

A Theory of Archaeological Knowledge Building by Using the Internet - The *DIASPORA* Project -

Igor Bogdanovic, Oriol Vicente and Juan A. Barceló

Quantitative Archaeology Lab
Universitat Autònoma de Barcelona
Spain
igor@ramudavala.com

Abstract. The Internet has opened a new arena for many social, cultural, educational, and scientific activities. If we want this new technology to be really useful, we must change the way we do archaeology today. The DIASPORA project seeks to construct an archaeological knowledge base, not as a single database but as an 'intelligent' web of problem definitions and existing knowledge related to archaeological problems. Inspiration for the project came from already existing E-mail discussion lists, the appearance of Internet publications, and our year of experience with a web symposia system. DIASPORA would provide 1) a way of exchanging hypotheses to arrive at an agreed problem definition, based on communication among scientists, 2) a network of distributed databases and procedures to select relevant records, and to build a local database from distributed data, and 3) a series of methods and tools to process these data and to help solve the archaeological questions, which have been collectively defined in the first step.

Keywords: information sharing, internet, distributed databases.

1 Beyond the World Wide Web

Browsing through the WWW is a way of losing our time, not for the general idea of the Internet, but because there is nothing to be searched, or, sometimes, we do not know what we are looking for. Archaeological Internet documents are dead monuments, without real interest beyond the advertising of some archaeological materials. Internet is generally understood as a colourful showroom, with no other goal but to present partial information on or interpretation of archaeological research. It could be argued that these pages are not built for archaeologists themselves, but for the public. If this is the excuse, then we should say that even this task of popularisation of archaeological results on the Internet is not done properly. On the other side, Internet is lived out of the process of archaeological investigation, as an interesting, but useless gadget. We have a 'blind window', pure decoration, which cannot be opened for a breath of fresh air, or to look through to get to know the world. We can see only the window itself.

As a consequence, the WWW is nowadays almost exclusively used for presenting information, which is found and read by a patient *passive user*. The intention of using the web actively, outside of the Internet, for the purpose of professional archaeological life, is represented by 'discussion' or 'mailing' lists, directed at a more or less open circle of professional archaeologists, and defined by fields of interest. But the shortcomings of these lists are obvious. Take for instance the Archtheory list. In one year, from February 1998 to February 1999, the predominant utilisation was a demand for articles and books! This fact hides an important aspect of archaeological reasoning: archaeologists require information in order to explain past remains, and this in-

formation exists in books and papers. Internet and the Web are only a way to look for printed texts!

2 Archaeological knowledge building should be a collective, interactive and distributed task

Social effects of using the Web remain to be studied in many aspects. The use of network possibilities in science, particularly in modern archaeology, is a perspective, which draws our attention now. We are convinced that archaeology should confront the need to take a new step in understanding its goals and concepts. And the WWW is not understood here as a new gadget of yuppie archaeologists, but as an element for a new philosophy of archaeological research.

For any navigator on a journey the greatest desire is to arrive at a safe port. On the Internet, this principle should be translated into the necessity of quickly finding information somewhere on the web. But, as we commented before, it doesn't always happen this way, especially if we look for some topic related to archaeology.

In general, when we are interested in archaeological information we should go to some on-line magazines, the better if it is not necessary to subscribe, and hope to find some article on our topic of interest. Or we can see what search engines can find for us on the web. But even then we cannot be sure to find what we are looking for. Sometimes we can find a page that contains the word we have specified, and it may even be interesting for us, but we may have lost half a morning trying to find something!

It is difficult to find a really interesting website with important contributions to archaeological interpretation.

There may be a lot of material about archaeology, but most of this is without real interest for research. The Internet is being used by archaeologists for advertising, and not for knowledge exchange. Archaeology on the Internet could be an excellent place for a *collective problem discussion*, but not one of the specialized email lists is used in this way.

