

Grand challenges: Grand opportunities? archaeology, the historic environment sector and the E-science programme

*William Kilbride, Glasgow Museums
Thursday, 01 June 2006*

0. About this document: Please read me!

This document is intended to stimulate debate and discussion in preparation of the E-science Scoping Study Expert Seminar for archaeology and the historic environment. It is a tentative, partial and therefore controversial delineation of the challenges and opportunities which the E-science agenda presents archaeologists and the historic environment sector. It proposes areas of specific activity that the sector may benefit from addressing. It also delineates a series of barriers – some of which are beyond the scope of archaeology to resolve.

It begins with five discussion points of what could be achieved using the tools and approaches of E-science.

1. Archaeology as Virtual Research Organisation
2. Archaeology as Spatial Data Infrastructure
3. Visualisation and simulation: interpretation and methodology
4. Resource discovery and delivery
5. Research *sans frontières*

This list is not intended to be exhaustive and participants are invited – indeed encouraged – to make new proposals or to propose that items are removed.

Participants are invited to consider an action plan of what we might need to do to achieve the goals outlined above. This might be done by noting and reviewing the ‘barriers’ and giving each one an ‘owner’. Some of these will end up being outside our sector, and others will be within the sector, and therefore in our power to resolve.

This document does not represent the views of ADS nor AHDS nor even Glasgow Museums. Moreover, because it is intended to provoke discussion it does not necessarily represent the views of the author. Consequently participants at the workshop are asked not to circulate this document: a more rounded discussion will form part of the reporting of the project, and participants will be sent a copy of this in due course.

1. Introduction: What is E-science?

One of the characteristics of E-science is a confusion on what can be ruled in and what can be ruled out. This is in part because the E-science community is replete with its own jargon. Moreover, the discursive attraction of association with E-science gives it the feel of a bandwagon. Consequently, numerous projects and teams are actively re-describing their processes and tools as E-science in the hope of sharing in the relatively large investment that government is making through its E-science programmes. This is

especially true in the humanities which has not yet seen much expenditure under the banner of E-science. The problem is exacerbated by the lack of a precise and comprehensive definition of E-science. The question then arises, can any IT application be proposed as an 'E-science solution'.

I have resorted to the familiar mechanisms of archaeological typology by proposing a set of criteria that characterise existing E-science projects. Based on the list provided before the seminar, E-science projects display at least one of the following characteristics:

- Collaborative but distributed interaction between scholars based on necessarily bespoke IT applications
- Collaboration that meets 'grand challenges' in a discipline
- Massive and distributed computing power
- Massive and distributed storage and transport of data
- Virtual access to shared scientific facilities
- Data mining and analysis
- Hypothetico-deductive reasoning

Crucially, none of these criteria are antithetical to archaeology. Some of our language may be different: hypothetico-deductive reasoning may be part of a broader interpretative approach; shared scientific facilities such as mass spectrometry and electron microscopes are often shared with colleagues in other disciplines; we may call our grand challenges 'research frameworks'. In short, the prospects for archaeology and the historic environment are strong. Uniquely in the arts and humanities, archaeology would be able to present research projects which matched *all* of these characteristics.

2. Archaeology as Virtual Research Organisation

Characteristically, E-science is about collaboration. Superficially, therefore, archaeology is well placed to contribute and benefit from the programme since almost all archaeological research projects are shared endeavours. Archaeology is smaller in terms of number of active researchers than many of the scientific disciplines that make collective decision-making and agenda-setting easier there is also a tendency to localisation of research interests.

Collaborative tools, often termed though loosely specified as Virtual Research Environments, have been used in a number of E-science projects to share activities. Such document sharing and personalisation tools have been adopted in many archaeological projects and in cases such as the Silchester project the site database has been explicitly re-described as a Virtual Research Environment underlining the collaborative nature of archaeological excavation to the E-science community. Work on portals-based tools are relevant here, especially the development of portlets for personalisation – such as implementation of WSRP and JSR286 – which are designed to allow remote content to be re-cut according to the needs of distributed but personalised working environments. Such work extends several generations of research and development tools to enhance archaeological field recording and post excavation analysis. There is no question that archaeology can benefit from adopting some of these tools, and by adopting the language of E-science make existing archaeological applications available to other researchers.

Over and above collaboration on specific research projects, arguably archaeology is closer than almost any other discipline to transforming itself into a complete virtual research organisation. By this I mean the whole discipline acts as persistent consumers and producers of information through a continuous collaborative cycle of analysis and interpretation. The roots for this transformation exist in the AIP and DES projects which gathered and published descriptive details of every archaeological intervention in England and Scotland respectively. OASIS has allowed both of these projects to emerge in digital form. Although development is painstaking, the use of the OASIS data pipeline means that contracting fieldworkers, local government curators, national heritage agencies and academic researchers are linked into a single virtual organisation that supplies and supports research outputs.

