Stuart Eve – Guy Hunt

ARK: A Development Framework for Archaeological Recording

Abstract: Archaeological recording practices vary from site to site, throughout the world. Different systems are developed to address site-specific issues and to satisfy individual research agendas. In our opinion, this is to be encouraged and embraced – recording systems should be easily adaptable to ongoing or new issues, and data structures should not limit or even influence the range of questions that can be investigated or the resolution of data that can be recorded. This paper introduces the Archaeological Recording Kit (ARK), a web-based system built on commonly available open source software that is designed to be able to adapt to any digital or paper-based recording system. ARK includes data-editing, data-creation, data-viewing and data-sharing tools, all of which are delivered using a web-based front-end.

Introduction

This paper introduces the Archaeological Recording Kit (ARK), an open source project for archaeological recording and documentation. The original impetus behind the creation of ARK revolved around attempting to provide a flexible system that allowed us to put our site records into a database. As is the case with most projects, ARK developed over time and our aims and objectives have changed dramatically since the project's inception. We were naive when we first embarked on the project and had little concept of the complexity of creating such a system. We have undertaken a long journey from being archaeologists to being web-developers, with a great number of lessons learned along the way. This paper documents our journey and also provides some background to our theoretical standpoint on the nature of archaeological recording and describes the structure of ARK.

Previous Databases

We work as commercial archaeologists within the Cultural Resource Management (CRM) sector in the UK but are also involved with a number of university-run archaeological research projects. The use and nature of the databases we created varied greatly from one implementation to another and the system that we wanted to create needed to be able to adapt and scale to the kind of uses we have for it at L – P. We also hoped others would be able to invent uses for the system in areas outside archaeology. We felt dissatisfied with the database systems that we had used in the past and the systems that we ourselves had written, and whilst there was no concrete reason behind this feeling, we thought that something better was possible.

In creating archaeological recording systems we are also constrained by a number of factors such as the necessity to comply with museum archive standards, international data standards and our own legacy data compatibility, but in so far as is possible, we have tried to treat these issues as entirely secondary to the development of ARK and develop a system that responds in the first instance to archaeological needs.

One of the major problems for archaeological data managers is keeping the computer subservient to the practice of archaeology and not vice versa. The creation and use of databases often requires specialist training and the data structures are constrained by the necessity for referential integrity. In our work simple technical issues like choices about checkboxes or drop-down menus were gaining precedence over the more archaeological questions such as what should be recorded and how. In general, the field archaeologist is becoming separated from the creation of data, as the data structures are being designed by technicians, rather than field archaeologists themselves. This problem is compounded by the need to define early in a project what should be recorded for any given record. Because the database must be ready before the excavation begins, there is a necessity to second-guess what is important to record about the archaeology before a trowel has even hit the soil. Of course it is possible to change the database as the excavation progresses, but even quite minor changes can be time consuming and naturally restrictive. It is unlikely that the field archaeologists will want to contact the database technician every time they want to add a new criterion or field to the database. In summary, the database systems we had previously used were a nuisance at best and were actively limiting the imagination of the archaeologists working in the field at worst.

Hypertext

From the outset, ARK was to be delivered via a web-based interface. This allows it to be used on any computer system, from anywhere in the world, without the need for special software. This web-based approach led us to a much more detailed examination of "hypertext", both in terms of its original meaning and as a way of thinking about the creation of documentation, texts and media.

The World Wide Web is the best known implementation of a hypertextual system, although it lacks most of the features and functions of hypertext as it was first imagined. The term hypertext was coined by Ted Nelson in 1965, when he set out to express a new way of thinking about text and publishing (Nelson 1981). Nelson's aim was to break with the "paper paradigm" and to create entirely new forms of textual documents.

Docuverse

Nelson proposed a *docuverse* where all data is stored once and no data is ever typed twice. This forms the basis of the idea of *transclusion*. There is only one copy, which is the master copy of everything; every other copy that is viewed or distributed is a manifestation of this original. This view "repurposes" the entire computer system into a box that maintains the connections between all of the transitory and cached pieces of information, whose identities are maintained with their originals. The idea of this ***docuverse*** fits perfectly with our overall vision for ARK, both paradigmatically and practically.

Hyperlinks and Eternal Revision

Navigation through information in this ***docuverse*** would be non-linear, depending on each reader's choice of links. But more than this, the reader, by actively creating the links, would become a participant in the process thereby becoming an author in this creation of new narrative structure. Unlike in HTML, the links would be two-way, allowing the document to be read in any order at any time. The links themselves would provide a non-linear branching and looping of the narrative, which could be unique for every user of the document.

