

This paper was presented at the conference:

Oxford eResearch 2008

11-13 September 2008

University of Oxford

Conference website:

<http://www.oii.ox.ac.uk/microsites/ereseach08/index.cfm>

Conference papers collection:

<http://ora.ouls.ox.ac.uk/objects/uuid%3A64aa6f39-7e81-4d42-a008-ee2d7524bd67>

Conference organisers:

Oxford Internet Institute

University of Oxford

1 St Giles, Oxford OX1 3JS

<http://www.oii.ox.ac.uk/>

email: enquiries@oii.ox.ac.uk

Oxford e-Research Centre

University of Oxford

7 Keble Road, Oxford OX1 3QG

<http://www.oerc.ox.ac.uk/>



'A Virtual Research Environment for the Study of Documents and Manuscripts'

A.K. Bowman (Presenter), *Centre for the Study of Ancient Documents, University of Oxford*

C.V. Crowther, *Centre for the Study of Ancient Documents, University of Oxford*

R. Kirkham, *Oxford eResearch Centre, University of Oxford*

J. Pybus, *Oxford eResearch Centre, University of Oxford*

Abstract

The scholar interpreting a degraded documentary text has a broad range of relevant electronic tools available; but the interaction is largely in one direction and the experience is fragmented by the dispersal of both the original documents and the electronic resources. Decipherment and transcription are scholarly activities that naturally benefit from shared visualization and multiple perspectives, but the potential of ICT to mediate collaborative investigation has yet to be fully exploited. A VRE for the Study of Documents and Manuscripts is a two-year JISC-funded project designed to explore these challenges through the construction of an integrated collaborative environment for the study of ancient documentary artefacts.

In this paper we describe the background to the project, the need for VRE tools and technologies in documentary and manuscript studies generally and the specific impetus provided by the concerns and experience of two of the authors in working with ancient documentary materials; the potential of VRE technologies is illustrated through a detailed case study involving the collaborative decipherment of a Latin writing tablet from Tolsum in Frisia. In the second part of the paper, we discuss the construction of the technological framework for a VRE for documentary and manuscript studies from a variety of perspectives: the creation of a working environment within a portal framework: the specific needs of documentary and manuscript scholars for a customizable and interactive viewing frame and annotation tools; the integration of existing electronic tools such as Vindolanda Tablets Online¹ and the Lexicon of Greek Personal Names² xml database; the potential integration of specialized image management and manipulation tools (some of which are being developed in a parallel and current AHRC-funded e-Science project, Image,

Text, Interpretation: e-Science, Technology and Documents); and collaboration with the VERA archaeological VRE based around the Silchester Roman Town project.

Introduction and Objectives

The VRE for the Study of Documents and Manuscripts (VRE-SDM)³ is a two-year JISC-funded project designed to complement research into damaged and illegible documents, funded by a major e-Science grant, at the Classics Faculty's Centre for the Study of Ancient Documents (Image, Text, Interpretation: e-Science, Technology and Documents (eSAD)).⁴ The aim of the VRE-SDM project is to construct a pilot of an integrated environment in which the data (documents), tools and scholarly *instrumenta* will be available to the scholar as a complete and coherent resource. Although initial development has been driven by the requirements of researchers drawn from the Papyrological and Epigraphical communities, the VRE has been designed from the outset to be extensible to a wide range of humanities disciplines.

The objectives of the project are to create an environment within which a researcher will be able to:

- View, manipulate and enhance digitized images of documents and manuscripts within a portal framework
- Search across multiple, distributed data sets, images and texts
- Select, store and organize items from the above, in a 'personal workspace'
- Add annotations to these items to store personal thoughts and responses
- Support collaboration by allowing multiple researchers in separate locations to share a common view of the workspace, in conjunction with real time

communication via Chat, VoIP and desktop integration with AccessGrid

- Allow a collaborator to comment, point/highlight, discuss and annotate the items in the shared workspace
- Gain comprehensive user requirements and expand the use of the VRE for documentary and manuscript scholars in other fields of humanities research

Background

An initial survey carried out by the Building a Virtual Research Environment for the Humanities (BVREH) team between June 2005 and September 2006 defined the range of services that a Virtual Environment for the Humanities should offer – from information about researchers and their interests and about conferences, lectures and seminars, to integrated communication and collaboration tools to support advanced research. Through building a number of ‘demonstrators’ or mock-ups the project team was able to consult humanities researchers directly about their needs. One of these demonstrators was A Virtual Workspace for the Study of Ancient Documents (VWSAD) which addressed the specific needs of ancient documentary scholars. Building on a range of needs highlighted by the original survey this demonstrator has now been expanded into the pilot VRE for the Study of Documents and Manuscripts (VRE-SDM), an environment adaptable to the needs of documentary scholars across the humanities.