Though incomplete this transformation worthy of analysis by other disciplines: although it has only come to the table late, archaeology may well be the first complete disciplinary grouping to transform itself into a comprehensive virtual research organisation.

Barriers to using the OASIS pipeline to create a comprehensive virtual research organisation are as follows:

- Training needs and communication
- Intellectual property rights
- Need to extend beyond literature into other data formats
- Need to capture ‘second phase’ research
- Varied adoption of OASIS tools
- Need to digitise or acquire the grey literature backlog
- Dependence on development-control system
- Mixed quality of research outputs

3. Archaeology as spatial data infrastructure

One of the familiar metaphors in E-science is the idea that an entire network (or grid) of computing power acts as a single but infinitely extensible computer. Following this metaphor, if different parts of the grid behave as components of a single system, then the duplication of services becomes wasteful in more than simply financial terms. As well as reducing revenue costs de-duplication frees up processing power and storage and allows researchers to move to substantive questions more quickly. The geospatial data (GD) sector presents a useful analogy: in part because it is relatively more advanced in its preparedness for E-research, and in part because it is a sector with which archaeologists routinely engage.

Of late the concept of a ‘Spatial Data Infrastructure’ (SDI) has emerged to describe a global model of shared services, protocols and processes that are envisaged to facilitate interoperability in the geospatial data sector. The result – at least in theory – is a comprehensive, persistent and distributed set of tools that reduces the barriers to integration and use of diverse data sets. Research is thus assisted through the removal (or diminution) of repetitive and elaborate preparatory work. This strategic infrastructure is based on the deployment of common and relatively simple standards and also on a shared

understanding of roles and responsibilities. Moreover, because systems are able to default to mutually comprehensible spatial syntaxes, diverse and specialised data forms and processes can be reconciled. Arguably, SDI has many of the characteristics envisaged for the semantic web.

There are two lessons we can draw from the thinking behind spatial data infrastructures. Firstly, because it is inherently geographic, archaeology should expect to contribute to and draw from services configured for use in this way. Indeed, future developments in archaeology should be required to have these geographic standards embedded within them.

Barriers to archaeology developing as a contributor to a wider spatial data infrastructure are as follows:

- Training and communication
- Intellectual property rights
- Standards development and implementation

Perhaps more interesting is the opportunity for archaeology to model itself self consciously on the GD community towards the development of an *archaeological data infrastructure*. The GD community is a multi-billion pound, global industry in which commercial and public agencies compete for the massive profits and catastrophic losses which can be made. It is all the more remarkable therefore that the sector has emerged from two decades of expansion with such a strong sense of unity. This has in part been because the Open Geospatial Consortium – a membership organisation and forum for discussion – has allowed many of the issues to be resolved in an open and practical way.

How might an archaeological data infrastructure emerge? Using OGC as our model it is clear that organisations like the Forum for Information Standards in Heritage will be critical. However there is not the same sense in archaeology that information standards matter, fewer examples of the benefits that accrue from the use of standards and consequently nothing like the urgency or universality in debate as there historically has been in the OGC. In short, standards have to be much higher on the agenda.

Barriers to an archaeological data infrastructure:

- Engagement with standards development

4. Visualisation and Simulation: interpretation and methodology

Many E-science projects have used or developed visualisation and simulation tools as a basis to solve problems, especially in disciplines where visualisation and simulation are familiar methodologies. Uses include using E-science resources to visualise, analyse and share the outputs of CT scans, microscopes and other sensor arrays. In addition, grid computing power means that visualisation need no longer be a solitary experience: collaborative virtual environments allow groups of scholars to interact in real time with a shared visualisation.

Although computer-based reconstructions are now a popular cliché of archaeology, they tend to be educational or entertainment-related outputs. Research uses of 3d visualisation are piecemeal. Other forms of visualisation such as geophysics, spatial analysis and remote sensing may be less familiar to a public audience but much more recognisable in archaeological research. In the last decade, a new generation of affordable 3d survey equipment has emerged which allows for detailed recording and has principally been deployed to record upstanding monuments. Other devices such as CT scanners, ground penetrating radar and sonar survey have also become more familiar. Analytical simulation is rarer in archaeology though it has a longer pedigree, originating in cybernetic theory of the 1960s and continuing through intelligent systems and predictive modelling in the 1980s and 1990s. Crucial to all these technologies is the degree to which they contribute to the research process. Arguably many visualisation and simulation projects have in fact been technologies looking for a problem to solve.

The question is to what extent e-science tools and approaches should be deployed to enhance or extend these existing approaches, and what the impact may be.

