Torsten Madsen

With Jens Andresen and Mads Holst

Centre for Archaeological Data Theory and Methodology

Excerpt from proposal to the Danish National Research Foundation in 2001for a centre for Archaeological Data Theory and Methodology. The excerpt includes only the presentation of the basic ideas.

Centre for Archaeological Data Theory and Methodology

The advent and development of computers during the last decades have caused profound changes with consequences far beyond the computer sciences. It has opened up for completely new ways of representing and analysing data, and directly and indirectly has inspired the development of new theoretical directions within the social sciences and the humanities. For instance, Agency models in cultural theory seems to be inspired from object orientation in computer science, and indeed resemble the newer agent based computational models.

Within archaeology, the development of cultural theory has followed the general trend in the humanities, but its concept of data has not changed to any notable degree. Still the view dominates that data are distinct "building blocks" with meaning build into them in the past, a meaning that may be uncovered through description and classification. This has led to a growing gap between on the one hand abstract models relating to empiric data and theoretical models relating to prehistoric societies on the other, even though the inherent duality of data and theory calls for a close correlation (for further discussion see the section "Theoretical considerations" below). If we cannot bridge this gap, we will end up with two archaeologies: theoretical archaeology based on data myths and field archaeology based on administrative practice and hardly anything else.

It is important to understand that empirical data from archaeological excavations by themselves are historical documents not relating to the past, but to the excavation, and that they cannot be reproduced, because the act of observation – the excavation – destroys their source. Hence, data in archaeology attain a right of their own raising questions of how to structure and preserve them for the future. Without this perspective, the goals of this research proposal cannot be fully understood. Further, it should be stressed that post-excavation analyses of data uncover patterns, structures, correlations, etc. leading to new data that must be added to what have already been recorded, available for further analyses. Archaeological data are thus a dynamic phenomenon, where the results of one set of analyses become part of the data for another set of analyses. It should also be understood that the storage and maintenance of an archive of archaeological data accumulated through time is not just a "stamp collection". As data accumulate, new patterns are created that may be uncovered through new analyses.

What we need is a development of data theory: The perception of data; the modelling of data for representation and storage; the manipulation of data as part of the research process. Archaeological data theory has two important preconditions. One is cultural theory that defines our frame of reference and determines the way we perceive data and their role. The other is the technological level of our information handling capabilities that sets the limits for how we can represent and store our data and how we can manipulate them.

The overall purpose of setting up Centre for Archaeological Data Theory and Methodology is thus to develop data theory adapted to current cultural theory and based on the achievements of computer science. Further, to implement the structures designed, and in connection with archaeological case studies to develop an appropriate methodology. More specifically, we will address the following four themes:

1. *Development and implementation of a meta-structure for data storage and preservation based on object-oriented principles*. The meta-structure must be flexible (able to handle widely different recording structures), relational (in principle everything can relate to everything else), dynamic (can adapt to changes in data structures as they are introduced), and have a memory (it should be possible to reconstruct the state of recordings at any point back in time). The aim of the meta-structure is to provide a sound basis for data recording in connection with excavations and post-excavation studies and for the storage of these data for future use (for further discussion see the section "Towards an object-oriented meta-structure for

archaeological data" below). From a heritage preservation point of view, this is a crucial issue (for further discussion se the section "Danish archaeology and the Information Age" below).

Based on our current experiences we will be able to start right away with a working version of such a meta-structure, and further develop this to its final state within the first year of the project. This meta-structure will become the foothold for the archaeological projects.

2. *Accomplishment of archaeological case studies in close association with the development of meta-structure and methodology*. It is imperative that there is a close interchange between theory and practice. The actual use of the meta-structure will be its test, and the use will provide guidance for changes needed. The development of methods is primarily a matter of actual needs in particular types of archaeological studies, but at the same time it is a question of bridging the meta-structure with these needs. It should be noted that computational potentials and limitations affects the formulation of the methods, thus forming a dialectic process between the archaeological component and the computer component.

Archaeological research is carried out at many levels, and in many contexts: There is the excavation producing primary empirical data; The post-excavation analyses, searching for patterns among the observations of contexts and structures, and the finds brought from the excavations; Studies of form and stylistic composition in artefact materials; Studies of artefact type compositions in various types of structures; Studies of structures and artefacts across sites within a region; Studies of sites in the landscape.