Some years ago, Jean Claude Gardin explained that archaeology is what archaeologists do, and archaeologists excavate, and write and publish texts. Archaeological reasoning is in a way connected with writing. Archaeologists 'reason' (or explain data) when they write. And they use data when they write; therefore they read books and papers to be included as reference data or to test hypotheses. The Internet could provide an easily distributed text+image file or a method for looking up texts and images. Nobody uses the Internet to learn, or as a tool for explanation.

Archaeologists are not isolated individuals working like 17th century antiquarians, but they are scientists connected to the real world. Archaeological knowledge building should be a collective, interactive and distributed task. Text analogy is then a wrong basis for modelling archaeological reasoning. The structure of the World Wide Web is a better analogy of the way in which archaeologists *should* reason, because it offers a chance for a really interactive and distributed exchange, beyond the classic forms of information distribution such as symposia, conferences, publications, etc.

The Internet should not be a database or presentation system, but a technology for archaeological knowledge building. In the same way as we use text data for building explanations, we should use information exchange in many formats (qualitative and quantitative data, images, 3D models, etc.) for rethinking the way in which we are doing archaeology.

We think that the most important obstacle to the use of the Internet in archaeological knowledge building is archaeology itself. Our discipline has grown into a descriptive topic, interested only in good-looking and spectacular objects. Archaeologists are more interested in publishing their own excavated data than in explaining social dynamics. Archaeologists have a tendency to limit themselves to the publication of data, and to minor explanations (context, chronology) or their own data. If they do look for other data it is only for comparative purposes.

Perhaps the World Wide Web is the best vehicle for a post-modern archaeology only interested in narrative histories, and appearances, but if one really considers archaeology as a way of historical problem solving, then it is obvious that much more than excavated data is needed. We need Knowledge, and for the moment this knowledge exists in the form only of scattered printed material in books and journals. The Internet as it is now, is a means of knowledge visualisation and non-systematic accumulation, but not a means of knowledge building.

3 The D.I.A.S.P.O.R.A. project

3.1 Archaeological Web Symposia

Our project is a development of an earlier initiative by Igor Bogdanovic and Nenad Tasic, who tried to transform the usual e-mail discussion lists and Web presentations into a Web Symposia system. Papers were to be submitted, discussed and updated by archaeologists participating in a 'Virtual' Symposium. The results obtained in this way could be easily integrated into an information library distributed via the Internet, CD-ROM and multimedia support. In this system, 'papers' are not passive depositories of information, but distributed files interactively updated. In order to moderate updating, an International Editorial Board was nominated.

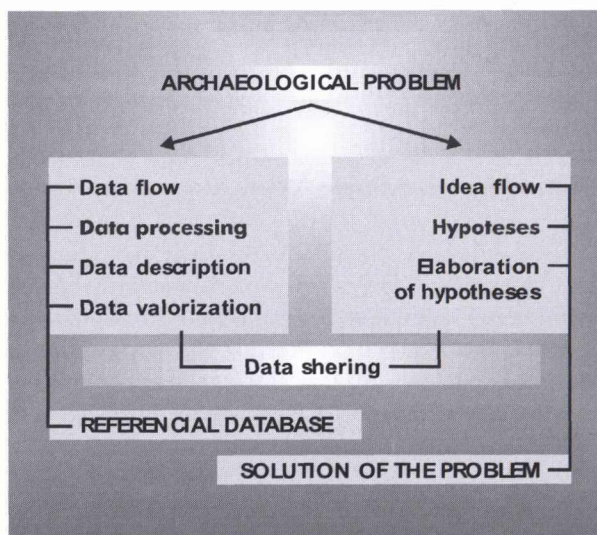


Fig. 1. Solving archaeological problems

Unfortunately, from the very beginning, the realisation of the project was hampered by various problems and difficulties, which resulted in a small number of registered participants. Therefore Web Symposia was 'put on ice' to await better times. In further developing the idea, we 'defrosted' this project, and combined it with an existing project of referential databases at the Universitat Autònoma de Barcelona, resulting in the DIASPORA project.

