The Robert Graves Diary (1935-39): a TEI Application using an XML Database (eXist)

Chris Petter *(cpetter@uvic.ca)*

University of Victoria

Elizabeth Grove-White *(grovewhi@uvic.ca)*

University of Victoria

Linda Roberts *(l-roberts@shaw.ca)*

University of Victoria

Spencer Rose *(srose@uvic.ca)*

University of Western Ontario

Jessica Posgate *(jposgate@uvic.ca)*

University of Victoria

Jillian Shoichet *(shoichet@uvic.ca)*

University of Victoria

O ur paper will cover both academic and technical development of the Graves Diary Project (1935-39). The prototype can be found at <http://www.tapor.uvic.ca:8080/cocoon/graves/>.

Academic Development

R obert Graves' 1935-39 diary is part of the prized Graves collection in the UVic Libraries' Special Collection. The diary has been used extensively by biographers and scholars, but it has remained inaccessible to a wider readership until recently. The Robert Graves Trust in Oxford owns the copyright to the diary and in 2001 they agreed to allow the University of Victoria Libraries to publish the diary as both an electronic and a print edition. Beryl Graves, Robert's widow, transcribed the diary into text and a copy of this text was deposited with the University of Victoria. In addition the Trust encouraged Chris Petter to scan an annotated version of the transcript, prepared by Karl Goldschmidt, Graves' long time secretary. William Graves, Robert's eldest son by Beryl, offered to contribute notes that he kept on the Deyá portions of the diary.

Robert Graves (1895-1985) is a major twentieth-century English poet, novelist and essayist. After surviving the First World War and subsequent shell shock, he married, studied at Oxford and began to publish poetry. In 1926 he met Laura Riding, the American poet whose work he had admired from afar. She became an enormous influence on him, and on his writing, and their intense working relationship lasted for over ten years. They founded the Seizin Press together, and in 1929 they moved to Deyá, Majorca. The novels that made Graves famous — *Goodbye to All That*, *I Claudius* and *Claudius the God* — were written in this period. The diary is an important document illuminating their life together and that of the little coterie of writers and artists they gathered around them.

The diary's permanent project team consists of Elizabeth Grove-White (English Department) who is responsible for the introductory material; Chris Petter (Library), project manager and Linda Roberts M.A., who is responsible for encoding, abstracting and annotation. Spencer Rose, and later Martin Holmes of UVic's Humanities Computing and Media Centre have developed the interfaces. Elizabeth was successful in landing a two year SSHRC project grant for the diary for 2004-2006. Dr. Patrick Quinn kindly contributed monthly abstracts for 1935 and 1936 diary entries.

TEI Development

Work began in 2002 when Chris Petter was granted a study leave from the Library. The manuscript was digitized and an index created which links the file title to the date. Chris traveled to the University of New Brunswick Text Centre and then to Oxford. At UNB Chris was able to restructure the text files into day entries within month divisions. In Oxford, Sebastian Rahtz advised on using the TEI.corpus dtd and an XML database to present the diary. The reason for this advice was because of the structural difficulties of the diary with its 115 enclosures and numerous letter logs. Chris also set up databases which could store information on the names, places and titles mentioned in the diary. These included the annotations of Karl Goldschmidt and the notes contributed by William Graves.

Markup

A guiding principle of the Graves diary markup procedure is to approximate the original document as closely as possible, so that the character of Graves' 'diary style' is preserved along with its content. Fortunately, XML (Extensible Markup Language), with its capacity to convey emendations such as deletions (crossed out) and supralinear additions allows us to produce an authentic version which reflects to some extent the immediacy of the diary mss. It has been necessary to work constantly with the mss in order to identify and adjust any changes made in the transcript which diverge from the copy text, including paragraphing, spelling and punctuation. Any exceptions will be accounted for in the editorial notes. The markup process allows

us to include annotations for names, places, titles, foreign words, and emendations, as well as notes and editorial comments.

Technical Implementation: Web Interface Development

Work on the web interface began in the fall of 2002, and has since become a platform for testing client-side xml processing in the rendering of XHTML documents using XSL stylesheets. Spencer Rose's contribution to this project, through the Humanities Computing and Media Centre, has involved transforming updated TEI-conformant XML documents into a simple and transparent web interface that is intuitive and useful for researchers.

The first prototype developed by Spencer Rose in 2003 made use of client-side XML processing, but was expanded to accommodate more complex XML markup. These xml processing capabilities became available with advanced web browser software. Some desirable features of client-side XML processing included the offloading of processing from the server to the client, and the direct access of XML files for customizable display. However, unlike server-side XML-to-XHTML transformations, client-side processing depends on the compatibility of the web browser to parse and render using XSL stylesheets, which, until recently, had been an unstable feature of standards-compliant browsers.

Web Prototype

The interface design involved two phases. The first phase was to build a static web display that allowed for easy browsing of the diary text. The second phase would allow users to perform complex search querying of the XML documents. For this prototype, the interface design involved a number of separate components. Of these components, some might be considered common to most web development projects such as using CSS and javascript to web-enable the site; others required special work. The static components of the site design included the general web design, XHTML layout and styling using XSL rendering and Cascading Stylesheets. The dynamic components involve using javascript for client-side interactivity. These components are brought together to form a document that is web-enabled.

XSL Templates and XHTML

The Graves Diary xml documents strictly conform to the *Text Encoding Initiative* guidelines and therefore use standard tags and attributes that describe typographic and analytic structures

of the text. Attention to detail in the XML markup was reflected in the detail of the XSL-Transformed representation such that the diary's wide range of styling features — all encoded using TEI elements — were reproduced in the transformed XHTML document. As well, the interactive features of the interface — including image scan and spot-of-reference pop-ups, as well as other dynamic display elements — were developed using client-side javascript.

