

**NATIONAL ENDOWMENT
FOR THE HUMANITIES**

SAMPLE APPLICATION NARRATIVE



Preservation and Access Research and Development
Institution: University of Maryland

Project Description

To create the next generation of the technical infrastructure supporting image-based editions and electronic archives of humanities content, we propose through the NEH Preservation and Access: Humanities Collections and Resources program (research and development focus) a new web-based image markup tool, the **Text-Image Linking Environment (TILE)**, to be developed by the Maryland Institute for Technology in the Humanities, Indiana University Bloomington, the Royal Irish Academy, the University of Oregon, and Harvard's Center for Hellenic Studies. Despite the proliferation of image-based editions and archives, the linking of images and textual information remains a slow and frustrating process for editors and curators. **TILE**, built on the existing code of the NEH-funded AXE image tagger, will dramatically increase the ease and efficiency of this work.

At the end of two years, we will have produced software interoperable with other popular tools (including both the popular Image Markup Tool and the Edition Production and Presentation Technology suite) and capable of producing TEI-compliant XML for linking image to text and image to image with some level of automation. We will also put the image linking features of the newest version of the Text Encoding Standard (TEI P5) through its first rigorous, "real world" test, and, at the close of the project, expect to provide the TEI with a list of suggestions for improving the standard to make it more robust and effective. **TILE** will be developed and thoroughly tested with the assistance of our project partners, who represent some of today's most exciting image-based editions projects, in order to create a tool generated by the community, for the community, with the expectation that, unlike so many other tools, it will be used by the community.

Table of contents

Narrative	1
History of Grants	23
Participants	24
Budget	25
Appendices	39

Introduction

To create the next generation of the technical infrastructure supporting image-based editions and electronic archives of humanities content, we propose through the NEH Preservation and Access: Humanities Collections and Resources program (research and development focus) a new web-based image markup tool, the **Text-Image Linking Environment (TILE)**, to be developed by the Maryland Institute for Technology in the Humanities, Indiana University Bloomington, the Royal Irish Academy, the University of Oregon, and Harvard's Center for Hellenic Studies. Despite the proliferation of image-based editions and archives, the linking of images and textual information remains a slow and frustrating process for editors and curators. **TILE**, built on the existing code of the NEH-funded AXE image tagger, will dramatically increase the ease and efficiency of this work. At the end of two years, we will have produced software interoperable with other popular tools (including both the popular Image Markup Tool and the Edition Production and Presentation Technology suite) and capable of producing TEI-compliant XML for linking image to text. We will also put the image linking features of the newest version of the Text Encoding Standard (TEI P5) through its first rigorous, "real world" test, and, at the close of the project, expect to provide the TEI with a list of suggestions for improving the standard to make it more robust and effective. **TILE** will be developed and thoroughly tested with the assistance of our project partners, who represent some of today's most exciting image-based editions projects, in order to create a tool generated by the community, for the community, with the expectation that, unlike so many other tools, it will be used by the community.

Significance

The Tool

TILE will be based primarily on the Ajax XML Encoder (AXE) developed by project co-PI Douglas Reside and funded through an NEH Digital Humanities Start-up grant. During the course of this project we will extend the functionality of AXE to allow the following:

- Semi-automated creation of links between transcriptions and images of the materials from which the transcriptions were made. Using a form of optical character recognition, our software will recognize words in a page image and link them to a pre-existing textual transcription. These links can then be checked, and if need be adjusted, by a human.
- Annotation of any area of an image selected by the user with a controlled vocabulary (for example, the tool can be adjusted to allow only the annotations "damaged" or "illegible").
- Application of editorial annotations to any area of an image.
- Support linking for non-horizontal, non-rectangular areas of source images.
- Creation of links between different, non-contiguous areas of primary source images. For example:
 - captions and illustrations;
 - illustrations and textual descriptions;
 - analogous texts across different manuscripts

We are especially concerned with making our tool available for integration into many different types of project environments, and we will therefore work to make the system requirements for TILE as minimal and as generic as possible. Two projects unconnected with TILE, NINES and the Son of Suda Online (SoSOL), have already expressed interest in integrating our tool into their environments. Please see the **appendix** for their letters of support.

History of Images in the Digital Environment

Texts, from the earliest classical inscriptions to most twentieth-century correspondence, were originally inscribed on such physical objects as stones, papyrus scrolls, codex manuscripts, printed books, and handwritten and typewritten letters. As editors transfer a text from its original inscription, some of this context is necessarily obscured. Further, editors must often make potentially questionable decisions as they interpret the unclear or damaged text on the original artifact. A good editor will, of course, highlight such interventions in textual notes, but such notes, usually in small type and inconveniently separated from the main text, often go unread. The inclusion of page facsimiles can make the editorial process more transparent, but in print editions the reproduction of multiple, high quality images is often prohibitively expensive. Digital facsimile editions, on the other hand, may be distributed far less expensively, and so many editors are now choosing to publish their facsimile editions online.

