AHRC ICT Methods Network Expert Seminar on History and Archaeology

VIRTUAL HISTORY AND ARCHAEOLOGY

Humanities Research Institute, University of Sheffield, 19 – 21 April 2006.

ABSTRACTS

Using GIS to Study Long-Term Population Change
Ian Gregory, Queens University Belfast, Northern Ireland.

The study of long-term change in population has long been hindered by the complexity of the available data which are published for areas where boundaries change, making long-term comparisons impossible. GIS allows us significanty to address this problem by uniting the attribute data, that tell us *what* was happening, with spatial data that tells us *where* it was happening. As we have data for many points in time we also know *when* it was happening. Even within this comprehensive database, exploring change through all three of these components remains difficult. This paper explores approaches to this problem by studying changing patterns of infant mortality in England and Wales from the mid-nineteenth century. It shows how methodological innovations can gain new insight from data that have been extensively studied for over a century.

Which; What; When?
Manfred Thaller, University of Cologne, Germany.

In current discussions about metadata most proposals for simplifying descriptions reach the conclusion that most subject areas may be served by a scheme which starts from core categories (such as "What?", "Who?", "Where?", "When?") and builds upon these by qualifying the basic concepts into continually more refined descriptors.

While this is straightforward for bibliographical databases, it is considerably less so for the type of material used in history or archaeology, where a "What?" like the Acropolis, may actually require many "Whens?" for individual component parts, while other constructs—like a sketch of Atlantis—although this may actually have a convincing "When?" for the production of the sketch, raises quite a few problems when one wants to date the time to which it relates.

I propose that a general solution may be found when we assume that future information systems will be comprised of "objects" in the sense used by Information Technology. These have certain characteristics, which will be presented in this paper, which allow them to map the relationship between the component parts of archaeological "objects".

Using the CIDOC CRM, the most advanced ontology for the description of archaeological artefacts, it will be demonstrated that time and space may be seen as dimensions into which such "objects" may be mapped independently of the relationships between them, providing a wider context for database operations.

Visualization: Pretty pictures or enabling technologies
Vince Gaffney, University of Birmingham, UK.

The past decade has seen a revolution in digital technologies. To some this may appear to be limited to the benefits conferred by the web, increasingly powerful computers and cheap data storage. However, the historical disciplines, along with most of society, are increasingly subject to the pervasive influence of alluring digital visualization. What are the implications of this development and how should we react? Will these technologies provide us with new and useful insights into our research areas? Will visualization technologies be formative in research terms or will they just provide pretty pictures, useful for dissemination to the wider public or to impress funding bodies. This presentation will consider these issues and discuss their implications.

Spatial Technologies in Archaeology in the 21st Century
Paul Cripps, University of Southampton, UK.

GIS has undoubtedly made a significant impact on the archaeological discipline; its use in cultural resource management, development control, and research contexts in the public and private sectors in the last decade has greatly improved quality of and access to data whilst at the same time providing new opportunities for researchers and other end-users, including the general public. In the field of prehistoric landscape studies, considerable work has been successfully undertaken, demonstrating the benefits of such techniques as visibility analysis. At the same time, there has been an improvement both in availability and quality of source data combined with an increase in available computer processing power to work with such increasingly large and complex datasets. Having said this, the use of GIS in archaeological scenarios is often limited by its two-dimensional Cartesian, predominantly static view of the world. This paper will examine some of the ongoing developments in the broader field of spatial technologies in archaeology aimed at tackling some of these issues as relating to landscape studies. This will include applications of novel forms of GIS-based visibility analysis including dynamic and probabilistic viewsheds, use of laser-scan data for landscape analysis, the use of three-dimensional applications such as 3D Studio for analytical purposes in the field of visibility analysis and the potential for back-end database structures and middleware which facilitate temporal reasoning about spatial objects in all of our cultural heritage systems.

SUBJECT 2. THE PAST AND THE VIRTUAL REPRESENTATION OF TEXTS

Imaging of Historical Documents
Andrew Prescott, University of Sheffield, UK.

