

Virtual Vellum – final report

Virtual Vellum (VV) set out to provide distributed access to research-quality high-resolution digitisations. The 'test bed' corpus was a group of complete digital surrogates of illuminated manuscripts of the *Chronicles* of Jean Froissart, the principal historical source for the first phase of the 100 Years War between England and France (1325-1400). Six complete and closely-related manuscripts dating from around 1408-1415 were digitally photographed on site at the partner libraries: Stonyhurst College, the Bibliothèque Royale Albert 1er, Brussels, the municipal libraries of Besançon and Toulouse. These archival files were brought home on hard drives, and stored on a University of Sheffield server.

Prompt access to the images within the Virtual Vellum viewing environment is achieved by storing them in tile-based data structures. We started with JPEGs of around 8MB, then began to explore the potential of the JPEG 2000 format, for reasons explained below. The files are enabled in a viewer which will form part of a more complex Virtual Research Environment. When seeking to view an area of a given image at very high resolution, it is not necessary for a user to access the whole of that image; he/she can simply retrieve the relevant portion as tiles, at a level of magnification the user can specify, from the pre-programmed tile structure. A 'hotspotting' facility will be added later, whereby information about the MS (translations, transcriptions, notes on the art and calligraphy) will be accessible by clicking on an icon for a given folio/image. These will soon be available to scholars from the internet via an SRB-enabled portal.

ICT Guides method taxonomy classification(s) (with acknowledgements to support staff at the Arts and Humanities e-Science Support Centre, and its Website)

2D scanning; Manual transcription; Coding/Standardization; Data modelling (network); Data modelling (object-oriented); Image enhancement; Markup/text-encoding – descriptive – document structure; Markup/text-encoding – descriptive – linguistic structure; Record linkages; Graphical collaborative publishing; Graphical resource sharing; Image optimization for the web; Server scripting; Textual resource sharing; Web browser scripting

Bibliography and links

<http://www.shef.ac.uk/hri/projects/projectpages/virtualvellum.html>

Project contact details

Professor Peter Ainsworth, project PI
Dr Michael Meredith, Technical Consultant
Mr Colin Dunn, Digitisation and Imaging Consultant

Description of aims

Manuscripts such as those containing Froissart's *Chronicles* are immensely valuable, both financially and in terms of their research content. Libraries and other publicly-funded institutions charged with their curation and care are reluctant to allow scholars – still less the general public – even limited access to them, especially since regular or even periodic exposure to light results in (incremental) damage. The public Library at Toulouse operates a policy whereby vellum manuscripts are only exhibited in show cases for a maximum of three months at a time, in very reduced lighting conditions. Researchers are allowed access to the manuscripts, but for necessarily short intervals, and are typically obliged therefore to rely for intensive, long-term research on surrogate media, usually in the form of monochrome microfiche copies or b&w microfilms (though colour microfiches or 35mm filmstrips are available in some libraries). In-house digitisations are occasionally made available, but these attract very high direct and overhead costs. Furthermore, there are complex practical and political problems for researchers wishing to employ their own photographers, and to enter into dialogue with them.

The aim of the EPSRC Virtual Vellum demonstrator project was to overcome these very severe limitations by providing free access for scholars (subject to the usual copyright and IP safeguards) to high-resolution digital photographs of the manuscripts, using open source and grid-supported delivery software to enable their use in a variety of environments including individual users' desktops, Access Grid venues and live displays at conferences or seminars. From October 2007-March 2008 the virtual manuscript material will be displayed alongside an exhibition of contemporary arms and armour at the Royal Armouries, Leeds, as a complement to the presence and display of at least one real manuscript. As the original concept developed, it became clear that there was a niche for a more generic product, especially since some of those widely available as part of office software suites are not sufficiently configured to academic purpose. In sum: the art historian who used to show colour slide using two 35mm projectors has been succeeded by the scholar or lecturer needing a data projector with software suitable to their requirements, or by someone (for example) wanting to talk over the AG to a colleague about the contents of a particular image or set of images, in real time. Display and manipulation mode software hasn't kept pace, we would argue, with hardware developments.