The growth of the Internet as a public space in the early 1990s led to the first generation of widely-accessible scholarly electronic archives, and even at this early stage many projects integrated images into their work in significant ways. The Valley of the Shadow (1993), provided images for some of letters in the collection (in relatively low resolution), and the Rossetti, Dickinson, and William Blake Archives brought together encoded texts and images for parallel viewing and study¹. The relationship between image and texts in these archives is quite simple: for example, the page image of the source of the edited text in the *Valley of the Shadow* or the *Rossetti Archive* may be opened in a separate window (see **appendix** for screenshots), but the links go no deeper than the page level. One cannot, for instance, link from a word in the edited text to its location in the image or click an interesting area in the image to read an annotation.

At the same time that these relatively open-ended online archives were under development, other scholars were taking advantage of digital technologies to build self-contained scholarly editions. Some of the earliest efforts include the *Wife of Bath's Prologue on CDROM* (Chaucer 1996), the *Electronic Beowulf* (Kiernan 1999), and the *Piers Plowman Electronic Archive*, Vol. 1, (Langland 2000). As with the online archives, these early editions were limited in how closely they linked image and text. The *Electronic Beowulf* did provide some annotations linked to areas on the manuscript folio image, but there are few of these as the coordinates for each had to be added to the HTML "by hand."

Throughout the 1990s, several tools were developed with the intent of making this sort of text to image linking easier. BAMBI (Better Access to Manuscripts and Browsing of Images) (1994-1997) (Bozzi and Calabretto 1997, Bozzi 1999) enabled semi-automated linking between transcribed words and words on a manuscript image and permitted users to annotate the transcription. Debora (Bouche 2000) and Inote (Inote) were both designed to support the annotation of image regions. None of these tools, however, became commonly used by the scholarly community. Some appear never to have been completed or to have made it past the prototype stage; most of them seem to have faded away at the close of the projects funded to build them. The most recent version of Inote, v. 6.0, is still available for download from its site at the Institute for Advanced Technology in the Humanities at the University of Virginia, but it is built to run on Windows 95 and the download site itself has not been updated since March 21, 1998 (as of July

¹ Valley of the Shadow: Two Communities in the American Civil War, Virginia Center for Digital History, University of Virginia (<http://valley.vcdh.virginia.edu/>); The Complete Writings and Pictures of Dante Gabriel Rossetti, A Hypermedia Archive, edited by Jerome J. McGann, University of Virginia (<http://www.rossettiarchive.org/>); Dickinson Electronic Archives, edited by Martha Nell Smith, Online. Institute for Advanced Technology in the Humanities (IATH), University of Virginia (<http://www.emilydickinson.org/>); The William Blake Archive. Ed. Morris Eaves, Robert N. Essick, and Joseph Viscomi. (<http://www.blakearchive.org/>).

23, 2008). Unfortunately, this is quite typical for digital humanities tools; those that are not actively used and supported are in danger of disappearing completely.

In 2002, *Computers and the Humanities* dedicated an issue to “Image-based Humanities Computing” at a time when “a majority of first generation image-based humanities computing projects [had] reached at least an initial plateau of completion.” (Kirschenbaum 2002, p. 3) In addition to articles describing several image-based projects, this issue contained an article describing prototype software developed through the Philectre project (Lecolinet 1998) for “image-text coupling”, “focus[ing] on the editing of image oriented hypermedia that contain many document images (such as original manuscript facsimiles) and require efficient means for relating these images with textual parts.” (Lecolinet 2002) It appears, however, that this software did not make it past the prototype stage or become widely used in the humanities editing community.

Nonetheless, this issue marked a watershed moment, illustrating a recognition within the digital humanities community that images are - or should be - an integral part of scholarly digital editing. Since that time interest in incorporating primary source images into digital editions has blossomed, as has the number of image-based projects completed or under development.¹

As the community of scholars developing image-based projects has grown, more tools have been created that are actively used for project development and that have been successfully maintained past initial creation. As of July 17, 2008, the project investigators know of no fewer than ten tools or collections of tools that allow users to edit or display images within the context of textual projects or editions. These range from those that simply display an image alongside a text, to very robust software suites which support the development of complete image-based projects with substantial functionality beyond simple text-to-image mapping.

The simplest tools enable the viewing of images alongside text transcription, either for editing or for display. Juxta, developed through Networked Infrastructure for Nineteenth-century Electronic Scholarship (NINES) <<http://www.nines.org/tools/juxta.html>> provides a window for viewing image files (if provided) alongside transcriptions, which could be very useful for an editor checking readings or adding annotations, but does not provide any method for connecting the image with the text beyond the page level. Similar is the Versioning Machine, developed by Susan Schreibman at the University of Maryland Libraries (<http://v-machine.org/>): a display tool for comparing encoded texts that also enables page images to be linked to the text at the page level. These tools are both useful, but for those scholars who seek to include more fine-grained linking in their projects they are not suitable.