The approach of the BVREH project, conceived and supported by enthusiastic user champions working within the Humanities Division at Oxford, marked a significant change for the humanities researchers surveyed. Respondents to the initial survey commented that this was the first time that they had been consulted directly about their ITC needs and they found it refreshing and stimulating to be invited to work alongside the technical team to provide user feedback through every stage of development.

So, what is a VRE?

What actually is a Virtual Research Environment and how will it help documentary scholars? The concept of a Virtual Research Environment is

constantly evolving as more and more work is carried out in this area, but JISC offers the following definition:

‘The purpose of a VRE is to help researchers in all disciplines manage the increasingly complex range of tasks involved in carrying out research. A VRE will provide a framework of resources to support the underlying processes of research on both small and large scales.’⁵

This definition closely matches the user requirements established by the BVREH survey in 2006 and has provided the basis of the VRE for the Study of Documents and Manuscripts. It also corresponds to the specific needs of ancient documentary scholars detailed below.

The scholar’s perspective

The initial impetus for two of the authors to construct a ‘Virtual Workspace’ for studying ancient documents came from a meeting at the Centre for the Study of Ancient Documents (CSAD) in 1996, described in the following account by one of its participants, Professor Dorothy Thomson:

‘In 1996 an important demotic papyrus from Rifeh (now housed at University College, London), was brought to the Centre for the Study of Ancient Documents (CSAD) to be scanned and digitized. Previously the text had only been readable through the use of a magnifying glass with many days spent poring over the text. Once the papyrus had been scanned, the team of papyrologists assembled in front of a computer to see what benefits digitization might bring. Far from the barely legible original, the images on screen made the “script come to life” and readings and suggestions from the group flowed freely leading one of the group, Dorothy Thompson, to conclude: “I have seen the future and this future works – at least so far. Work in different countries on the same text at the same time can now take place without problem and for a long and difficult text, where the writing is small and faded; the possibility of working on the image on the screen is in itself a great advance”.’⁶

This enthusiastic response details how advances in technology and digital imaging have transformed research in the field of ancient documentary studies. Not only have texts become more legible, but perennial problems of time and access are being significantly reduced. While the scenario above describes a number of researchers travelling to Oxford to gather around a single computer screen, the 'Virtual Workspace' demonstrator has been designed to take the possibilities further, enabling Professor Thompson's vision of the future to become a reality (as in the example of the Tolsum Latin tablet described below): to 'work in different countries on the same text' simultaneously allowing researchers to collaborate and form 'virtual gatherings' without being unnecessarily constrained by time, money or travel. The team envisaged from the beginning that such a workspace would be useful not only to ancient documentary scholars, but to those working with documents and manuscripts across the humanities, from specialists in medieval music manuscripts to English scholars working with digitized manuscripts of Jane Austen's work. This premise was backed up by the BVREH user survey which reported in 2006 that:

'Interviewees want tools to enable them to work collaboratively on documents, to share material with collaborators and to view material simultaneously with colleagues wherever they might be based. At the same time interviewees want to communicate either through video conferencing or real time chat facilities and to enable a collaborator to point/highlight and annotate items throughout the discussion.'⁷

The next stage then was to start to gather the requirements of documentary scholars in the Humanities. As the BVREH team had previously worked closely with the ancient documentary specialists at the CSAD, it naturally followed to utilize this already established user base with enthusiastic user champions as the first group with which to test the pilot on a continuous, iterative basis, before exposing it to scholars in the English Faculty and elsewhere.