As recently as 1979, the eminent medieval historian G. O. Sayles declared that administrative records provide the only means by which historians can escape the 'uniformed tittle tattle' of literary sources such as chronicles and newspapers. However, historians using administrative records have in recent years become increasingly aware that these records are also complex textual productions, frequently as artificial in their discourse and deceptive as literary texts, and by no means more objective because of their 'official' provenance. As V. H. Galbraith has forcefully pointed out, 'official' records are as much literary productions as conventional 'literary' sources. The exploration of the textual layers of administrative documents is a complex matter, since knowledge of scribal practice and procedure in the production of such documents is at present limited. One tool which can be very valuable in developing a critical approach to administrative records is the digital image. This will be demonstrated with reference to records of the suppression of the English Peasants' Revolt of 1381. In order to progress such studies, an essential requirement is an editorial environment which allows editorial text and commentary to be integrated more closely with digital images. The *Edition Production and Presentation Technology,* developed by Professor Kevin Kiernan of the University of Kentucky, provides such an environment. The application of this technology to administrative records will be demonstrated. Such an approach is not only relevant to medieval materials; its importance in analysis of modern texts will also be considered.

Data, Structure and Analysis: XML mark-up and its application to historical data
Donald Spaeth, University of Glasgow, Scotland.

This paper will explore the limitations of two dominant models of textual markup, the Ordered Hierarchy of Content Objects (OHCO) thesis, based upon a document view, and the use of bibliographies as the model for

data-based work. Both present an over-simplified view of data, because it is predominantly hierarchical. In the real world (a domain which includes historical documents), things are more complex and data are semi-structured. How can a hierarchical system such as XML deal with such complexity?

For several years, I have been working on a category of documents: probate records, which have proved more complex than expected. The probate inventory, a record of the moveable goods someone owned at death, appears to have a hierarchical structure: houses contain rooms, and rooms have goods in them. Historians have made extensive use of the content of inventories, but I am primarily interested in their formats. It is my argument that the inventory is a representation, rather than a description, of the contents of a house, which tells us as much about the men who created it (appraisers) as it does about the owners of the house. Appraisers commonly introduced another level to the hierarchy: the item, consisting of a group of (possibly related) objects, which together are assigned a single valuation. The practice of appraisers can be studied by counting rooms, items and objects to reveal the choices they made about how to represent what they found, but these frequencies depend entirely upon the choices made about how to mark up the data. As anyone who has marked up data knows, to their cost, there is no better way to discover the complexities of documents than to try to mark them up.

The remainder of the paper will draw upon specific examples from the documents to expose structural complexities and their implications. In brief, the main problem is that the appraisers sometimes violate our hierarchical expectations, and we must decide whether (and how) to impose hierarchy, bearing in mind the ambiguities of document structure. Fortunately, XML fulfils criteria of semi-structured data, but we run the risk of double-counting some data and miscounting others. The available querying tools do provide ways of getting around these structural complications, so long as the researcher knows to ask the right questions.

Encoding and the Scholarly Community
Harold Short, King's College, London, UK.

Digital editions of historical documents created using current methods are likely to be fundamentally collaborative, and the resulting digital system is likely to be an example of what John Unsworth termed 'second generation' digital resources. Such systems are likely to be multi-disciplinary in creation and/or content, and to involve multiple methods and technologies.

The Guidelines of the Text Encoding Initiative (TEI) are widely used by humanities scholars to attach encoding to the source documents that are the subject of their research. The TEI was the largest collaborative project there has been in the humanities, involving hundreds of scholars in many countries over more than 10 years. Its work continues via the agency of the TEI Consortium. It is important to recognize, however, that the principal role of the TEI is to provide a continuing forum for scholarly debate.