There are also tools that support the linking of transcribed text to images of that text. The Edition Production and Presentation Technology (EPPT), developed by Kevin Kiernan at the University of Kentucky under the aegis of the Electronic Boethius and ARCHway projects (<http://www.eppt.org/eppt/>) is a set of tools that have been developed in and run through the Eclipse software development platform. As described on the project website, the EPPT is a robust set of “tools for overlapping two images and studying and encoding their differences; analyzing and encoding script; accelerating the process of encoding variant manuscripts; collating different manuscripts; gathering and displaying statistical data; searching all structural and non-structural data; and customizing through XQuery,” including image-text linking as well as glossary development, paleographical description, codicological description, and final

¹ A number of image-based projects are described at the image tool wiki: <http://imagetool.pbwiki.com/>, although this list is self-selecting and is by no means exhaustive.

edition display. (<http://www.eppt.org/eppt/>, select "Features") For a scholar interested simply in linking text and image, then, the EPPT may prove to be too complex. In addition, the EPPT's linking process requires the creation of an "image catalog", a file separate from the TEI file which lists all image files related to the project in a unique XML format, with coordinates pointing to those files placed directly on the relevant tags in the TEI encoding. Essentially, the image-text linking output is not TEI conformant and is thus of little use to scholars who are interested in creating TEI conformant projects. It is, however, a very useful tool and thus TILE will be interoperable with this non-standard output. See **Standards** section for a discussion of the EPPT linking method.

There have also been some efforts to build tools to automate the creation of links between transcribed text and image of that text. One such tool was prototyped by Cheng Jiun Yuan as part of his PhD work at the University of Kentucky (Yuan 2003). It is described in chapter 5 of his dissertation as "correspondence linking," and used intelligent character recognition (ICR, a specific type of optical character recognition designed specifically for recognizing handwritten characters) to match plain text with the image of that text. Although this software could theoretically be trained to recognize any type of script, it did not have a simple front-end suitable for use by a humanities scholar and (as with so many other tools) did not develop beyond the prototype. TILE will include semi-automated text-image linking, provided by the incorporation of the Word Linking tool, originally developed by TILE co-PI Reside for the Shakespeare Quartos project, and described in further detail below (under **Our prototype tools**).

Perhaps the most notable existing tool is the Image Markup Tool (IMT), under development by Martin Holmes at the University of Victoria, BC. IMT is the first tool to output complete and valid TEI P5 XML, and has, as a result, earned widespread acceptance by the TEI community. Projects actively using the IMT for development include the Pembroke 25 Project (<http://www.rch.uky.edu/Pembroke25>) and eMunch: An electronic archive of Edvard Munch's written material (http://emunch.no/index_eng.htm). The IMT enables a user to place a series of annotations on an image, resulting in a file that validates against the regular (unmodified) TEI P5 schema, and then enables the user to create HTML for the display of those annotations online. The IMT is very simple and easy to use, and is in many ways a model of the type of tool that we will be developing in this project - it does one thing, and it does it very well. Unfortunately, the IMT runs only on Windows machines and cannot be easily ported into new web-based projects. TILE will interoperate with the constrained IMT TEI format, which is described in more detail in the **Standards** section.

It is also worth mentioning three projects that provide tools for end users rather than for editors and project developers. Networked Infrastructure for Nineteenth-century Electronic Scholarship (NINES), mentioned above in connection with Juxta, is powered through its online presence by Collex. Through Collex, scholars can search across several different projects and view and compare edited texts and images from them. It does not, however, provide a way for scholars to create links between text and image. The NINES editors have expressed interest in incorporating TILE tools into their system for end users, not for editors (Collex is a viewing system, not an editing system), and we will work with them to ensure the tools will be easily portable. Please see **appendix** for letter of support. Similar projects include MIT's Simile (<http://simile.mit.edu/>), which provides "robust, open source tools that empower users to access, manage, visualize and reuse digital assets" - although no tools as of yet that enable the creation of links between image and texts, and George Mason's Omeka, an open source system designed specifically for the online publication of collections and exhibitions.

Our prototype tools

The Ajax XML Encoder (AXE), developed by co-PI Reside at MITH, allows users with limited technical knowledge to add metadata to text, image, video, and audio files (see **appendix** for screenshots). Users

can collaboratively tag a text in TEI, associate XML with time stamps in video or audio files, and mark off regions of an image to be linked to external metadata. Now, at the end of the startup period, we believe the image and text tools in the suite are both the most potentially useful, but that they also require the most additional development. At present the web-based image tagger allows users to select regions in an image and store the coordinates of this region in a database, but it does not provide tools to make use of this data once it is stored. The text tagger allows a user to specify a relaxNG schema and then tag a text using this schema, but it requires users to enter coordinates for image links by hand (it does not, at present, interface easily with the image tagger). The tools in AXE were always intended to be interoperable and to have the functionality described in this narrative, and this current collaboration allows us to build on the groundwork that has been laid in the startup period to move the suite to the next stage of its development and to thoroughly test it on important, multi-institutional projects.