Ancient Documentary Studies – Case Study

Scholars who edit ancient documents are almost always dealing with damaged or degraded texts

and ideally require access to the originals, or the best possible facsimiles of the originals, in order to decipher and verify readings, and also to a wide range of scholarly aids and reference works (dictionaries, name-lists, editions of comparable texts etc.) which are essential for interpretation of their texts. The originals may be in widely scattered museum and institutional collections (individual parts of archives and even single documents have suffered this fate), and the full range of scholarly tools is immediately available in only the best public and university library collections. The decipherment of texts from facsimiles of the originals, such as conventional photographs or digital images, can also be supported by sophisticated tools for image-enhancement and signal processing. The storage and processing power of ICT, along with on-line access to remote sites and resources, offer the prospect that all of these facilities can be made available to the researcher in a VRE of the kind described here.

The development of technologies of image-enhancement and signal-processing is a vital component of a VRE for studying ancient documents. The use of multi-spectral imaging for texts written in carbon-based ink is comparatively well-known and has been used for many years, but there is still enormous scope for developing and refining techniques which may be applied to other kinds of written material which will not yield their texts to the naked eye: these include texts scratched on wooden stylus tablets, incised in metals such as lead or bronze and stone inscriptions on which the top surface has been wholly or partly worn away. The construction of user-friendly tools for imaging and enhancement is a key feature of the tool-kit which such a VRE will eventually offer.

User Requirements – Ancient Documents

In the spring of 2008 Professor Alan Bowman, Professor Roger Tomlin and Dr Charles Crowther began a collaboration over a number of months to decipher a Latin text written on a wooden stilus tablet found at Tolsum in the Netherlands in the early 20th Century. The team decided that this would offer an excellent opportunity to refine user requirements for the VRE for the Study of Documents and Manuscripts, and filmed the researchers'

collaborative meetings for use as a case study. The aim of the filming was to discover and document the inherent practices, tools and processes which are currently used to decipher ancient texts and to establish ways in which a Virtual Research Environment might emulate, support and advance these practices. The process was designed to gain requirements for the construction of the VRE, not to test the VRE interface itself. Between February and May 2008 the VRE team filmed four collaborative meetings between the three experts. It should be noted that these meetings were organized by the scholars to address a real research question (documented by Professor Bowman below) and the filming was designed to be as unintrusive as possible.

“The stilus tablet was found at Tolsum in the Netherlands in the early 20th century and contains a Latin text that is well known to Roman historians and experts in Roman law. The writing was transcribed, interpreted and published by the Dutch Classicist C.W. Vollgraff in 1917 and his text has remained the basis of understanding the document ever since. It was discussed and minimally revised in 1998 by another Dutch scholar (E. Slob), who was able to conduct a carbon dating test which supported the idea that it was written in the early Roman imperial period (probably first century AD).

Although Vollgraff’s text showed many peculiarities of formula, syntax and nomenclature, it has always been accepted as basically sound. One of the main reasons for this is that the subject matter, as he interpreted it, fitted very well into a known historical context. A contract for the sale of an ox is appropriate to the area and the period because we know from the historian Tacitus that in AD 29 the Frisian tribe revolted against Rome because of the heavy burden of taxation in ox-hides imposed on them (Tacitus, *Annals* IV.72 ff.) – this was (and still is) a region well-known for its famous Friesian cattle breed. However, there has always been some uncertainty about the exact date of the stilus tablet because the names of the consuls in the dating formula were not clearly legible; Vollgraff and later scholars thought they might be the consuls of either AD 29 or AD 116, but the earlier date has generally been favoured simply because it did make the text fit so well into the context of the Frisian

unrest described by Tacitus for the year 29.

Our re-investigation of this document began because Dr Klaas Worp, an old friend and colleague who is a papyrologist working in Leiden, knew of our attempts to develop imaging techniques for deciphering this kind of writing and asked us whether we could make some digital scans and finally establish the correct reading of the consular date of the document. The curator of the museum in Tolsum brought the tablet to Oxford so that we could do this. We did propose to concentrate on the dating clause, but even a quick look at the original tablet and the first set of scans showed that there were fundamental mistakes in Vollgraff’s original reading. Specifically and most seriously, his reading of the word ‘bouem’ (ox) in line 4 on the front, which is the only possible reference in the text to an ox and hence the only reason for thinking that this document is the sale of an ox, is simply wrong: the correct reading is clearly the latin relative pronoun ‘quem’ (‘whom’). It was thus clear that a complete re-reading and re-interpretation was needed”