XML Indexing System

The Graves Diary contains numerous enclosures — letters, poems, photographs — clippings that are components of the transcription. As with each diary entry, each enclosure has a separate digital scan that is indexed in XML documents. As well, each entry and enclosure also contains numerous biographical, geographical and bibliographical references that link to an external reference database. Because of this complex cross-indexing of media, reference information and enclosures, an important design issue was deciding on a suitable indexing system that linked these components in a coherent display.

One of the projects greatest innovations was the creation by Spencer Rose of two modular XML index files: one file cross-indexes the collection of digital image scans of the diary (including enclosures) with the main diary files; another XML file lists reference entries identified with reference locations in the diary text. Both of these external XML files originated in different file formats and needed to be transformed into XML documents. These XML files could then be included with the diary markup in the XSL templates, and as well made the creation of image and file index displays straightforward. Finally, XML pointer files for the diary entries were also used to isolate the XSL references in the document header from the actual document. This has the benefit of removing the diary files from a specific stylesheet reference.

eXist XML Database Late Breaking Development

The present phase of this interface project is to make the transcribed *Graves Diary* documents searchable online. For this, the implementation of the Open Source native XML database system eXist (<http://exist-db.org/>) has shown a promising start — with at least the proof of concept being established in a working prototype.

The eXist search engine makes use of an extended XPath query language called XQuery to search elements in a document. XPath is an established document syntax that is integral to XSL in that it defines the elements of XML documents for stylesheet transformations. eXist's enchanced querying includes basic

XPath expressions to search through the nodal structure of the XML document, but it is also capable of keyword searches on XML elements and attributes, as well as queries on the proximity of search terms and regular expressions. Analyses of nodal relationships (e.g. parent-child relations between elements) are also possible with eXist. One feature of the eXist search engine is that, for a wide range of XPath expressions, it uses stored index files that reference the structure of the XML document nodes. Information can then be retrieved without accessing the collections documents directly. This improves the speed and efficiency of information retrieval.

The Graves Diary eXist database is still in the process of development, with the rendering and placement of enclosures (some multi-page) alongside their digital images proving to be a challenge for Martin Holmes (Humanities Computing). In the meantime, the markup of the diary text and the creation of abstracts for each month by graduate students continues under the supervision of Linda Roberts. The project is scheduled for completion by July 2006.
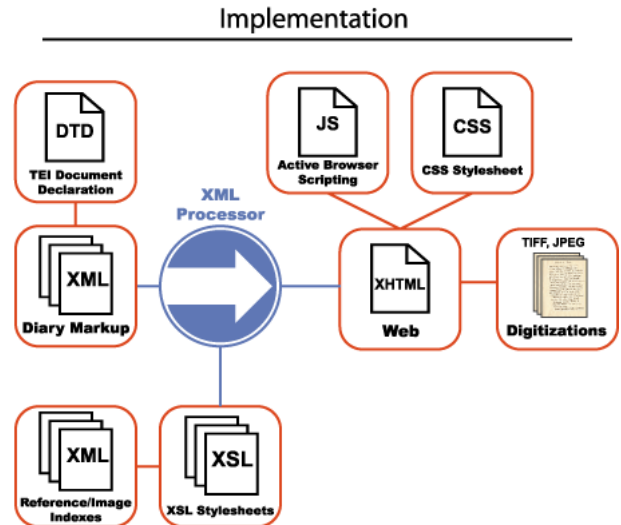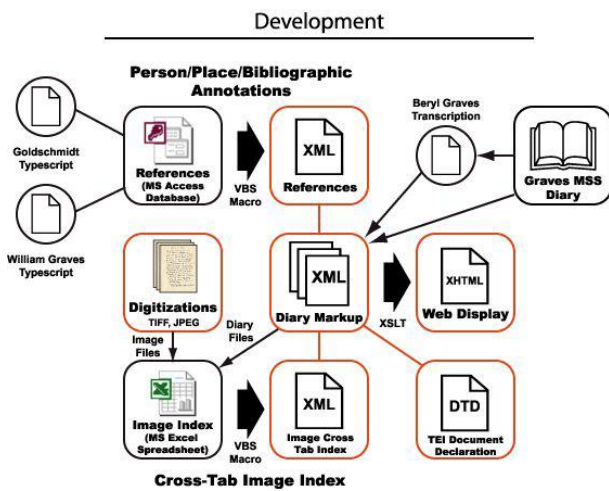

Figure 1

Figure 2

Figure 3

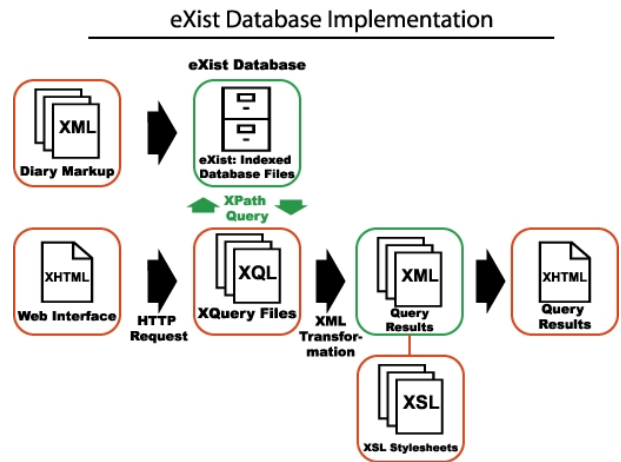# Bibliography

Meier, Wolfgang. *eXist: An Open Source Native XML Database.* . Accessed 2002-11. <http://exist-db.org/webdb.pdf>