As part of his work on the NEH and JISC-funded Shakespeare Quartos project, project investigator Doug Reside has built a small, web-based tool that automatically links the transcribed text of the quartos to page images with about 70% accuracy. This tool thus far has only been tested on early printed editions; our project will refine this tool to work with handwritten manuscripts, and will standardize the backend to generate the same XML-output as the other tools.

Weaknesses of Existing Tools

We see two major weaknesses with existing tools. The first is perhaps the most obvious: all of the existing tools (with the exception of Reside's Word Linking tool) are manual. They depend on a human user to create the relationships between the image and the text. The PIs of this project have used these tools in their own work and know first-hand how time consuming and tiring the manual tagging can be. Those fortunate enough can take advantage of student assistance, but that work must still be checked by scholars. Reside's tool is a proof of concept for the automated linking of words and lines of text in a transcription to lines of text in image files, and a primary goal of TILE will be extending its functionality and increasing the accuracy.

The second major weakness of current tools is that they are not, in general, interoperable. Individually, each of these tools allow a user to do interesting things. A scholar should theoretically be able to create a simple TEI transcription in an XML editor such as oXygen, bring those into the Word Linking tool for automated linking of the words, port those files into the EPPT for more fine-grained editing (for example, marking all scribal additions, deletions, and marginal notes), then port them into IMT to add some editorial annotations, and then either output the HTML from the IMT or import them into Juxta or the Versioning Machine for display. However, at present that type of interoperability is not possible. Each tool has its own unique output (not all TEI-conformant).

Our proposed tool suite will not only be internally interoperable (that is, the individual components of TILE will work well together), but will also enable the interoperation of existing tools by providing a system for converting between the outputs of frequently used tools (at least the IMT and the EPPT). It is important to remember that each of these tools has a community, however small, of existing users, and those users may wish to change to something new. We seek not to replace existing tools so much as to create a system that enables them to work well together. If one scholar is building a project using the EPPT and another scholar is building a related project in the IMT, our tools would enable them to bring their projects together and convert them to valid, well-formed and standard TEI P5.

TILE: A National Priority for the Humanities

The outcomes of TILE are all of high priority for the future of digital projects development and the humanities in general. Interest in building image-based projects will only grow as more museums and libraries implement digitization projects and more high-resolution digital images become available. Scholars will need simple, effective tools that enable the efficient creation of links and metadata in standard formats, and that also support the interoperability of tools and files. The TEI community, notably the Text and Graphics SIG, provides a ready audience for the project. We also recognize, though, that if our tool is useful only for those already entrenched in the TEI or Digital Humanities communities, they really aren't that useful at all, and we will therefore make special efforts to make this tool available to the less technically inclined in our respective fields through presentations at subject conferences (see **Dissemination** for details).

This project has the potential to help change not just digital editing, but the way software in the humanities is developed and considered. Many tools created for humanists are built within the context of a single project, focusing either on a single set of materials or on materials from a single time period, and this limits their ability to be adapted for use by other projects. By design, our project cuts across subjects and materials. Because it will be simple, with focused functionality, our tool will be usable by a wide variety of scholars from different areas and working with a variety of materials - illustrations and photographs as well as images of text. Therefore we have brought together several collaborators from different projects with different needs to provide advice and testing for our work.

History, Scope, & Duration

History of the current project

The earliest developments of the current project can be traced back to the Digital Atheneum, a project funded 1999-2003 through the National Science Foundation's Digital Library Program (<http://www.eppt.org/digitalatheneum/>). This project, directed by W. Brent Seales and Kevin Kiernan at the University of Kentucky, developed a suite of software tools for the creation of image-based editions. This software suite was further developed and tested under the ARCHway project (Architecture for Research in Computing for the Humanities through collaborative research, teaching and learning; <http://beowulf.engl.uky.edu/~kiernan/ARCHway/entrance.htm>), funded 2003-2005 by the National Science Foundation, and the Electronic Boethius project, funded 2002-2006 by the National Endowment for the Humanities and the Mellon Foundation. Both of these projects were directed by Kevin Kiernan at the University of Kentucky. The current project co-PI Porter was a co-PI on the Electronic Boethius 2003-2005 and was the project coordinator for the ARCHway project. Co-PI Reside was a software developer during his tenure as a graduate student at the University of Kentucky, 2004-2005. Both Reside and Porter contributed to the development and testing of the software suite, which was then called the Edition Production Technology (EPT), and has since 2005 been further developed by Kiernan and Ionut Emil Iacob, and is now called the Edition Production and Presentation Technology (EPPT; <http://www.eppt.org/eppt/>). TILE is not affiliated with the EPPT or its current developers.