During 2006-2007 the BVREH project held a number of workshops funded by the Arts and Humanities Research Council (AHRC) entitled ‘User Requirements Gathering for the Humanities’.⁸ Along with recommendations to use iterative development with humanities researchers and to establish user champions from within humanities disciplines, one of the most important practice suggestions was to:

‘Encourage developers to attend meetings with researchers and to understand the nature of their research and the material with which they are working’

It was clear, therefore, that the pilot VRE would only be effective if developed in constant consultation with documentary specialists (in this case study, Classicists) and that a dialogue should be encouraged between the developers and researchers on a regular and continuous basis. The presence of the developers at collaborative sessions benefited the project greatly as they were able to form a direct understanding of the user needs having witnessed their processes at first hand: simple subconscious actions such as tracing a letter in

the air, or following the line of a down stroke on a digitized version of the text on a projector were observed by the project team, and allowed us to consider how these collaborative and expressive practices might be emulated in the VRE when individuals collaborate in separate locations. Without the benefit of these meetings and subsequent study of the videos of the sessions, important user requirements could easily have been missed.

The technical perspective

Working in the first instance with the user requirements from the above case study, the pilot VRE has adapted Open Source tools to enable annotation and sophisticated document viewing, making use of existing VRE tools to facilitate communication and collaboration between scholars. The project is keen to reuse and re-purpose tools and software wherever possible, choosing to concentrate on the developing user requirements of humanities scholars rather than writing proprietary software that would be specific to our system. With this in view, we have established an instance of the uPortal framework which offers interoperability with other Virtual Research and Virtual Learning Environments and allows us to reuse JSR-168 portlets from other projects whilst making our components easier for others to reuse (JSR-168 portlet containers have been used by a number of other VRE pilot projects)⁹. The uPortal open source software, developed by a consortium of Universities additionally allows us to provide a framework which can be customized directly by the users who will be able to compile their own interfaces using portlets which offer the tools and services relevant to their own research.

This means that in the long term the VRE will be able to provide tools to researchers across the humanities. Some, such as the viewing and annotation tools, will be relevant to the broadest range of scholars, while other more specialist tools can be added by the individual users or groups as and when needed. For the ancient documentary specialists the VRE will be an environment which emulates the current decipherment process, with the extra benefits that scholars will be able to conduct real time meetings and to annotate digitized, degraded texts. Additionally, the VRE will offer an area within which other more specific tools may be deployed, such as those from the eSAD project

(to aid the decipherment of degraded documentary texts through character recognition and decision support software) along with the functionality to search across specific datasets such as the Lexicon of Greek Personal Names (LGPN), the Vindolanda Tablets Online texts and more besides. This approach of adding elements relevant to the individual scholar or specialism creates a customized workbench re-workable and reusable across the humanities.

Document viewing and annotation

One of the integral parts of the VRE for the Study of Documents and Manuscripts for all users is a superior image viewer which can handle high resolution images and can download images at high speed over the web. The viewer was developed from the Giant Scalable Image Viewer (GSIV)¹⁰ open source code. This viewer allows large, high resolution scans typical for digitized views of documents and manuscripts to be viewed in a web browser. By splitting the original image up into tiles and using javascript within the browser, only those parts of the image necessary to show the current view are downloaded and the user can pan or zoom the image, viewing it in detail. The viewer has been implemented as a portlet and extended to allow the user to select and annotate regions of the image. For the purposes of annotation we have extended the Annotea¹¹ vocabulary to deal with the needs of complex images. This allows researchers to make specific comments on areas of the text which they can save within their own personal workspace or share with others through the web based communication and collaboration tools.

