ARCHAEOLOGY, DATA STRUCTURES AND COMPUTER SCIENCE

Torsten Madsen

Introduction

The computer has been around for some forty years now and it has been acknowledged and tentatively used in archaeology for more than twenty years. Only recently, however, has it begun to exert a real impact on our profession. Slowly, but inadvertently, the situation has changed from one, where the computer was an optional tool for the few, to one where it is the obligatory tool for everyone. At the same time the computer is changing from being a monstrosity, difficult to access and difficult to use, to an all round, personal tool with seemingly infinite powers and prospects. Today, nobody in their right mind will dispute that the computer is going to influence the way we work with archaeology. Yet, how and on what levels it influences us right now, or is going to influence us in the future, is probably far from clear to most of us.

I suppose that everybody has a notion of the impact on our work procedures – a change in techniques if you wish – and many probably share the false notion of rationalisation and saving of time. Many are probably also aware that new ways of dealing with problem solving in archaeology – new methods – will be seen. Methods that were utopian without a computer becomes obvious choices. Few, however, have probably considered that the advent of the computer influences the theoretical level in archaeology as well. This has little to do with the computer as such, but a lot to do with the enormous theoretical work on data structures and process handling that underlies software development. When we apply software as a solution to our problems we accept the theoretical background of the functionality that the software possesses.

One of the most important theoretical issues in archaeology is the nature and structure of archaeological data. Depending on how we perceive archaeological data, and what structuring principles we use for its organisation, we end up with different views on archaeological research, and its objective–past cultural systems. The nature of data and their structure is a key problem to any discipline that bases itself on observations of the real world.

Despite the fact that data has always been a focal point in archaeology, and has drawn much theoretical attention, archaeologists have seldom, if ever, been noted for their theoretical contributions to the understanding of observational data from the real world. Computer scientists on the other hand have been in a key position where a considerable part of their efforts have been centred upon attempts to build theories of data structures that will allow for more realistic representations on the computer of the real world. This has resulted in entirely new ways of structuring and describing data compared to what was available 20 or 30 years ago.

Traditional hierarchical organisation

In principle it is generally accepted that culture and indeed archaeological data used as an information base to former cultures are multidimensional (e.g. Binford's famous claim "that culture is neither simple nor additive" 1968: 24), and that polythetic rather than monothetic principles should be used for their structuring (Clarke 1968: 35-38). In practice, however, multidimentionality and polythetic structures are impossible concepts to realize on a piece of paper, when it comes to the description of actual data.

The traditional way to describe data is through the use of two-way tables, in which instances of an entity are described through attributes. Every line in the table holds one instance, and every column

represents an attribute of the instance. Despite the ideas of multidimensionality, however, each row in the table represents a one- dimensional description.

If two-way tables are going to be used for detailed descriptions of complex material, their size in terms of numbers of columns will grow almost endlessly. The solution to this problem of boundless descriptions in two-way tables has traditionally been the application of hierarchical trees. The hierarchies impose a predefined structure to the items described, vastly reducing the number of possible co-ordinated combinations in the descriptions, and placing most of the actual description in minor tables at the leaves of the hierarchy. The reduction occurs because at any child level in the hierarchy there can be a number of coordinated records that relate to the same parent record. By allowing for one-to-many relations down the hierarchy the size of the recording scheme can be reduced drastically. A hierarchical tree produces a very clear representation of the described reality, which is easy to grasp. However, it does not overcome the problem of one-dimensional descriptions of data. Any hierarchical tree structure may, in theory at least, be unfolded to a simple two-way table. I have once tried to unfold a hierarchical description for pottery decoration, in order to make statistical analyses in SPSS. I ended up with more than 500 columns.

Hierarchically organised descriptions are almost universal in archaeology, and despite the generally accepted adherence to the multivariate reality concept, it is often seen that archaeologist tend to think of hierarchical organization as being a natural characteristic of archaeological data, and a true way of describing them (Carver 1985: 50). Even in standard literature on database systems within computer science, the use of hierarchical systems are explained with reference to hierarchies as a natural characteristic of the physical and natural world (Elsmasri & Navathe 1989: 253)

This is also obvious when we turn to another area where hierarchies play an important role the definition of classifications. Here again, a classification based on an hierarchical organization is easy to grasp, and easy to apply to new material. But still it is a one-dimensional instance of a multivariate reality and it is, indeed, the realization of the monothetic type concept. Two items – far removed from each other on the hierarchical tree – may be more alike than any other two items in the classification. Hierarchical classification is as far removed from the polythetic type concept as can be.

Although we accept that cultures are multidimensional and that our types should be polythetic, archaeology nevertheless continues to describe data one-dimensionally and form monothetic types. Obviously, the problem is that there has been no way in which a multidimensional description or a polythetic type definition could be implemented. Descriptions in flat tables, even if they are organised hierarchically, are one-dimensional, and the traditional hierarchical classification systems are certainly monothetic in nature.