Much like in "wiki" technology today, there would be no deletions of data, just eternal revisions. These revisions ought to be transparent so that earlier edits and versions would be available to the reader for simultaneous side by side comparison.

In an important way, this idea of a hypertext and hypermedia has influenced us on both a practical and a theoretical level. In one way, we have tried to undertake some crude implementation of true hypertext and transclusion in ARK. But in many ways we are restricted to what we can do by the limits of currently available web protocols. We have also begun to think of true hypertext as a model for the way archaeological data and media could be created. We think of ARK less as a database system and more as a means of creating archaeological texts and media.

Post-Processual Theory

Those of us who work with data often bear the brunt of criticism from post-processual archaeologists. We have a tendency to work in a way that "scientificises" and formalises what is essentially interpreted data. Especially in the field of site recording systems, our tendency is toward abstraction and removal of the voice of the excavator. We cannot claim to have totally answered these criticisms, but in our work we are certainly aiming to target those areas in which we can have most success.

Multivocality, for example, is one of the key areas that ARK can approach. It should be possible to record different interpretations for the same data and groups of data. We realised that ARK ought to provide these tools for readers and users, which can in turn empower the reader to become the interpreter and to contribute to the project. Reflexivity is also something that ARK can easily address. The instantaneous nature of ARK and its web-based front-end means that people working on the project can interact with each others production on an almost immediate basis. ARK also makes it possible for individuals to group and interpret data in their own way, to present conflicting and differing interpretations of the same data.

ARK, therefore, is a response to our dissatisfaction with the status quo, influenced by hypertext, post-processual theory and many other factors that have been present throughout the last 5 years.

What is ARK?

In software terms ARK is a collection of PHP pages and a MySQL back-end with some Javascript / AJAX techniques used on certain pages. But more importantly, in conceptual terms ARK is a loose collection of tools, a development framework to allow the recording of any type of archaeological (and even non-archaeological) data, from a simple image catalogue to a complicated multi-relational, multi-site excavation database involving many specialist datasets.

Items and Fragments

ARK works by atomising the recorded data into the smallest possible units and then providing methods to agglomerate these into larger units for the creation of a narrative. In ARK the basic units of data are known as "items" and "fragments". An "item" is a record, the content of which the field archaeologist creates. Items can represent real-world phenomena, such as contexts, or conceptual constructs, such as bibliographic references. ARK can have as many of these items and as many different types of items as are needed. They can be linked directly to each other in a non-hierarchical way.

"Fragments" are bits of data that can be attached to items; again, there can be as many or as few as necessary. There are various ways to visualise the networks of fragments that are then created. Quite simply and intuitively, we like to imagine the fragments as bunches of "Post-It" notes that can be stuck to and removed from items as needed. Each fragment of data has a type, so, for instance, if we are recording a context and want to attach a short description of it, we can choose to attach a piece of text of type "short description" directly to that context's item record. We can then attach different fragment types to each item *(Fig. 1)*.