Integrating electronic resources

The annotations and other metadata used within the system are stored as RDF¹² with the jena¹³ software used to manage an RDF triplestore. The use of RDF allows greater flexibility in the types of data which can be stored and integrated. One demonstration of this was importing data from the Vindolanda Tablets Online project. The images within the resource are available as XML marked up to the Epidoc standard, a flexible but rigorous standard for the digital encoding and exchange of scholarly editions of ancient texts. Because the texts were marked up in this way, it was a relatively simple

process to extract the known readings of the image and import them into the RDF triplestore for scholars to use within the VRE. The Lexicon of Greek Personal Names, a significant international resource which collects and publishes all known Greek personal names drawn from all currently available sources from specific geographical regions, is also now storing entries in XML. The project is in the process of making a set of search services available as web services which will make it possible for the VRE team to construct a portlet which will offer a search box and return results that link directly back to the LGPN servers.

For the ancient documentary scholars in our earlier case study, adding services such as those described above marks a significant step forward in the use of the VRE in the decipherment of unread or unrecognized texts. Although the case study focuses on a Latin rather than a Greek text, the process of referring to resources similar to the LGPN and the need to cross-search electronic resources across the web are relevant to documentary scholars across the humanities. Providing access to these resources within the research environment will allow the scholar to pull together relevant research in a personal workspace or, indeed, to share research with others in a way that is currently not possible.

The wider (ancient) world

The project is also working to extend the context of documentary study by treating documents not as disembodied texts but as artefacts with an original archaeological or physical context which can, in principle, be recovered or reconstructed. The significance of this approach is that the construction of a VRE appropriate for texts as artefacts opens up the possibility for the archaeologist and the textual scholar to work both separately and together within a unified environment. Currently, documentary scholars can sit in studies or libraries deciphering their texts, but it is important to remember that all such texts come from specific physical contexts on ancient sites. Knowledge of the details of archaeological context can often aid decipherment and interpretation by suggesting what sort of a text one might be dealing with or what other objects or structures are associated with it. Conversely, the archaeologist in the field may be led to a better understanding of the

physical context by having immediately at hand information which a deciphered document can offer. Provision of a full ranges of resources, text-and-image databases, scholarly reference works and technical tools, in a way which makes them mutually interactive and accessible to communities of documentary specialists and archaeologists, will be an enormous benefit of a successful VRE implementation.

To meet this need, the project has been investigating how the VRE-SDM might interoperate with other Virtual Research Environments through a collaboration with the the Silchester Roman Town VRE (VERA), based at the University of Reading, which has already developed a sophisticated system for registering, tracking and analyzing data recorded in the field to allow efficient recovery of information on any given artefact and its original context. The projects have created a system to allow the cross-database searching of archaeological data, through the XDB-Arch project¹⁴. This involves the use of the Dublin Core for Collections standard¹⁵ to describe groups of objects held by each collaborator (Vindolanda tablets in the case of Oxford, and objects within the IADB (Integrated Archaeological Database) at Reading). A search specified in CQL (the Common Query Language, a formal language, based on the semantics of Z39.50 used for representing queries to information retrieval systems such as digital libraries, bibliographic catalogues and museum collections) is then distributed to each database using the Tycho middleware¹⁶ and the results combined to return to the user in one list. For the purposes of the XDB-Arch project the VREs are focusing on objects that might be of interest both to archaeologists and ancient documentary scholars in order to demonstrate the wider potential of VRE collaboration and cross database searching.

The VRE team have also been working closely with the eSAD project based at the Oxford eResearch Centre. The project is working on creating tools which can aid in the reading of damaged texts like the stilus tablets from Vindolanda and, indeed, the Frisian tablet described above. At the same time, the project is exploring how an Interpretation Support System (ISS) can be used in the day-to-day reading of ancient documents, keeping track of how the documents are interpreted and read.

The benefit of this collaboration for ancient documentary specialists is that the projects are creating a seamless link between tools designed specifically for the decipherment of ancient documents and the Virtual Research Environment through which these tools can be utilized alongside communication and annotation tools by a broader base of users. Working collaboratively with projects such as eSAD and others across the humanities is vital to ensure not only that the VRE is relevant to the specific research needs of humanities researchers, but that the experience is as coherent and as integrated as possible.