3.2 An Overview of the D.I.A.S.P.O.R.A. project

The DIASPORA project is a computer experiment that seeks to simulate the use of the Internet for a social construction of archaeological knowledge. It is not the usual series of nicely linked webpages, but a theoretical investigation into *how* to use the Internet in a dynamic and constructive way.

We are trying to integrate archaeological reasoning and data processing into a single system. We follow the problem-solving theory, and we have tried to create a computer architecture that reproduces this general framework of reasoning.

D.I.A.S.P.O.R.A. is

DISTRIBUTIVE
INTERACTIVE
ARCHAEOLOGY for
SYNCHRONIZED
PLATFORMS
OF
RESEARCH
ACTIVITIES

'Distributive Interactive Archaeology' means that archaeological knowledge building is a collective and dynamic series of tasks and processes. An individual archaeologist cannot explain his/her data because the explanatory process requires knowledge as raw material, and this knowledge as such does not exist in the individual mind of the scientist but it does exist as a global set in the research community. Explanation is not only a mechanical logical operation, but also a social process. Therefore, archaeological knowledge has to be distributive, which means that it should come from many different sources, and that it should also be interactive, as all scientists should be able to transform collective knowledge.

'Synchronized' means that knowledge produced and transformed by individual scientists should be made accessible to the rest of the research community in real time. Knowledge use is then synchronized in such a way that, what I am using and transforming here and now, will be used, and transformed by you there, now, and thereafter.

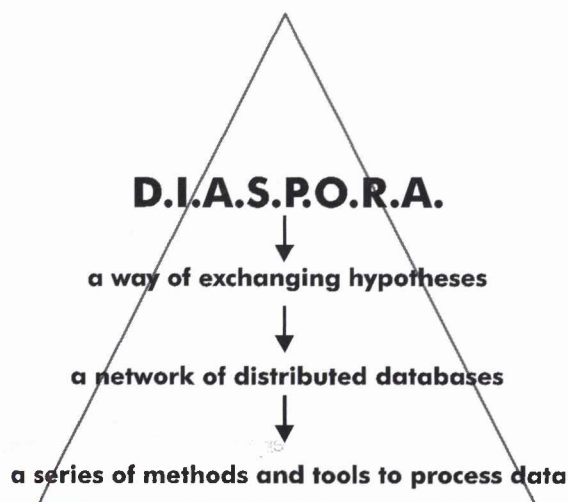


Fig. 2. A schematic view of DIASPORA

Platforms Of Research Activities' means that DIASPORA is something more and something less than a computer program. It is a set of assumptions and formal requirements for exchange of data, hypotheses and information. It is a set of linked databases, and explanatory rules. But it is not a tool that automates the process of explanation.

DIASPORA is then:

- a way of exchanging hypotheses to build collective archaeological problems based on communication among scientists
- a network of distributed databases
- a series of methods and tools to process those data and to help solve the archaeological questions, which have been collectively defined in the first step.

3.3 Diaspora as a platform for debate

A *Debating platform* is the creative potential of DIASPORA. It represents the creation and distribution of hypothesis, discussion, and data flow. *Debating platform* is an entrance door to DIASPORA, where archaeological questions are defined through debate. The main part of our system is *Academia* – a part of the debating platform, where *Web Symposia* take place, with a flow of hypotheses, as well as suggestions for methods for problem solving, and data relevant to the subject and possible conclusions. Several constantly open doors lead out of and back to this component. DIASPORA should contribute to formalise questions and to offer hypotheses and data, to organise the system of referential databases, and to build knowledge bases for the solution of a specific archaeological problem. In this way, a researcher contributes with his/her data and hypothesis, and has the entire corpus of distributed information at his/her disposal.

This method of research calls for a significant change in attitude towards archaeology. To begin with, DIASPORA supports a constructive opposing of thoughts, and not the more usual senseless attack and defence. The Debating platform is not limited in space, thus facilitating total coverage of the problem in both a geographical and a methodological sense. In order to really reach the knowledge base, it will be necessary to offer a systematic and complete presentation of available data, without hiding arguments known only to the author himself, which religiously happens among some archaeologists.