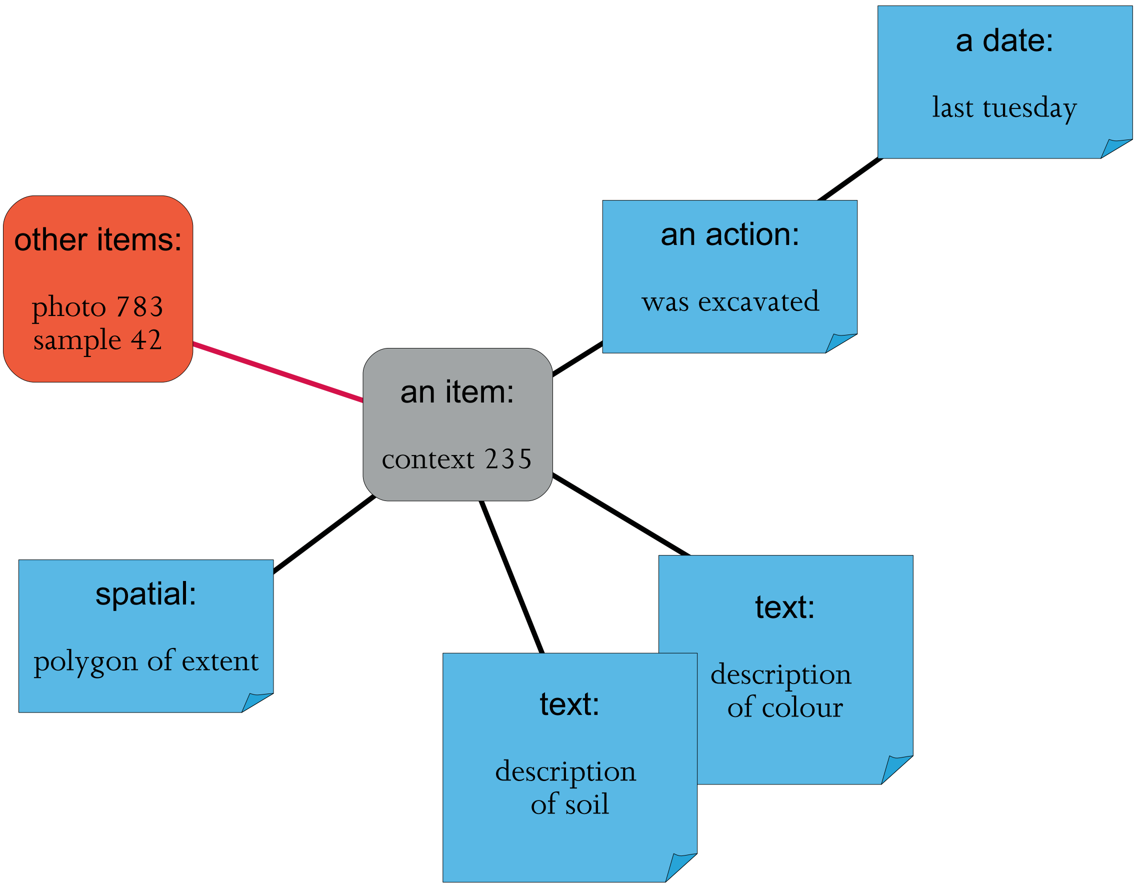

Fig. 1. Showing how fragments are attached to items.

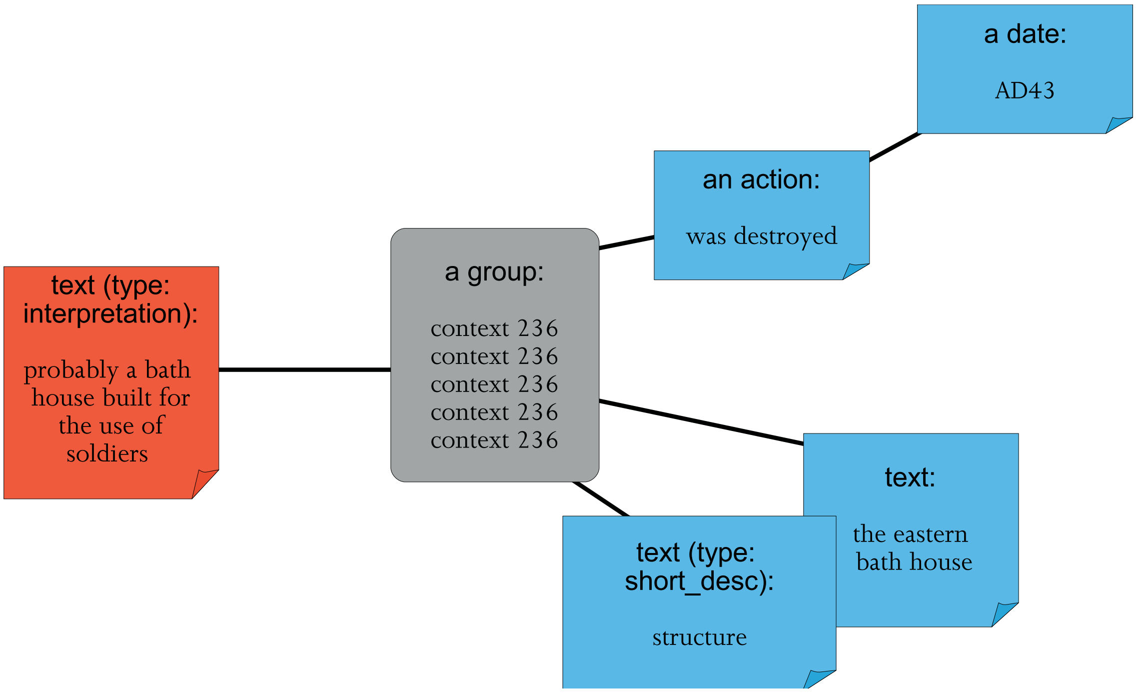

Fig. 2. Groups.