Description of methods

The manuscripts were photographed on site at 500 dpi resolution. The large image files were split into sections ('tiles'). At high levels of magnification, a small number of tiles can be viewed in one screen, although at lower levels, using just this technique, more tiles would be needed. Therefore to increase efficiency at lower magnifications, the image is rescaled to half its original size, and tiled again. The rescaling is done several times to produce a collection of tiled images at different resolutions, thus allowing the viewing software to select the most efficient set of images to display based on the overall magnification level required.

Two data storage arrangements have been employed. The first uses a collection of JPEG images to tile and resize the larger image into fragments. The second encodes the tiled image and rescaling technique as a single JPEG 2000 file, thus entailing the redundancy of the tiled JPEG method.

Description of outcomes, or intended outcomes

The project has delivered the promised demonstrator, confirming the functionality of the Virtual Vellum viewer. The software has been showcased at UK e-Science All_Hands meetings, AHRC ICT seminars, AHRC AG workshops, HASTAC seminars in California, and over a live Grid with Urbana-Champaign, UCL, Lancaster, Bristol and Glasgow. A licensed version of the viewer is currently under consideration for commercial deployment. Academic researchers will have access to a copy free of charge, but will sign a license and open source software agreement.

Details of any training provided/needed

A user-friendly manual in layperson's language plus a more technical reference guide have been prepared and are now available online. The software itself will be obtainable by download from a URL hosted by the Humanities Research Institute, University of Sheffield.

Technologies

The viewing environment does not currently reside on the Grid, but an SRB client and WUN Grid nodes at Sheffield and Urbana-Champaign are being developed to ascertain the possibilities of making the viewer available over the data grid and AG. Access Grid deployment was trialled successfully on December 6th 2006 during an AG workshop forming part of an AHRC-funded series ('Access Grid for Collaborative Arts and Humanities Research').

Standards employed; data formats and protocols

The project has scoped the use of the JPEG2000 standard which permits multi-resolution tiling within a single file. This largely automates the tiling process. For the data structures, XML is used.

Tools/software used, including information on development

The software is written in Java using only the APIs available to Java version 1.2 or above. The decision to use an older version of Java than currently available was taken to ensure maximum compatibility with machines wanting to run the software without having to update their installation of Java. The software is completely contained within Java code, which includes a JPEG2000 decoder.

Flash 8 was also considered at the outset of the project, but there was not enough flexibility to facilitate the inclusion of a JPEG2000 decoder.

Services or resources used

A Gridsphere and INCIPIT portlet for hosting a VLE\VRE that hosts a copy of Virtual Vellum for viewing the Froissart manuscripts.

Closing remarks

This has been an exciting project, and the PI and Technician wish to place on record their sincere appreciation to the EPSRC e-Science programme committee for supporting the work. The demonstrator has fed into applications to the JISC VRE2 programme (outcome: highly-rated but not funded), the joint EPSRC/NSF initiative (grid and e-Science; outcome pending), the AHRC/EPSRC/JISC/e-Science programme (outcome pending) and the AHRC Resource Enhancement scheme (re-submission of an A+ unfunded project; outcome pending). Since inception, considerable interest in Virtual Vellum has been expressed, and we hope to develop the project further. Projects at Bristol and Reading have expressed their wish to use Virtual Vellum.

We should like to acknowledge here the invaluable input of the "Christine de Pisan Queen's Manuscript" project team (Glasgow, AHRC; PI Professor James Laidlaw), all of whom visited us in Sheffield for a workshop on Virtual Vellum, and for the related AG workshop later in the day on 6 December. We should also like to place on record the input of the "Shahnama Persian Books of the Kings" project (Cambridge), and of CARET, for technical and conceptual advice, and for their invitation to present to their team from the University of Cambridge (workshop, 19 October 2006).

The PI was fortunate indeed to find in Dr Mike Meredith a programmer of considerable talent, expertise and determination, with interests in VR, the grid and the potential of both. Mike has since made several innovative contributions to my own and a couple of other grid projects based at Sheffield. I hope to have the privilege of working with him on at least one continuation project!

In conclusion I would underscore the immense stimulus of working on such a challenging interdisciplinary project with a considerable knowledge transfer value involving a student of ancient medieval manuscripts, a programmer, imaging and digitisation consultant, and a group of forward-looking research libraries, not to mention grid technicians and research colleagues from several countries.

Peter Ainsworth, PI, 27 February 2007