The case studies may be delimited to one of the above aspects, or it may reach across several aspects. It is not possible to determine the actual archaeological content of the individual studies in advance. It will fully depend on the archaeological interests and backgrounds of the persons involved in the studies. We plan to offer two PhD and four two-year post-doc scholarships for these studies. We will provide a detailed guideline outlining our demands that the studies must be based on data stored in the object oriented meta-structure and be focused on relational information, and our expectations for use of methodology, but the applications will decide what the archaeological content will be. As both PhD-students and post-doc scholars will come with limited knowledge of the structuring principles of data and the type of methodology that they will have to work with, they will be given education and training at the beginning of their work, and throughout as needed.

We exclude one particular type of investigation from the scholarships, and that is excavations, primarily due to cost of these. Instead, we establish a formal cooperation with Moesgård Museum, where selected excavations becomes integrated with the project. The projects then provides the recording system for the excavations and scholarships for post excavation analysis.

We do not restrict the archaeological studies *a priori* to deal with specific periods or specific regional areas. What matters is the complexity of the problem addressed. We wish to focus on subjects that present archaeological problems due to the complexity of the underlying data. Subjects such as composition in form and decoration on archaeological objects, complex time and space structures on archaeological sites, conflicts and ambiguities between logical and geometrical information on accumulated sites, problems relating to data on various scales of spatial and/or temporal resolution, etc. Post-excavation analysis is another major area in the archaeological research process with its own complex set of problems that we will address.

3. *Development and implementation of analytical methods to handle data described and stored using the meta-structure*. To make the meta-structure operational for archaeological studies it is necessary to write interfaces and applications to handle data stored in the structure. Thus, it is essential to the project to have programmers available. Especially in the beginning, there will be a pressure for programming expertise. We will try to make programming as modular

as possible, and try to develop user interfaces that as far as possible can adapt dynamically to different recording systems stored in the database. Still there will be a need for individual programming in relation to different archaeological studies. As the project progress there will be less programming needed to bridge meta-structure and archaeological studies, but at the same time the creation of the full object–oriented solution will demand more programming time in this part of the project.

4. *Development towards a true object oriented, agent based structure and methodology for archaeological research*. The meta-structure that will form the foundation of the archaeological studies and the methodological development is in a long-term perspective a middle step only. A true object orientation with full support for abstract data types, encapsulation of methods, and software agents is the goal, but there is no way that we will be able to achieve this from the outset. We intend, however, to turn the development in this direction.

   We see this development work as a gradual process over three years, where initially we will explore the possibilities and limitations of what can be done within existing software environments. Next, we will start experimenting and make partial solutions. Finally, in the last 1-2 years of the project we will try to establish a working prototype (for further discussion see the section "Towards an object-oriented meta-structure for archaeological data" below). The long development phase should be seen in relation to the simultaneous need to develop applications for the archaeological studies in relation to the meta-structure. During the development phase we need contributions and advice from computer scientists. We already have the following personal contacts in Europe:
   Nick Ryan who teaches Databases at the University of Kent in England, and for many years has worked with archaeology.
   César González Pérez, formerly at the University of Santiago de Compostella in Spain, but now privately developing software solutions with an object-oriented background for heritage organisations.
   Professor Jim Doran, University of Essex, who has a keen interest in archaeology. His research area is theoretical foundations of agenthood seen from a space-time perspective and agent-based social simulation.
   We will also seek contact with the department of computer science at University of Aarhus, renown for its work with object-orientation and complex data structures.

The following three sections are to be seen as extension to issues of central importance to the content of the proposal.

Theoretical considerations

Major changes have occurred in cultural theory within archaeology over the last 10-20 years, following the same trend as in most other disciplines. Generally, a deconstruction of the idea that general rules govern the trajectory of culture has taken place. This has consequences not only to our view of how human culture works, but also to our view of the research process in archaeology and the nature of archaeological data.

A recent, more constructive development is the notion of Agency. Although still not a well-defined theoretical approach, it holds some promising prospects. Very briefly: Agency sees the world as inhabited by agents, being individuals or groups of individuals of different size. Each agent is constrained by structural conditions (distribution of material resources, available technologies, systems of symbolic order), and they are in command of some structuring principles. The latter being knowledge based methods to manipulate the structural conditions. Agents act to

apply their methods, but at the same time they impress themselves on other agents through their actions, and they receive impressions from the actions of other agents – they "send and receive messages" that influence the structuring principles. Action therefore expresses the social context, the identity and the capabilities of the agent.

The Agency model defines human societies as complex, dynamic, self-organizing and self-reproductive systems. There are no teleological causes, no covering laws, and no "outside control" governing their function. The agency model is clearly part of a general trend in the humanities and social sciences today. We need just mention dynamic semiotic systems within the linguistics as an example.