After several conversations at the 2008 Digital Humanities conference in Oulu, Finland, the PIs came to the realization that each needed a better tool for image markup. Porter, whose work with existing tools on several image-based projects had convinced her of the need, not just for a new tool, but a new, collaborative model for tool development, conceived the grant and presented it to Walsh and Reside. Walsh, whose work has included several image-based editions as well as the Comic Book Markup Language, strongly agreed with Porter's assessment of the need for a new tool and, given his experience with a number of image-based projects with various needs, seemed the ideal candidate to direct its testing. Both were impressed with what they had seen of Reside's automatic word tagger and AXE tool, and hoped to include his work in the tool suite. Reside, whose Digital Humanities Startup Grant was

drawing to a close, was keen to move AXE beyond the prototype stage and into a period of final development and beta testing, and so readily agreed to join the project. Information about AXE can be found at the MITH research page: <http://mith.info/mithresearch/>

We expect the tool will take two years to complete and will be maintained beyond the funding period as described in the **Sustainability** section of this document.

Methodology and Standards

Methodology

A new model for tools development

TILE represents an innovative approach to tools development within the context of the Preservation and Access: Humanities Collections and Resources grant program. In fact, our project may be described as a *metaproject* rather than as a project in and of itself: rather than creating collections, we are building resources that will enable members of the scholarly community to build and strengthen their own collections. Directors of image-based projects will be involved with the design of the tool at the start and will contribute recommendations for refining the tool during an intensive testing period during the second year of the project (see the **Work Plan** for details).

Technical methodology

TILE will be built primarily in JavaScript and will therefore be capable of running almost entirely on the client machine without need of a network connection. A few functions, however, will require some server side scripting. The most crucial of these is data storage and management. Standards compliant JavaScript, for security reasons, is not permitted to write files to the client machine, and it would therefore be impossible for users to save their work automatically. For this reason, we will also provide PHP code that will store this data in an XML file in several formats including (at least): XML compatible with the IMT, XML compatible with the EPPT, and a more general form of TEI P5. Along with our JavaScript library we will therefore provide an API for programmers, allowing them to write their own server side scripts to process and store the data.

The automatic word to image linking requires image processing functions currently unavailable in standards compliant JavaScript. We will build this code in PHP using the GD library and host it on servers at partner institutions. We will publish the API for processing image and text through this program, and will provide the PHP code for download.

In addition to running as a stand alone application, TILE will also include an interface-independent JavaScript library able to be ported into other web-based projects. Developers of two web-based environments have already expressed interest (Son of Suda Online, a web-based editor, and NINES, described above; see **appendix** for letters of support) and we expect that others will follow suit as the tool is proven to be easily portable and useful for scholars.

Developing Standards

Despite the evolution and expansion of digital imaging and editing technologies and the steady growth in the number of projects under development, the lack of a single standard for describing relationships between texts and images in machine-readable encoding has led to a situation where editors and tool

developers have been forced to develop their own ad hoc methods for creating links. The situation, however, is improving. The Text Encoding Initiative Consortium is responsible for developing and promulgating the TEI Guidelines, the international standard for humanities text encoding (TEI Consortium 2007). The TEI recognized that this proliferation of methods was a problem for the interoperability of data and represented a significant hole in the recommendations, and it also saw that this issue was of significant interest to the TEI and digital humanities communities. The development of specific recommendations for image-text linking was made a priority during the final two years of development of the most recent version of the TEI Guidelines, TEI P5. Work on the recommendations took place in 2006-2007 through the TEI Technical Council, and this work was spearheaded by TILE co-PI Porter and Conal Tuohy from the New Zealand Electronic Text Center (TILE co-PI Walsh was also on the Council during 2006-2007 and contributed to the recommendation).

A new TEI Special Interest Group on "Text and Image" was co-convened by Porter and Walsh soon after the release of P5. The SIG serves to bring together members of the community developing image-based projects to discuss, develop, and document various strategies and best practices for implementing the Digital Facsimile tags available in TEI P5, especially as relates to graphically rich texts, and for developing methods to extend the TEI when existing structures are not suitable. After the project partners, the SIG community is one of the main initial audiences of the current project and the tool will be released to them for testing after initial development is complete (in May 2010).

With the release of TEI P5 in November 2007 there is now a standard for specific methods for building links between encoded text and image files, both at the page level and more finely-grained (between sections of text and areas on an image). Technical details describing how the standard works can be found in the **Standards** section below. Practically, the TEI recommendation gives tool developers and scholars a single method for expressing relationships between images and texts. Extra-textual annotations, links between transcription and image, and links within images can now be expressed using the same syntax. A standard is emerging, then, but tools that comply to that standard are necessary if it is to be truly useful. In turn, tool developers no longer have to create their own methods - they can take advantage of the standard, and the standard will in turn be strengthened by the use provided by tools.