These fragments might include spatial geometry, event records, or dates and values from controlled lists. By allowing a fluid process of creating and attaching of data fragments to items, we are going a long way towards a structure that is once more in the hands of the excavator, whilst at the same time reducing database redundancy. The data recording system is not set up by ARK, but is configured uniquely for each project. Each instance of ARK, then, can be very different from the next.

Filters and Groups

As we start to build these units of data into bigger blocks, it becomes necessary to filter and group the records. In this way, the user begins to take the raw data and build his or her narrative. Filters can be made up from a variety of different search mechanisms: free-text search, "attribute" search and even a spatial search. Each filter is stacked on top of the previous one, filtering the data until the desired result set remains. This result set can then be viewed either in a tabular format, as text fragments or even plotted onto a dynamic map. Once the user has a configuration of satisfactory filters, they can be saved to the user's personal account or made global so that all users can reload the filter criteria. In this way, during an excavation users can create, save and re-run their own custom queries or borrow those of their collaborators on the actively changing dataset.

A result set can also be saved as a fixed snapshot or "group". Groups then become ARK "items" themselves, and can have other fragments attached to them. This is essential as the project begins to build narrative, since stratigraphic groups or other groups can be commented on as a unit.

Spatial Architecture

From the outset we considered that spatial data should be an integral part of the ARK architecture. Virtually every excavation now has some form of GIS or spatial data, even if this is just the site plan digitised into CAD. ARK uses a combination of open source software to present the spatial data. The University of Minnesota's Mapserver program (http://mapserver.gis.umn.edu/) is combined with OpenLayers (http://www.openlayers.org) to provide simple and clean access to the spatial data alongside other forms of data in the same display. This obviously has benefits in terms of ease of use and in terms of the integration of the display elements. However, the spatial architecture of ARK goes beyond producing attractive maps and making the GIS easier to use for non-GIS users.

Geometries are considered an integral class of data. Where records have a spatial component, this is always stored as such. If we take the example of a context, in all single context systems on paper and on computer, contexts have three dimensional location information. We also hold the spatial extents of other archaeological / spatial constructs on the site (for instance, trench extents, rooms or grid squares). From this base we can then use ARK's PHP functions to query what contexts are in any given trench or room, without ever specifically attaching that information to the record itself.

This kind of approach allows a much more fluid thinking about the areas of a site both as practical excavation constructs and also as archaeological ones. New spatial "areas" may be created during the course of the excavation or old areas may be renamed. The excavators themselves may not agree on the extents or even the existence of a particular area and ARK facilitates this kind of multivocal and reflexive approach. It can allow different users to present their interpretations to others.

The second major architectural idea is the ability to define map layers by themes. We can use any of the fragments attached to a record as the criteria for either display or symbology. In practical terms this means that the user can very easily create themed plans such as phase plans based on the values held in the database. On the architecture level, this means that we are handling the idea of these themes or layers in a new way, detaching the theme that the reader wishes to see from the geometry itself.

Transclusion

On a certain level ARK simulates *transclusion*. The limitations of HTML mean that true *transclusion* is impossible, but as a concept, *transclusion* is fundamental to ARK. As all records are viewed as non-hierarchical it is essential that elements of one record such as a photo are able to be viewed from within another such as a context. In a basic way, our model for this is a form of *transclusion*. This idea references Ted Nelson's idea that the links between fragments of hypertext would be as important as the data they reference.

As the archaeologist begins to aggregate individual primary records into ever larger units or hypertext, it becomes essential that he is able to make seamless reference to other elements of the hypertext, even quite complex elements such as phase plans. *Transclusion* enables an author to write about something and at the same time include the data being described within the new document. Narrative can be written alongside a live feed from the database, a current phase plan or a local stratigraphic matrix. Simple tools enable this kind of data to be served to systems outside ARK such as a wiki or a blog, which itself becomes part of the hypertext.

Conclusion

Throughout this paper we have intentionally focused on the theoretical underpinnings of the ARK system rather than on actual implementation. ARK is a continually developing project, and a number of projects are currently using it in the field. We will be making the source code freely available in the second half of 2007, and are hoping that the user community will take the project on and develop it in new and exciting ways.

References

NELSON 1981

T. NELSON, Literary Machines. The report on, and of, Project Xanadu concerning word processing, electronic publishing, hypertext, thinkertoys, tomorrow's intellectual revolution, and certain other topics, including knowledge, education and freedom (Sausalito 1981).

Stuart Eve
Guy Hunt

L – P Archaeology
The Truman Brewery
91 Brick Lane
London
E1 6QL
United Kingdom
stuarteve@lparchaeology.com
guy.hunt@lparchaeology.com