If we look at the agency model from an archaeological point of view, the remains on which we base our knowledge of the past are leftovers from actions carried out by agents applying structuring principles to structural conditions. We find all elements of the model embedded in the archaeological remains, not as meaning, but as structure.

Traditionally we view archaeological data as bits of information each telling their part of the story of the past, and if we continuously unearth them, we will progressively build up a picture of the past. It is the "big puzzle" idea, as it clearly emanates from famous books like Childes "Piecing together the past". The current movements in cultural theory, as indeed expressed through the agency model, see data differently. The bits of data do not posses an inherent meaning, but taken together they posses an inherent structure owing to sequences of actions in the past.

This changed attitude to the nature of data has recently led to attempts to change the practice of archaeology. The current meticulous recording of archaeological excavations is futile because we only record the bits, not their structure, the claim is. Further, to record structure using formalised recording systems is not possible because the computer is too rigid a machine to handle the complex delicate interrelationships representing the structures of archaeological data. Only the human mind can hold and oversee the complexity. Therefore excavators should build up the structure of what they are unearthing in their minds and transform it directly into a narrative text describing their "feel" of what has happened in the past.

We do not share this view. First of all it should be stressed that archaeology has always been concerned with structural and contextual information, but given the nature of its paper based recording system, has never been able to handle the complexity of the information in a formalised recording system. Beyond simple context recordings, structural information has therefore never played a major role. With the growing theoretical awareness of the importance of complex structural information in archaeological data, a disparity has arisen. Archaeology has not managed to alter its recording systems to reflect the new theoretical views, but has kept on with the traditional recording structures even if computers are in full use. This has led to the fallacious stance that formal recording is of no use, because it cannot cope with complexity. This trend is ironic, because if we look at the general background of models of complex self-organising systems, then developments in computer science have played an important role. Archaeology has caught on to the new ideas, but has not fully understood their background.

Danish archaeology and the Information Age

Danish archaeology has a long and deeply rooted tradition for research of high quality. Together with an extremely rich archaeological record and a tradition for dissemination of results and for contact to the public, this has earned it a high status within the European archaeological community. The good public relations have also resulted in a large number of archaeological institutions. No less than fifty museums across the country have archaeological obligations, which has lead to a much more decentralised structure than anywhere else in Europe.

Danish archaeology is currently in a volatile situation with potentially profound changes due to the introduction of modern information technology. This applies not least to data management. Due to the nature of archaeological data, as presented in the introduction, they fall under the laws of heritage protection. Therefore, the museums carefully store paper files with data from archaeological excavations. To access this information you have to visit the museum and read the files.

Information technology has turned our attention to the possibilities of network-based access to data. We are getting used to the idea that whatever information we need is only a few keystrokes away. It is politically a hot issue, and there is a growing demand for making heritage data available online. Further, the reliance on digital data has introduced a completely new problem - how do we secure digital information for the future, physically and especially with respect to readability. In both cases arguments tends to focus on standardisation and centralised management.

To archaeology, standardisation traditionally means standardisation of content. If only we all used the exact same description systems and classifications, then securing, sharing and comparing data would be easy, is the argument. The problem, however, is that description and classification is part of the research process, part of the way that researchers ascribe meaning to observations. If everybody were restricted to use the same description and classifications systems, research would fossilize.

Strong administrative interests, however, may dictate standards of content. A couple of years ago the Dutch ministry of culture decided that all excavations with a funding based on heritage legislation (as in Denmark almost 100%) must use a single authorised recording system with fixed description categories and classifications for reasons of compatibility, ease of control, and ease of information exchange. The ministry considered it a simple administrative decision, but from a research point of view, it was a catastrophe.

In Denmark, a reference group has been working with the problem of securing and sharing digital data over the last year. Due to the decentralised state of Danish archaeology, its recommendations on standardisation will probably be modest. The same situation seems to apply in Great Britain where a tradition for widespread autonomy is also predominant. Still, in both countries, you hear strong voices demanding standards of content, and if ever the administrative demands for standardisation become imperative, then with the current state of the art it will surely be a standard of content.

An alternative approach is to create standards of form, which is what you do with a meta-structure. Through the use of a meta-structure some of the key arguments for standardisation is met: It will be easy to secure data since there is only one well documented structure; It will be easy to share data, as everything is stored within a common structure; Comparisons of data will not be hampered by structural incompatibility, but only by differences in the way researchers have chosen to describe and classify data, and that is as it should be.