In addition to standard general TEI output we will also enable multiple different formats of XML output as well as create an XSLT-based system whereby any user of our tool can specify his or her own output by including a stylesheet that describes the desired output. One popular standard in the digital library community is the Metadata Encoding and Transmission Standard (METS), which provides a wrapper for combining pointers to external files (including TEI files and images) with many different types of metadata into a single file. METS provides a method for linking areas of image files with any type of metadata included in the file (including, potentially, TEI), and providing output in METS which will extend the usability of TILE to the digital libraries community.

Accounting for two different types of output in the first year, TEI and METS, will give us the opportunity to develop a method for enabling output in almost any XML format, given an XSLT to provide the mapping from TEI to the other format. This functionality is absolutely vital to creating the type of interoperability TILE seeks to support. During the second year we will work with the project partners to expand the output system to allow for a range of XML output including files conformant with the EPPT methods of text-image linking, the IMT-specific TEI format, the Comic Book Markup Language, and EpiDoc (TEI XML for Epigraphic Description, which is used by the Integrating Digital Papyrology project; see **appendix** for letters of support from Joshua Sosin and Gabriel Bodard).

We will also take advantage of the <appInfo> tag in TEI P5 which is designed to contain a log of applications that have been used to edit the TEI file (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HDAPP>). When a tool opens a file, by referring to the information contained in

<appInfo> it can determine the last piece of software used to edit the XML, and then make whatever modifications are needed for it to be used by the current tool. This will be very useful when files created in TILE are exported to other tools and vice versa.

Evaluation

Our development process and TILE itself represent serious innovation in the practice of tools development for the humanities, and we have arranged for both internal and external evaluation to measure our success. Partners at the University of Oregon, Harvard's Center for Hellenic Studies, Indiana University Bloomington and the Royal Irish Academy will work closely with the PIs during the first year as we develop the tool, and will test the tool during the second year and provide suggestions for further refinement. We are also fortunate to have as a volunteer adviser Martin Holmes at the University of Victoria, the developer of the Image Markup Tool, who has agreed to work with us to ensure the interoperability between our tool and the IMT and to provide general feedback throughout the project.

We have also arranged for an external evaluator, Melissa Terras, senior Lecturer in Electronic Communication at University College London School of Library, Archive and Information Studies, who will attend annual meetings with the development team and project partners, contribute to the project wiki and documentation, and will provide recommendations during the development and testing process. Terras will also compose independent reports to be included in the annual and final reports to the NEH. With extensive experience both in software development and in usability testing, Terras is a valuable member of our team, and her knowledge and experience will be invaluable to us as we move forward in the development of TILE.

Project Partners

The project partners have been selected purposefully from amongst our larger group of collaborators to represent the many different key needs of image-based projects. Such a varied group of initial partners will enable us to thoroughly test the usability of the tool. During the course of the grant period, the project partners will meet twice at MITH to discuss their experiences using TILE on their own work. The projects involved are described below and letters of participation are included in the appendices.

Mapas Project

The Mapas Project, part of the larger Wired Humanities Project at the University of Oregon, directed by Stephanie Wood and Judith Musick, has as its goal the digitization and close study of textual and pictorial Mesoamerican manuscripts dating from the sixteenth through nineteenth centuries. The project, underwritten by a Collaborative Research Grant of the National Endowment for the Humanities, seeks to represent the manuscript images with transcriptions, translations, and analysis of their texts. The goal is to achieve a better understanding of the histories of indigenous communities of Mesoamerica under foreign domination. As of Summer 2008, the project is just completing two years of intensive work on four manuscripts in the Kislak Collection at the Library of Congress.

While these manuscripts show European stylistic influences, owing to the Spanish colonial and the neo-colonial context in which most of them were produced, they also contain artistic styles, compositional elements, and content from pre-Columbian, non-Western traditions. The manuscripts contain both textual and illustrative material. In many cases the texts serve as descriptions of illustration - that is, the text on the page is actually linked to "image" (illustration) on the page. This requires a different sort of linking than that required by more textually-focused projects--linking not just text to image, but also image to image.



FIGURE 1: Mapa de Tolcayuca

The Mapas Project has found it a challenge to encode image and text in a way that captures their relationships adequately. The Mapas themselves contain a large variety of text types, and each one may not relate directly to image types on the same page. Some textual expressions might be similar to a gloss or caption, but there are also descriptions and narratives. For instance, some manuscripts include speech scrolls emerging from people's mouths (somewhat reminiscent of bubbles in comic strips).

The Mapas Project has several needs that TILE will be designed to address, including linking areas of images to other areas of images (and on to descriptive metadata or encoded text) and text that follows non-horizontal, non-rectangular boundary areas. Its inclusion as a test project will ensure that TILE can support some of the most advanced image-text linking requirements.