It is also important that the wider community of scholars (and that includes the TEI community) look beyond the work already done to consider encoding in a broader context. In part this is to do with the 'old' question of whether text encoding or database encoding provides the 'better' framework for modelling the source materials, and making the resulting resources appropriately accessible; in part it is to do with the source document as a cultural artefact; in part it is to do with the role of the digital resource within the 'global' digital library.

Finding Needles in Haystacks: Data-mining in distributed historical datasets
Mark Greengrass and Fabio Ciravegna, University of Sheffield, UK.

This paper examines the application of Semantic Web technologies to the mining of historical data in distributed data-sets. The potential of such applications is clear. The problems, however, arise from working in a domain where ontology construction is constrained by the malleability of many of its conceptual frameworks. Using the experience gained so far from the Armadillo Project, which applies these technologies to distributed data-sets related to eighteenth-century London, this paper examines the extent to which these problems are likely to be overcome, so that the research dividend from data-mining can be realized.

Digital Searching and the Problem of the Ventriloquest's Dummy'
Tim Hitchcock, University of Hertfordshire, UK.

Most historians are the love-sick victims of manuscript. They sit, wrapt and lonely, communing with dusty papers and noisy parchment. If they are good, they use three or four archives and promiscuously sample the ideas and personalities of a range of institutions, but most are more monogamous than this and focus their every emotion on the clerk who constructed *their* archive. As a result, most history is driven by archival structures. We give greater weight to the role of governments and armies, parishes and hospitals, precisely because they form coherent archives. At the same time, we denigrate loose communities and informal connections, because they do not speak with a single voice. However, as it becomes ever more possible to search across archives, to hear other voices quietly whispering in the background, to mechanically combine the whispers of a thousand individuals into a stentorian shout that can be heard over the august tones of the clerk, the way we do history will change. This paper will suggest that our models of social order and change give unnecessary privilege to the institutions that by happenstance or good luck or good order, created the archives in the first place. It will suggest that the reconfiguration of the archives through digitization will radically re-invent history as a genre and as a social practise. As a generation of historians, we are faced with the challenge of removing the inky hand of the manuscript clerk from up our backsides, in order to engage in a broader, more promiscuous, and ultimately more humane, conversation with the past.

Using Computer-Assisted Qualitative Data Analysis Software (CAQDAS) in Historical Research: Some methodological issues from the experience of the 'Health of the Cecils' project.
Caroline Bowden, Royal Holloway, University of London, UK.

This software, developed mainly for use by social scientists, has been little used by historians to date; however closer examination of its potential would suggest that, with some adaptation, it has much to offer. This paper will focus on key stages of the project, which has been running for the past two and a half years, to consider the experience of using CAQDAS and its implications for historians. The Health of the Cecils (c.1550–c.1660) project based at Royal Holloway, funded by the Wellcome Trust for three years (Sep 2003–Sep 2006) is studying the experience of health care in one of the great aristocratic households of early modern England. The study extends to the wider family and the household as well as their wider political and social contacts. The family has been chosen for a number of reasons: extensive collections of archives survive, particularly for the first two generations in the study; members of the family suffered significantly from several medical conditions including gout, scoliosis, and agues on which they sought and were offered advice on many occasions. They were in a position to spend very large amounts of money on health care, often employing the same physicians and apothecaries as the Royal Family. The equivalent of £315,956.17 was spent on treatment at Bath for Robert Cecil, Earl of Salisbury in the period just before he died in 1612. At the same time the Cecils received advice and offers of remedies from family and friends. For example, Mary Countess of Shrewsbury sent Robert Cecil her quintessence of honey and laudanum and the Bishop of London sent William Cecil advice on how to treat his painful back. Their political status meant that physicians from other countries, including Germany, wrote offering medical services in return for Cecil support in helping them to establish a career in England. The Cecil households consisted of very large groups of servants, stewards, wards and for the third generation living at Hatfield, children. Account books for the seventeenth century include evidence of health care costs as part of the overall budget. As well as illnesses, other health experiences relating specifically to age or gender such as childbirth, ageing and death appear in the letters. The manuscripts offer a significant opportunity to study the experience of health care over an extended period of time when attitudes to medicine and medical knowledge were changing substantially. The main manuscript collections studied are the Lansdowne Papers held at the British Library, the Salisbury Papers still owned by Lord Salisbury at Hatfield House, where they have been since the collection was first formed, and State Papers. Evidence of health care has been found in a wide variety of documents including bills, accounts, medicinal and culinary receipts, medical reports and letters. In this paper, I examine the problems and potential research conclusions to be gained from applying CAQDAS to this kind of historical material.