Future Challenges

With the pilot VRE for the Study of Documents and Manuscripts still in its relative infancy there are inevitably challenges still to be faced both in terms of user uptake and continuing development. The first priority will be to continue to broaden the scope and appeal of the service to researchers across the humanities. Work is already in progress to extend the viewing and annotation capabilities of the system to a project at Oxford University, led by Professor Kathryn Sutherland, concerned with the digitised manuscripts of Jane Austen ('Jane Austen's holograph fiction manuscripts: a digital and print resource'). This collaboration is particularly interesting as the English scholars are keen to make use of cross-database searching and have prioritised four electronic resources; Eighteenth Century Collections Online (ECCO), Eighteenth Century Books Online (EBBO), the Chadwyck Healey Collections and the Samuel Johnson Dictionary for this purpose. As the VRE team learn more from the XDB-Arch project and from our collaboration with the LGPN, we will be better equipped to consider similar work in fields across the humanities and to work with individual projects to meet their growing and varied needs. The work on the XDB-Arch project offers the potential to scale up and to allow the connection of multiple resources. It remains to be seen whether the current design is capable of meeting the needs of this wide range of other projects or will need to be extended to do so. Even if the XDB metadata schemas are extended to cope with all these resources, it will remain of course a challenge to gain widespread deployment among archaeological and classical resources.

It is therefore essential that the VRE provides both tools and standards which projects across the humanities can make use of, and also that it provides the impetus for both smaller projects and larger electronic resources to present their own tools and services for reuse as portlets within the environment. For this to be possible, projects will need to consider how they will work together to make use of compatible standards. The most significant way that other projects can interoperate with the VRE is by creating versions of their tools which are exposed as portlets conforming to the JSR-168 standard. This would allow the portlets to be deployed directly within the VRE framework and enable VRE users to add these components into their workspace. It may be that other projects create separate portlet versions of their services, or they may choose to base their services on portlets in the first place, in which case they would have the opportunity to reuse VRE-SDM and other portlets within their own site. The challenge for the VRE team, then, is to persuade others through successful demonstration and collaboration with projects such as eSAD and the LGPN that the VRE is a useful and integral part of the researcher's toolbox and as such a worthy environment through which to provide tools and services.

Conclusion

The potential for scholars to make use of a Virtual Research Environment for the Study of Documents and Manuscripts is apparent across the Humanities, from those studying ancient documents to those working with the texts of Jane Austen. We have attempted to address this potential by focusing time and resources on gathering sound user requirements, focusing in the first instance on user champions in the field of ancient documents. The project has chosen to use Open Source tools, reusing and not reinventing what is already available. The project hopes to work with and encourage other projects to consider the adoption of mutually beneficial standards to enable the VRE to provide the best tools and services to researchers in the most coherent and beneficial way.

References

¹ Available at: <http://vindolanda.csad.ox.ac.uk>

² Available at: <http://www.lgpn.ox.ac.uk/>

³ Available at:
<http://bvreh.humanities.ox.ac.uk/>

⁴ Available at :<http://esad.class.ox.ac.uk/>

⁵ Available at the JISC Website:
<http://www.jisc.ac.uk/whatwedo/programmes/vre2.aspx>

⁶ The full account of Dorothy Thompson's experiences regarding the Demotic Papyrus are available at:
<http://www.csad.ox.ac.uk/CSAD/Newsletters/Newsletter2/Newsletter2a.html>

⁷ Kirkham, et al. (2007) *Building a Virtual Research Environment for the Humanities JISC final report* [online] available at:
http://bvreh.humanities.ox.ac.uk/files/Microsoft%20Word%20-%20JISC_Final_Report_Web.pdf

⁸ Available at:
http://bvreh.humanities.ox.ac.uk/news/Requirements_Gathering_Workshops

⁹ Available at:
http://www.jisc.ac.uk/whatwedo/programmes/programme_vre.aspx

¹⁰ Available at:
<http://www.mojavelinux.com/projects/gshiv/>

¹¹ Available at:
<http://www.w3.org/2001/Annotea/>

¹² Available at: <http://www.w3.org/RDF/>

¹³ Available at: <http://jena.sourceforge.net/>

¹⁴ Available at: <http://xdb.vera.rdg.ac.uk/>

¹⁵ Available at: <http://dublincore.org/about/>

¹⁶ [Mark A. Baker](#), Matthew Grove (2007): *Tycho: a wide-area messaging framework with an integrated virtual registry*. [The Journal of Supercomputing Vol. 42](#), no. 2, pp. 83-106.