DIASPORA should link dispersed projects, synchronise its activities, and distribute all available information on the subject, overcoming the vanity of the researcher. We think that that the remark is not justified that this kind of communication tends to render classic expert meetings superfluous – 'in life', where one can eat and drink, and switch to standard publications which

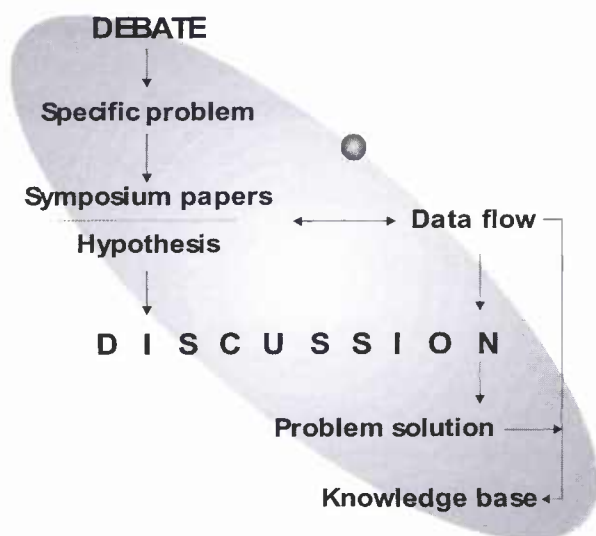


Fig. 3. DIASPORA in action

are so important for personal bibliographies and curricula. On the contrary, it will be more constructive to meet periodically with already discussed questions and a certain quantity of data available, and to present the conclusions to the problem, which are the results of teamwork. Standard publications will also keep their prestige, but this time at much higher level.

3.4 D.I.A.S.P.O.R.A. as a Knowledge base

DIASPORA is also a network of distributed knowledge bases. However, knowledge should be defined as Data in a Problem-Solving Framework. In other words, data that can be used to solve specific problems. We need something more than simple data to build a knowledge base.

In one sense, we could say that the system is a *data-base* containing *concepts*. Consequently, it is a structured set of facts, hypotheses and scientific laws. All this knowledge has to be built up by the interactive, synchronous and distributive participation of many archaeologists. However, the system should be something more than a large database full of scientific ideas; it should be able to manipulate all it contains, and to solve different problems by associating and transforming existing knowledge units.

DIASPORA has two main components:

- a body of explicit knowledge
- a set of actions to manage that body of knowledge.

Both components have to be defined in computational terms, that is to say:

- data structures
- procedural instructions

Although these are the classic components of relational databases, DIASPORA is much more than a single set of data. One would not use a relational database to 'solve' problems but to retrieve a specific data unit when the database contains a large number of data: the user *consults* information. To do such a task, query languages contain only constraints, a fixed number of features to specify which unit to retrieve. Instead, with DIASPORA we try to solve scientific problems, that is to say, we do not retrieve data units but we 'instantiate' a solution to the problem.

Of course, the World Wide Web is not an Expert System, although perhaps in the near future a network of linked Expert Systems will distribute the flow of archaeological data and hypotheses. We therefore suggest the idea of Reference Databases, in the same sense as it has been used in other disciplines. A Reference database is a structured set of data, which can be used directly for solving problems.

Part of DIASPORA will be therefore a series of Reference databases distributed in different sites.

A Reference Database contains normalised data, in such a way that it contains all possible solutions to a problem. For instance, in use-wear analysis, it will include all possible working activities and use-wear related to that activity. In Charcoal Analysis, for example, it can contain a description of how wood was gathered in specific environments and social conditions. The user is not looking for charcoal samples, but for 'instances' of the same concept: given the same environment, how differently was wood used in different comparable societies.