The relational model of data structures

The initial work in computer science concerned with database systems was concentrated on hierarchical organized systems. But as the limitations of these systems in relation to the real world became increasingly clear, computer scientists began to work with other models, first the network model and then the relational model. Today the relational model is a *de facto* standard for professional data base systems. To archaeo- logists, however, it is a new concept and to many probably just a word, not a concept.

There is no doubt that the relational model will become a key issue for archaeologists as well and eventually substitute the hierarchical model. The reason for this is quite simple: the relational model can do what the hierarchical cannot – describe a multivariate reality. The relational model works by way of a kind of divide and conquer principle. The reality to be described is divided into a number of logical entities, which individually are described in two-way tables. These tables are placed in relation to each other by way of other tables that describe what goes with what. As these cross reference tables may define many-to-many relations between the entities, and as tables that relate

entities may be set up in such a way that entities are completely cross-linked, a truly multidimensional structure is created.

To view the information on a piece of paper in even a simple relational data structure form is almost impossible because so much of the information is just cross references. It is only through the computer media that such a complex web of related informations can be controlled and accessed. When the content of a relational database needs to be presented to the user, it has to be done through a query. Through the query the information in the tables is combined and joined into a new two-way table with data that represent a limited one-dimensional view of the data held in the database.

Basically, the data that is retrieved from a relational data structure is just as one-dimensional as those retrieved from a hierarchical tree structure. The major difference, however, is that the one-dimensional structure of data in the hierarchical tree is laid down when the structure of the tree is defined, whereas ideally the one-dimensional structure of data from the relational database is created at the moment of query only.

The whole idea of keeping data in the computer in what appears to the eye to be an inscrutable structure, and then have it presented in various one-dimensional views on request, is new to archaeologists and therefore the relational database concept is difficult for most to understand. It is much easier to understand the concepts of the hierarchical representation and it will take time before relational organisation of data is the rule more than the exception in archaeology.

So far we have accepted the claim that the reality is multidimensional, but we have not learnt to describe it in a proper multidimensional way. Our association with the computer is currently teaching us how we can describe and store things multidimensional, but still we have to accept that we can only access the stored descriptions in one-dimensional views. The great challenge for us is to learn to use the computer as a means of constantly viewing the described reality through a multitude of one-dimensional views conditioned by our queries and not just through one *a priori* defined view, which we unconditionally accept as representing reality.

This challenge has a second important bearing, namely, the polythetic concept in connection with classifications. Archaeologists are accustomed to view a classification as a logical result of an analytical procedure applied to a description of reality. They are not accustomed to see it as a one-dimensional view conditioned by a sequence of choices that in principle are arbitrary. The computer may accomplish a change in this area as well. Given the high speed with which it is able to establish a view of a set of data based on rules laid down for the viewer, it encourages the user to experiment with different views of the data and helps him to gain insight into the structure of the data instead of forcing a structure upon the data in the first place.

In this area the computer may help us to change from a practice where order, most often in terms of rigid hierarchical classifications, is imposed upon the data and considered as the result, to a practice where no finite order of the data is sought, but where an understanding of the data is obtained through the application of a multitude of views on the data.

The object oriented model of data structures

The relational data model is currently the one of interest to archaeology, but it is not the model that currently is in the minds of many computer scientists today. Here we find the object-oriented model, which in some aspects constitutes a complete theoretical break with all previous models.

Since the sixties archaeologists have systematically been taught that the precondition for a description of data is a systematic decomposition of the data into its constituent elements (Gardin 1967). By the so-called attribute identification (Malmer 1963), the basic elements, on which the description must be based, is defined, and only then is the description made in terms of these elements. Likewise, classifications are based on combinations of the atomic elements defined through the attribute identification. The splitting of data into basic elements is a prerequisite for current data-structure models (the hierarchical, the network and, not least, the relational model).

They all work from the assumption that reality is best described in a bottom-up fashion, where tiny basic units of information are first defined, and then used as building blocks to describe reality in a structured way.

The object-oriented model turns everything upside down. In principle the object oriented approach starts at the top by defining what is common to all objects within the universe described. Then, gradually, a specialisation of objects is obtained by adding properties to these in various development lines, where each new level inherits the properties of all levels above it. In some systems the approach is limited to hierarchical descent trees, but systems where inheritance can occur from more than one line of parents are also in use, giving a network in terms of a directed graph and, indeed, creating obvious structural similarities to the real world.

In some systems it is possible to work in a bottom-up fashion as well, where one starts out with specialized objects and proceeds to create more generalized objects. This approach makes the difference from the relational model less obvious, but there are other properties of the object-oriented model which separate it from the relational model. The most important of these is that the data described is not just reduced to dumb numbers or strings in tables. Properties and meaning can be stored as a part of the data itself, which means that when the content of the database is processed the data handled will "know" which qualities it posses.