Shared Spaces: Lbrary and archive metadata, encoded documents and research needs
Susan Hockey, University College, London, UK.

Library and archive catalogues are very differently structured, but they share a common objective: they are means to find materials on shelves or in repositories. However as more materials become available in electronic media, the role of the catalogue and finding-aid has to change to reflect research needs. Using the LEADERS project as an example, this paper examines how archival finding-aids, electronic transcriptions and digital images of

documents may be integrated to create a single flexible research environment, and speculates on the differences this mode of delivery is likely to make to research practices.

Attempts to Construct a Common Platform for Archaeological Reports
Julian D Richards, University of York, UK.

Each year millions of pounds are spent on developer-funded archaeology in the UK. Following the introduction of PPG16, from 1990–99 there were over 28,000 separate archaeological investigations (Darvill and Russell, 2002). Little of this work fed back into academic research and most was lodged as grey literature reports in local planning offices. The OASIS project (http://ads.ahds.ac.uk/project/oasis/) is endeavouring to provide an index to grey literature and to create an online virtual library. However, for research results to be made readily locatable and retrievable it would be necessary to develop a standardized protocol or template for fieldwork reports. The majority of archaeological fieldwork reports follow an accepted format and implicit structure, but archaeologists (in the UK and the USA) have resisted attempts to impose a common platform. Gray and Walford (1999) advocated an XML-based approach to creating standard site descriptions, and Meckseper & Warwick (2003) and Falkingham (2004) each investigated the XML-tagging of grey literature. In 1999 UCLA undertook a failed attempt to introduce a digital imprint template. In Norway the Museum Documentation project has undertaken TEI mark-up of antiquarian accounts for many years. Most archaeological data collected in the field are highly structured but would require detailed semantic mapping in order to render them truly interoperable on a cross-project basis. In the United States a cyberinfrastructure for archaeology has been advocated, with automated harvesting of metadata allowing cross-searching of distributed archives and e-prints. These approaches have the potential to contribute to an archaeological Semantic Web but require agreed ontologies before they can proceed.

Darvill, T. and Russell, B. (2002) *Archaeology after PPG16: archaeological investigations in England 1990–1999.* Bournemouth University School of Conservation Sciences Research Report 10..

Crossing an 'Information Divide': The OASIS project and its use of XML schema
Catherine Hardman, University of York, UK.

It has long been recognized that the information flow between fieldworkers, local government archaeologists, national heritage agencies and the rest of the community could be improved. The Online Access to the Index of Archaeological Investigations (OASIS) project aims to use IT to ease the flow of information from those undertaking fieldwork to the wider archaeological community. The OASIS system captures the data, holds it in a database, and then allows different heritage professionals to access it. This paper will describe that process, the ways in which the project has progressed since its launch in 2004 and the way in which XML schemas have been developed in order to facilitate the easy transfer of data from the OASIS database to a variety of different databases across the curatorial community.

SUBJECT 3. THE PAST AND THE VIRTUAL REPRESENTATION OF OBJECTS AND EVENTS

Digital Artefacts: Possibilities and purpose
David Arnold, University of Brighton, UK.