Barriers to visualisation and simulation in archaeology:

- Preservation of and useful access to previous models
- Training and technology churn
- Risk of disconnection from the mainstream research
- Publication and recognition of visualisation outputs
- Entry costs can be high (geophysics vs immersive VR)

5. Resource discovery, delivery and processing

E-science relies on the ability to locate, deliver and process information so a number of core applications and tools have emerged that support or extend these functions for the E-science community. In effect E-science is neutral about the approaches used to discover and deliver resources. Nonetheless, the dependence on the ability to do so means that changes in technologies which support these functions are important.

Resource discovery in archaeology is largely based on the metadata model. In this model, a distinction is retained between the resource descriptor and the resource itself: the former is indexed and searched, the latter is delivered. This model works well for mixed analogue and digital resources and for data formats that cannot be indexed efficiently by computer. It is a familiar model from the library community and is the model used by 'metasearch' Internet search engines. But this is not the only model for resource discovery that exists nor is it necessarily the most suited to our needs. Full text keyword searching has supplanted metasearch as the dominant paradigm for Internet search engines and the boundaries between data and metadata are not clear-cut. Moreover the emergence of conceptual reference models (such as the CIDOC CRM) holds the promise of semantic interoperability between different information communities. Commentators are divided on whether such ontologies will be able to fulfil that promise.

A great deal of attention has been paid to resource discovery but relatively little attention has been paid to issues associated with the completion of the transaction after discovery. For example, persistent identification of resources remains problematic and such identifiers as exist tend not to operate below the level of filenames. Consequently it remains difficult to cite a data point within a file and be able to access it consistently. This is not a failing of resource discovery per se, but of fulfilment. Likewise accessing data points within a file normally means disclosure of capabilities via an independent schema and markup. Markup standards have received quite a lot of attention in archaeology, but tools remain largely under-deployed, making completion of transactions an ad hoc affair. These issues are by no means unique to archaeology, and it is not yet clear whether the E-science community has working solutions.

The relative immaturity of digital services in archaeology also constrains and hampers what we can deliver and process. Non-digital source material is still arguably more important across archaeological research than the entire digital corpus. It is unlikely that the E-science will resolve our analogue backlog digitisation issues, though it is clear that action is needed. Moreover, the tendency for parallel services to provide duplicate records without revealing their authority and the surprising proliferation and fragmentation of services makes it hard for users to keep up with services that exist, let alone make sense of data sets that they find through them. All of this points again to the potential benefits of a widely agreed archaeological data infrastructure which supports organisational and semantic interoperability.

Barriers for resource discovery, delivery and processing in archaeology:

- Persistent identifiers and ‘deep’ linking
- Research skills training in resource discovery
- Depth of cataloguing and indexing
- Intellectual property rights vs web services
- Markup standards and (practical) implementation
- Completion tools to import and connect data sets
- Non-digital backlogs
- De-duplication of similar sources
- Fragmentation and proliferation of services

6. Research sans frontières

E-science allows researchers in many parts of the world to collaborate in real time over sustained periods. Archaeology is international and the modern borders of the nation state would in most cases have been meaningless to the prehistoric populations that crossed freely the political boundaries that we now police. The internationalisation of archaeology means that differences in methodologies and institutions in different countries need attention. In particular question then arises of whether the massive computing power available through the grid can be deployed to resolve some of the multilingual issues that cross-border working creates.

At first inspection the prospects are not good. There have been a number of European multi-lingual activities in archaeology, including the development of a number of

European thesauri and word lists. These projects have taught us three lessons: that multilingual translation efforts are not trivial; that we should expect this work to throw up more questions than answers in the short term; and that the perceived benefits are modest as against the scale of the task in hand. As the harmonisation of heritage management policies and higher education provision across Europe grinds on, so there will be an expectation for the harmonisation of management tools – such as research outputs and data.

Although the prospects are bleak, we can at least sketch a number of prompts to continuing these efforts. We can be sure that, unlike geospatial interoperability, we cannot wait for anyone else to sort these issues for us. Characteristically international infrastructure development presents complicated technical challenges which we know we can meet. Moreover, the very fact that other humanists – especially linguists – are now involved in E-science means that there is likely to be attention paid to the linguistic challenges of the single European Research Area.

Challenges for working internationally are:

- Multiple languages
- Different traditions of recording and reporting and standards
- Cross border funding
- Organisational interoperability

7. Priorities for archaeology?

Some of the barriers discussed above are within the power of archaeology to resolve, so we need to consider how we would go about resolving it, and whom would be best placed to take it on. Other barriers are outwith the power of archaeology to resolve. Nonetheless we need to identify whom would be best placed to take them on. We may even decide there is merit in archaeology taking on this barrier on behalf of a wider scholarly community.