I have little doubt that the object-oriented approach from computer science will eventually exert an enormous influence on the theoretical thinking in archaeology. It does, however, demand considerable rethinking of the way that we should process data in archaeology. A rethinking that turns us away from the idea that data are best described and handled as a set of atomic attributes in two-way tables.

The archaeological research process

The advent of the computer is also about to change the archaeological research process. The nature of this process is dependent on how we individually adhere to different notions of scientific approaches, and it is not possible to point to a standardised way in which the research process is carried out. However, most models have one thing in common. They all have a loop from the level of interpretation back to the basic level of observation of the archaeological record (as an example one may take Clarke 1968: 36). This loop is logical of course, but it has little to do with reality.

The observation of the archaeological record is a slow piecemeal affair, continuously adding new information to our frame of reference. This information is described and categorised by various persons and, occasionally, used by them to present hypotheses and interpretations. But seldom, if ever, do the same persons go back and generate completely new material from the archaeological record in order to check and test their former interpretations. Rather, other persons using much the same material, perhaps with some extensions, and most likely described in a different way, will discuss other peoples interpretations and present their own. Only at a very abstract level can this loop in the research process be said to exist. Most of the time the research process is very much an act of balance in the mind of the researcher between what they know in terms of information about the archaeological record and what they imagine in terms of the structure and content of prehistoric societies. Thus the loops that exist are those going on in the minds of people between perceived information and stated interpretation.

The computer, as everyone's desk top tool, will help us to place these inferential loops on a more formal level. Our descriptions of data held on a computer may be queried, and the results of the queries may be used as a basis for interpretations, which may then be confronted with differently stated queries, and so on. It is in principle not different from what has been going on all the time in the minds of people, but the possibility of making the process formal, and of introducing formal analysis such as multivariate methods into the processing of the query results, opens up entirely new perspectives.

Conclusion

It is indeed difficult to separate the levels of methods and theory because no method is a-theoretical. By accepting and using a method some part of the theoretical thinking behind the methods is also accepted. Some may claim that the computer is just handing us new techniques and methods, but I believe there is more to it than that. I hope that this paper has been convincing in its claim: The advent of the computer in archaeology is going to change not only our daily work procedures, but also more basic issues. As we take up new and obviously superior methods offered through the computer, we are bound to have our theoretical views adjusted in accordance with the theoretical framework behind the methods.

References

BINFORD, L.R. 1968: Archaeological Perspectives. In: S.R. Binford and L.R. Binford (eds.) NEW PERSPECTIVES IN ARCHAEOLOGY, 5-32. Chicago.

CLARKE, D.L. 1968: ANALYTICAL ARCHAEOLOGY. London.

CARVER, M.O.H. 1985: The friendly user. In: Cooper and J. Richards (eds.) CURRENT ISSUES IN ARCHAEOLOGICAL COMPUTING. BAR International Series 271, 47-62

ELMASRI, R. - NAVATHE, S.B. 1989: FUNDAMENTALS OF DATABASE SYSTEMS. Redwood City.

GARDIN, J.-C. 1967: Methods for the descriptive analysis of Archaeological Material. AMERICAN ANTIQUITY 32(1), 13-30

MALMER, M.P. 1963: METODPROBLEM INOM JÄRNALDERENS KONSTHISTORIA. Acta Archaeologica Lundensia. Series in 8° No 3. Lund.

MIDDLE-RANGE THEORIES AND ARTIFICIAL INTELLIGENCE

Juan A. Barceló

I

Archaeologists usually say they need to discover the "meaning" of archaeological artifacts and ecofacts. They have also developed an enormous set of techniques to be able to obtain such "discoveries". In this context, "meaning" looks like a necessary category of real entities, therefore the task of scientists will be to find out that hidden characteristic.

I propose a different definition for "meaning": "the uses of the artifact". Given the fact that those uses may vary, we will conclude that "meaning" is not an intrinsic property of any archaeological artifact. A single object can be used in many ways, depending on the context or their users' needs. Archaeological artifacts have not the same meaning in all circumstances, because there is not a single way of using it. It is not the object that chooses its utilization, but users according to contexts.

From this discussion we must deduce that any interpretation of archaeological remains is a theory, even though it makes no reference to theoretical issues (Figure 1). Archaeological Meanings are cognitive constructs, the result of some inference mechanisms. The meaning of the archaeological record cannot consequently be "discovered", but reasoned or calculated by archaeologists, and to achieve this reasoning process, archaeologists need "knowledge". We usually obtain this kind of knowledge through actualistic or ethnoarchaeological research. In this paper, the "production" of such knowledge does not interest us, but it's use to solve a specific problem does.