DIASPORA contains:

- a set of concepts describing some relevant scientific knowledge related to a scientific problem and its meaning;
- an adequate 'active' representation of these concepts and meanings to allow the reaction of concepts to messages sent by the user or by other components of the system;
- a set of rules which will manage the concept descriptions in terms of their representation;
- a set of operators for the representational language;
- three kinds of meta-knowledge:
 - knowledge of the problem to solve
 - knowledge of the structure of the system
 - knowledge of the strategy to solve the problem.

Differences between DIASPORA and standard databases are a consequence of the different computer representations used to build them. Databases contain many data, expressed in a very simple way, and relatively simple search procedures to retrieve some of these data.

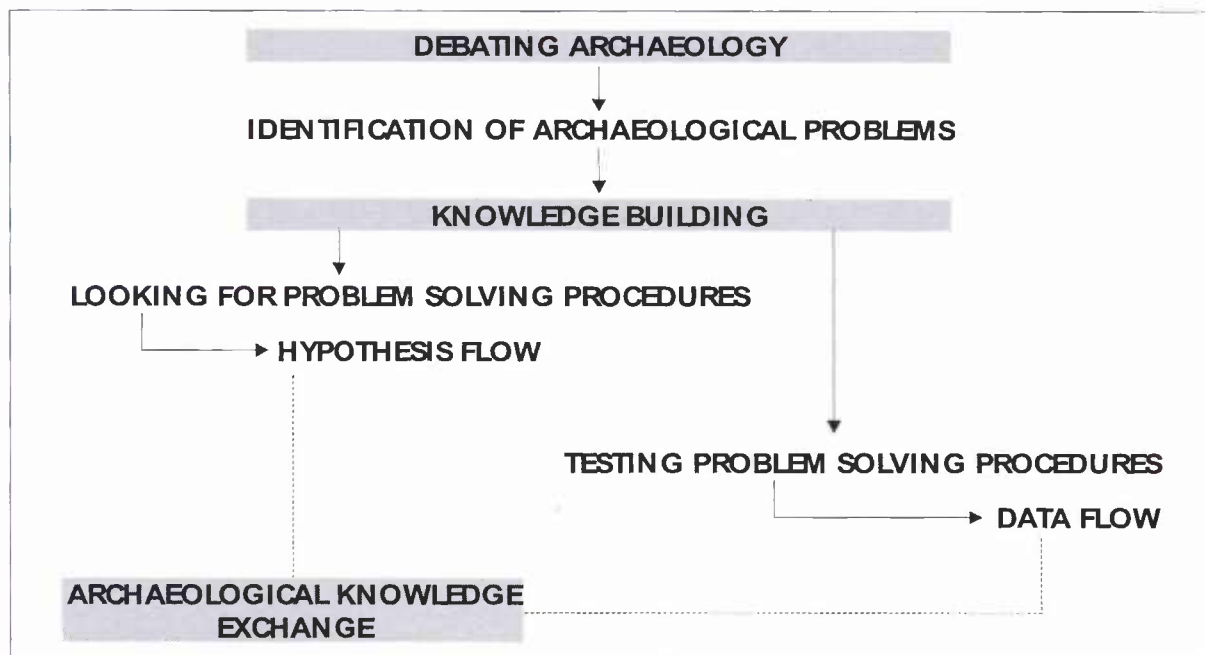


Fig. 4. DIASPORA diagram

In DIASPORA there are fewer data, but they are expressed by means of more complex data structures, because they represent scientific concepts. Query languages are substituted by problem-solving methods based in heuristic search, and using some contextual information (the meta-knowledge component) to reduce the amount of computer memory needed.

4 Conclusions

We cannot really speak about 'conclusions' when we are trying to comment on some theoretical points on the way to use the Internet and the World Wide Web for archaeological knowledge building. Machines are thinking more than humans. Technology is going ahead faster than we expected, and now we do not know how to deal with it.

Our thinking machine, the brain, has been developed to deal with a vanished world. If we do not change our way of analysing the world, if we persist with the good old archaeology of lost days, archaeology will become archaeological dust.