Developments in 3D scanning and recording technologies now mean that almost any level of accuracy is now achievable in digitization of historic artefacts. This capability poses challenges to cultural heritage professionals (archaeologists, historians, curators, etc.) who are now confronted with previously unimaginable opportunities and therefore must ask and answer questions on the underlying purpose of collecting digitized models of artefacts, archaeological contexts, historic monuments, buildings or ruins. As with many engineering challenges, there are significant issues of price/performance in undertaking data capture and the range of solutions may lend themselves to addressing different heritage applications. In this discussion I will address some of the potential purposes for creating and using digital artefacts, ranging from analysis to public dissemination, and pose more questions than answers in considering the fitness for purpose of data being collected and the challenges of re-purposing data collected for one purpose for use in a different context.

"Oh, to make boards to speak! There is a task!" Towards a Poetics of Paradata
Richard Beacham, King's College London, UK.

Scholars employing 3D visualization both as a tool and an outcome of their research need to determine how to document the process and the presentation of results in such a manner that other scholars can fully understand and rigorously evaluate them, thereby enabling such methods to acquire greater recognition and standing in the scholarly community, and driving up standards of such work throughout the academic and cultural heritage sectors. A major challenge for those working in this area is that we are producing visual "texts" that we and our colleagues in the scholarly community do not yet fully understand how to "read".   As part of its work under the AHRC ICT methodologies initiative, the Kings Visualization Lab recently hosted, together with the EU Network of Excellence, EPOCH, a symposium which aimed significantly to assist in providing the basis for future standards and methodologies in the field, both for enhancing the quality of the actual modelling process, and for establishing minimum levels of documentation necessary for users critically to assess visualization-based research processes. A major objective is to identify and disseminate the choices and decisions that occur during the complex process of modelling, which may include the reasons for choices made, as well as indications of possible alternative hypotheses. KVL have coined the term "paradata" to designate this process. The symposium produced a draft document, the "London Charter", which identifies a number of key guidelines and principles, relating to such areas as subject communities; aims and methods; the use of sources; transparency requirements; documentation; standards; sustainability; and accessibility. This presentation will discuss this initiative and briefly illustrate the type of issues it addresses.

Computing Aspects of the Corpus of Romanesque Sculpture in Britain and Ireland (CRSBI)
Anna Bentkowka-Kafel, Courtauld Institute of Art, London, UK.

The CRSBI (www.crsbi.ac.uk) is an evolving electronic archive of stone sculpture from the period c.1050–c.1200. Over 5,000 sites featuring Romanesque sculpture have been identified, 70 per cent of which have been recorded so far by the project's volunteer fieldworkers. 700+ reports have been published online on the project's website. The reports vary in scope and range from a single object report illustrated with a handful of images to a book-length entry for Ely Cathedral. The project involves an Access database as a project management tool and a repository of data, mostly image metadata, XML-encoding of site reports, digital photography, digitization of prints and negatives, image editing and processing for archiving and online publication, and Perl scripts for data linking. In this paper, I intend to compare the methods and solutions adopted by this project with others of this kind, and to discuss how a complex digital resource like this is likely (when completed) to enhance the study of Romanesque sculpture.

Virtual Restoration and Manuscript Archaeology: A case study
Meg Twycross, University of Lancaster, UK.

Manuscripts have a multiplicity of ways of becoming illegible, ranging from the accidental—wear and tear, mice, immersion, incineration—to the intentional, malicious or would-be benign. There is an equally wide range of (non-invasive) ways of recovering their readings, often developed for forensic science. Some techniques have been around professionally for some time, but in the last few years 'virtual restoration' may be carried out by anyone with a decent PC, a willingness to experiment with Adobe Photoshop, and the means of acquiring high-resolution scans, preferably taken under a range of light conditions. My case study describes some techniques I have deployed on images of a manuscript in York City Archives which has suffered more than its fair share of the depredations of time, accident, and archivists' repair. It is also a heavily-layered palimpsest, a patchwork of erasures and alterations by the medieval clerical staff of the City Council. One task is to recover what lies under the erasures, and to identify the hands which made the alterations. This form of non-destructive manuscript archaeology treats each level of text as significant, as a stage in the historical process which it records.