Homer Multitext

The Homer Multitext Project (http://zeus.chsdc.org/chs/homer_multitext), a project out of Harvard's Center for Hellenic Studies, edited by Casey Dué (University of Houston) and Mary Ebbott (College of the Holy Cross) with technical leadership provided by Christopher Blackwell (Furman University) and Neel Smith (College of the Holy Cross), seeks to present the textual transmission of the *Iliad* and *Odyssey* in a historical framework. Such a framework is needed to account for the full reality of a complex medium of oral performance that underwent many changes over a long period of time. These changes, as reflected in the many texts of Homer, need to be understood in their many different historical contexts.

The Multitext has published approximately 2200 digital images of three manuscripts of the *Iliad* housed in the Marciana Library in Venice, Italy: the tenth-century Marcianus Graecus Z. 454 (= 822), also known as "Venetus A", the eleventh-century Marcianus Graecus Z. 453 (= 821), and the twelfth/thirteenth-century Marcianus Graecus Z. 458 (= 841) (Dué et al. 2007). These have been published through the Multitext Manuscript Browser, and a small image of one of these (folio 100 verso of the Venetus A) appears below:

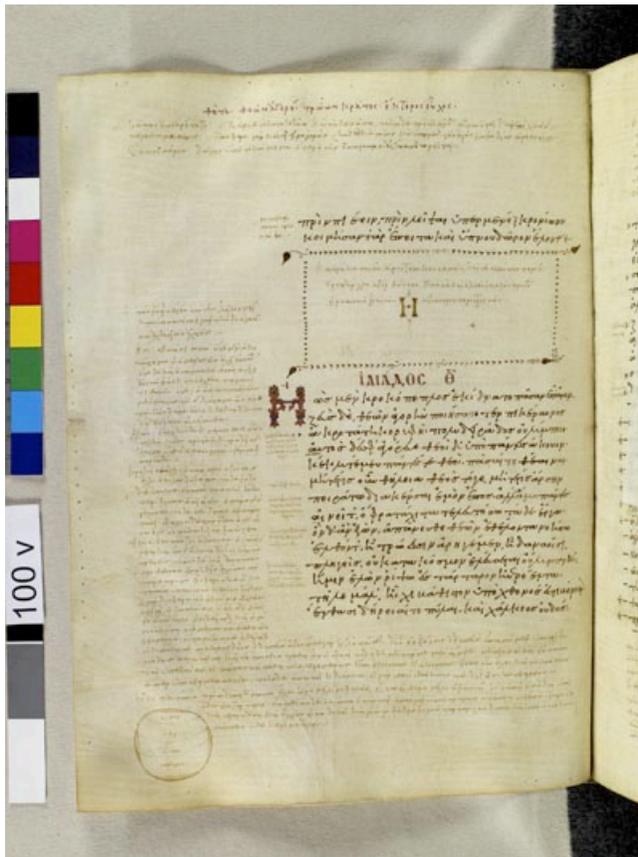


FIGURE 2: Marcianus Graecus Z. 454 (= 822) (Venetus A), folio 100v.

These manuscripts are significant not only because of their age, but because of the tens of thousands of scholarly notes, the "scholia", included in their margins. Each folio of the Venetus A includes, in addition to the Homeric Text, at least three discrete scholiastic texts, which appear as the main marginal scholia, the inner-marginal scholia, the intra-marginal scholia, and the interlinear scholia.

The HMT has produced TEI conformant transcriptions of both the Iliadic texts of these manuscripts and of the scholia. The latter, however, is based on published editions, which conflate the various scholiastic traditions, "harmonize" individual scholia which may appear differently from one manuscript to another, and add to, replace, or delete the lemmata--the quotations of Homer that begin each note. A full publication of this material requires a collaborative effort to inspect the manuscript images, edit the scholia to restore the transcriptions to an accurate state, and to translate the scholia (for the first time). The project requires a tool able to create links that associate the scholia in the manuscript images with individual passages of the scholiastic texts. The HMT could benefit from the semi-automated text-image linking provided by TILE, as well as from the image-image linking (to associate scholia with main text) also required by the Mapas project.

Indiana University Bloomington

Indiana University is home to a number of important and innovative digital humanities and digital library projects. Co-PI Walsh directs three encoded text and image projects that provide distinct use case scenarios for TILE and together comprise our partner projects at Indiana University Bloomington. Indiana is supported by a leading IT infrastructure and relevant development expertise in the Digital Library Program and University Libraries, the newly created Institute for Digital Arts and Humanities, where Walsh is a faculty fellow, and faculty research programs in the Schools of Library and Information Science, Informatics, and other departments and programs.

