they don't realize is, however, that they are in fact following one of three fundamental theoretical perspectives for databases. The three theories that I have identified that archaeologists use are: first, a Functionalist or Marxist approach that looks at the data as being nothing more than a representation of the material being described; second, a cognitive approach that gives the raw data to the observer in a form that the observer can then compile in his or her own way so the observer can build his or her own interpretation; and third, a stylistic or artistic approach that strives for greater clarity or aesthetic appeal in the presentation of the dataset as a means of providing visual support to the arguments based on that dataset.

These three approaches differ on one very important aspect. This aspect can be boiled down to how the researcher uses the database to construct his or her interpretations of the dataset. Someone using the Functionalist or Marxist approach views databases as either a means to an end, i.e. an analysis of the information that can be created through the manipulation of the data contained in the database (which doesn't differ in the slightest from the way a paper database is viewed), or as an electronic substitute for the actual objects themselves that then can be manipulated to help create an acceptable story. The key to this approach is the speed with which the final analysis can be created. A paper database requires much more time to "sift" for the juicy tidbits of data that are needed to make a compelling story. Rapid returns of analysis are dictated by the time constraints of contract archaeology and Cultural Resource Management in the US, and this then, requires the use of a method other than paper for their database. The financial concerns of not running over budget while still producing a thorough analysis dictates a methodology that requires the shortest amount of manipulation of data, and a Functionalist or Marxist approach fits this nicely.

The cognitive approach offers an alternative to the Functionalist or Marxist approach that some in the US and around the world are turning to. This approach normally provides a short overview of the data contained in the database, and then provides access to the entire dataset so that the dataset can be manipulated by anyone. For Cultural Resource Management in the US this allows the contract archaeologist a method of providing the desired information to their employers without having to go through the expense of actually doing the analysis, as well as providing a kind of "buffer" between the researcher and the financial managers. For others around the world this method is a way that they can expand their research by allowing multiple perspectives to be brought to bear on the dataset. It is felt that with multiple perspectives looking at the same data at the same time that a better view of the information contained within the dataset will arise. As one can imagine, most archaeologists using the cognitive approach to their databases are providing the interface to their data online via a web accessible design.

The third approach I am calling the presentational approach, though I feel that term is somewhat misleading. This approach is one in which the database is used in a manner that helps to create a better public presentation of the researchers conclusions, not simply as the data repository from which the information that makes up the conclusions were drawn. For obvious reasons most archaeologists who utilize GIS databases use this approach when creating the high quality maps that accompany their reports. By creating a better visual representation of the information contained in the database than a chart or graph provide, more information is conveyed with fewer typed words. This then in turn allows the researcher to spend more time on the actual results of his or her research than on the explanation of what the data contains. The drawbacks of this approach are high dollar costs, and time spent in the creation of the presentation. To utilize the presentational approach one must purchase the software and hardware that will allow the presentation to be created, which can be quite expensive. Then a considerable amount of time must be spent in training and the use of the software, and production of the actual final products. Once complete the rapid production of information rich visual representations of the data makes it all worthwhile.

Conclusion

What we find when looking at this is that most archaeologists don't consider their database as having any real bearing on the theories they use to develop their conclusions about their data. We have seen that this is not the case at all. A well-designed database can aid the archaeologist develop an analysis of his raw data that may not have appeared other wise. In fact most archaeologists develop their databases based directly on the desired end product that they are pursuing. Unfortunately for archaeology most of the end product being produced in the US is for cultural resource management, funded directly by the state and federal agencies that are only interested in how fast a project can be completed. This has caused archaeologists in the US to stop asking questions that require data design beyond a simple count of artifacts. Once archaeologists begin to recognize that they have been unconsciously constructing their databases biased on quite solid theoretical foundations they should be able to fine-tune their use of databases to generate more information than ever before.

References

Anderson, D. G. and Faught, M. K. 1998. *A North American Paleoindian Projectile Point Database.* http://www.anthro.fsu.edu/special/paleo/paleoind.html

Baker, Tony. 1999. *Digital Crabtree: Computer Simulation of Folsom Fluting.* Online publication. http://www.ele.net/algor/crabtree.htm

Barker, Richard. 1990. *CASE Method Entity Relationship Modeling* (Addison Wesley Publishing Company,), clothbound; ISBN 0-201-41696-4.