Using a meta-structure as the basis for recordings in general is thus an obvious solution to whatever administrative demands may be raised for data standards. The best insurance for the solution to have an impact is a close contact with and involvement of DKC (Det Kulturhistoriske Centralregister), who will be responsible for implementing any political/administrative decisions. A formal cooperation has already been agreed upon, and DKC will be directly involved in the planning of the meta-structure, and its use in connection with excavations and post-excavation analyses.

Towards an object-oriented meta-structure for archaeological data

Traditional recording of archaeological data is organised in lists and tables. The paper based media makes it difficult to record relational information and complex structure, as all cross-references are

merely bits of text. The introduction of the computer in practice has done little to alter this situation, but using relational databases of course is one way to create better structures for recording complex information. Surprisingly, perhaps, even solutions based on relational modelling cannot cope with the full complexity of data, nor meet our demands for flexibility.

From 1994-97, we worked on a project funded by the National Research Foundation for the Humanities to create a powerful, flexible recording system for archaeological excavations. We based ourselves on entity-relationship modelling, but as we progressed, and structural complexity grew, we experienced and realized the limits of this method when it comes to a flexible handling of complex data.

In recent years, we have experimented with designs on higher levels of abstraction within a relational database management system, and found the approach awarding. The strength of the object-oriented approach we have adopted is that what creates the structure of a particular recording system is by itself data entered into the database at user level. The database becomes a meta-structure for recording systems, which means firstly that you can design widely differing recording systems without changing the underlying data base structure and secondly that you can adjust any particular structure at any time, as long as the changes do not interfere with data already recorded.

In the process of changing our current experimental design into a productive meta-structure for archaeological recording, and creating and implementing methods to go with it, archaeological studies must be carried out as part of the development. We need a feedback from practical experiences with the meta-structure in order to ensure that it is in accordance with archaeological realities and we need the interaction with actual studies to formulate and implement the methods to go with the structure. We are certain that these studies will produce results of a novel nature to archaeology,

The novelty will primarily be manifest through a dominance of relational information in the studies. As stated earlier, archaeology is familiar with the use of contextual information, but has been unable to formally record relational information in a way where it was operational for further studies. With the object-oriented meta-structure, relational information will prevail and the methodology to be adopted will primarily fall within the realms of network analyses, which apart from a specific application to stratigraphic analyses is unknown to archaeology.

Our experiences tell us that a meta-structure inspired by object-oriented modelling will indeed be relational, flexible, and dynamic. The integration of meta-data and data ensures that no single bit is undocumented – and this is an achievement by itself. However, the relational model does not contain all the information needed for an integrated information-system. Additional qualifying elements have to be implemented in the form of rule-based application programs interacting with the meta-structure. Thus for each specific “view” of the data you have to provide an application program for their presentation and analysis.

The limitations of relational and object-relational databases are well known to computer scientists. In this context, one may stress difficulties with object syntheses (the aggregation of complex data objects stored in atomic data types across various relations), database navigation (retrieval languages supporting set-operations only), and the sheer notion of “dumb” data that does not know how to present or handle themselves (no class-definition, no encapsulation of objects, properties and methods).

In order to overcome these limitations vast efforts are invested in the development of truly object-oriented database management systems. These products have not entered the mainstream market yet, which means that the project will have to do investigations in the capabilities of rare, yet available systems. Contacts in our research network can definitely be extremely helpful here.

We are convinced that object-oriented databases will form the basis of the next generation of information system development and that the project gradually must direct its focus in this

direction. Besides overcoming limitations of the relational databases, object-oriented databases offer what is termed “generalization” or “class inheritance”, which means that it is possible to represent class-hierarchies of different levels of specification. Another important feature is the possibility of object-aggregation, which means that sets of objects may be lumped together in order to form objects of higher levels of complexity. Taken together these two features are of key importance because they formalise central concepts of archaeological thinking about data structures.

It may well be asked, why do we not aim at a full object oriented solution from the outset? If we did, we would not have the possibility to involve archaeological studies for the simple reason that we would be entering *terra incognita* on an experimental basis. For archaeological studies to be carried out we must have a firm basis on which the archaeologists can act. This basis we are certain to find in the meta-structure solution, and this alone will be a “revolutionary” step to archaeology, and one that we can make matter. Still, we wish to pursue the full object oriented solution, and will do so as a second step of development that will not be directly associated with specific archaeological studies.