Baumann, Steven M. 1999. *Integrating GIS and Cultural Resources Databases for Archeological Site Monitoring.* Journal Info Crm: cultural resource management. Volume 22, Number 933. SICI 1068-4999(1999)22:9L.33:IGCR;1-

Burrough, P. A. 1986. *Principles of Geographic Information Systems for Land resource Assessment.* Oxford: Clarendon Press.

Carroll, M. 1999. *Preserving Archaeological Digital Data: Report of the NCPTT Working Group.* From *SAA Bulletin*, vol. 18, no. 5.

Cathro, Warwick. 1997. *Metadata: An overview.* National Library of Australia. http://www.nla.gov.au/nla/staffpaper/cathro3.html.

Clarke, K. C. 1995. *Analytical and Computer Cartography.* 2nd edition. Upper Saddle River, NJ: Prentice Hall

Codd, E. F. 1974. *The Relational Approach to Data Base Management: An Overview*, Third Annual Texas Conference on Computing Systems.

Codd, E. F. 1990. *The Relational Model for Database Management* (Addison-Wesley Publishing Company,), 538 pages; clothbound; ISBN 0-201-14192-2

Coronel, Rob. 1997. *Database Systems: Design, Implementation, and Management. Course* Technology, a Division of International Thomson Publishing. Massachusetts.

Clarke, Keith. 1997. *Getting Started With Geographic Information Systems.* Prentice-Hall, Inc. Simon & Schuster / A Viacom Company. Upper Saddle River, New Jersey.

Crane, Gregory. 1991. *Hypermedia and the Study of ancient culture.* Journal Info Ieee computer graphics and applications. JUL 01 1991 v 11 n 445

Date, C. J. 1995. *An Introduction to Database Systems* (Addison-Wesley Publishing Company,), 839 pages; clothbound; ISBN 0-201-54329-X.

Duecker, K. J. 1979. *Land Resource Information Systems: a review of fifteen years experience.* In *Geo-Processing*, vol. 1, no. 2, pp. 105-128.

Eiteljorg, Harrison. 1997. *Electronic Archives.* http://intarch.ac.uk/antiquity/electronics/eiteljorg.html.

Eiteljorg, Harrison. 2000. *If We Preserve the Files, Who Will Use Them?* http://csa.brynmawr.edu/saa/saa-adap.html

EPA. 1997. *EPA Scientific Metadata Standards Project.* http://www.lbl.gov/~olken/epa.html.

Gilman P. 1997. *Securing a Future for Essex's Past. ArcUser* Magazine, http://www.ersi.com/news/arcuser/1099/essex.html

Heyworth, M. P., Richards J., & Winters J., Eds. 2000. *Internet Archaeology: Where Next?* http://csa.brynmawr.edu/saa/saa-ia.html

Hodder, I. 1999. *The Archaeological Process: An Introduction*

Lohse E.S. and D. Sammons. 1997. *DIGITAL STONES: A Basic Introduction to the Analysis of Stone Tools.* Beta version CD-ROM.

Lohse E.S. and D. Sammons. 1997. *A computerized Database For Lithic Use-Wear Analysis.* BAR Conference, 1996.

Lohse E.S. and D. Sammons. 1997. *Metalanguage or Metabaggage: Archaeological database constructions in cyberspace.* Paper presented at the 63rd annual SAA meeting, Seattle, March 25-29, 1998.

McCartney, P., Robertson I., & Cowgill G. 1997. *Using Metadata to Address Problems of Data Preservation and Delivery: Examples from the Teotihuacan Data Archiving Project.* http://csa.brynmawr.edu/saa/mccartney.html

Moyes, H. and Awe, J. 2000. *Spatial Analysis of an Ancient Cave Site. ArcUser* Magazine, http://www.ersi.com/news/arcuser/1000/cave.html

Robinson, D. 1997. *Digital Archiving Pilot Project for Excavation Records (DAPPER)* http://csa.brynmawr.edu/saa/dapper.html

Star, J. and Estes J. E. 1990. *Geographic Information Systems: An Introduction.* Upper Saddle River, NJ: Prentice Hall.

Weibel, S. 1995. *Metadata: The foundations of resource description.* D-Lib Magazine, July. http://www.dlib.org/dlib/July95/07weibel.html.

Wise, A. and Miller P. 1995. *Why metadata matters in archaeology.* Internet Archaeology 2. http://intarch.ac.uk/journal/issue2/wise_toc.html.

Wolle, A. 1997. *Çatalhöyük: Excavations of a Neolithic Anatolian Höyük* http://catal.arch.cam.ac.uk