Chymistry of Isaac Newton

Isaac Newton wrote and transcribed about a million words on the subject of alchemy, or chymistry¹, of which only a tiny fraction has today been published. Newton's alchemical manuscripts are graphically and visually rich, full of alchemical symbols; Newton's own illustrations; difficult scribal, editorial, and representational challenges; and complex graphic organizational schemes. TILE will assist the Newton project by providing simple methods for linking encoded transcriptions with text on the facsimile page images of Newton's chymical manuscripts, for highlighting and annotating figures and other illustrative materials in the documents, and for linking these figures and illustrations with Newton's annotations, editorial annotations, and user-contributed tagging and annotation.

Swinburne Project

The Swinburne Project (<http://swinburneproject.indiana.edu>) is a digital collection, or virtual archive, devoted to the life and work of Victorian poet Algernon Charles Swinburne. When complete the project will provide students and scholars with access to all available original works by Swinburne and selected contextual materials, including contemporary critical reactions, biographical works, and images of artwork about which Swinburne wrote.

The Swinburne Project project is entering a new phase of development that will include the addition of page images for the published print texts; transcribed and edited text and facsimile page images for manuscript material, and digital images for the many visual art pieces about which Swinburne wrote.

¹ As William R. Newman and Lawrence M. Principe have argued in several co-authored publications, it is anachronistic to distinguish "alchemy" from "chemistry" in the seventeenth century (Newman and Principe, 1998; Principe and Newman, 2001)

TILE will be useful to the Swinburne Project for general interaction with and annotation of the facsimile page images for the manuscript and print materials and will allow more novel opportunities to annotate the related visual artworks, and for linking between Swinburne's commentary and the art works.

Comic Book Markup Language

CBML, or Comic Book Markup Language, is a TEI extension for encoding comic books and graphic novels. In addition to providing support for encoding these rich and diverse media types, CBML seeks to solve general encoding, metadata, and representation problems for other media types with tightly integrated text and graphics, e.g., illuminated manuscripts, illustrated children's books, art books, liner notes and related artwork for audio recordings, etc.

Encoding and analysis of comics and graphic novels demands working with full-page images of comics pages, or two-page spreads, that maintain the compositional integrity of the comic page. But further subdivision of the page into individual sequential panels is also necessary. Ideally, the CBML/XML-encoded transcriptions include references to pixel coordinates and other metadata for relevant sections of the facsimile page images. A fully-developed and configured image tool would allow researchers to load images into the tool, graphically segment the image into various components, and add transcriptions, metadata, commentary, annotation, and markup to the components and the whole. TILE will serve as a editorial and development environment for the creation of CBML editions.

Together, the *Chymistry of Isaac Newton*, *The Swinburne Project*, and the *CBML* project will provide a rich and diverse platform for testing, evaluating, and improving TILE.

Digital Humanities Observatory, Royal Irish Academy

The Digital Humanities Observatory (DHO) is the centerpiece of the the Humanities Serving Irish Society (HSIS) initiative, a partnership of equals committed to developing an inter-institutional research infrastructure for the humanities in Ireland. Consisting of the Royal Irish Academy, six of the seven Irish universities, Queen's University, Belfast and the University of Ulster, the HSIS consortium is building a joint national platform for the coordination and dissemination of humanities research, teaching and training at an all-island level. As it supports digital humanities projects throughout Ireland, TILE has a number of potential testbeds available through the DHO's collaboration. The three projects described here provide a glimpse of the types of materials and requirements that will be furnished for testing.

St Patrick's Confessio (Royal Irish Academy)

The Latin writings of St Patrick are of crucial importance for Irish history and ecclesiastical culture. The Royal Irish Academy's Dictionary of Medieval Latin from Celtic Sources (DMLCS) project has been granted funding to construct an online archive that will present different aspects of St Patrick's work at various levels, closely interlinked passage by passage (<http://www.dho.ie/confessio.html>).

Of relevance to TILE is the centrality of linking for the *Confessio* project. The full-text of the original Latin text is already housed in the Archive of Celtic-Latin Literature database (ACLL), and the *Confessio* project will link each passage from the Latin text to many other realizations of the same passage, including a diplomatic edition of the ninth-century manuscript, a scholarly edition of that manuscript, the *apparatus criticus* from that text (showing variant readings from other manuscripts), an analysis of the structure of the work, several translations, and a glossary of the more interesting words found in the *Confessio*. In addition to these textual links, each passage will also be linked to a facsimile of the same manuscript transcribed and described in the other resources.

TILE has the potential to be immensely useful for the *Confessio* project as they begin building the links between passages in the transcription and the facsimile. The *Confessio* will also test our method for describing non-rectangular areas to align with the textual passages.

1641 Depositions Project (Trinity College Dublin)

The *1641 Depositions Project* (<http://www.tcd.ie/history/1641/>) aims to transcribe and digitize 3,400 depositions, examinations and associated materials located in the Library of Trinity College Dublin that describe the first-hand experiences of Protestants following the Catholic Irish rebellion in 1641. The project seeks to fully transcribe and digitize the collection, which totals 19,000 pages, and to make the collection available online. TILE would enable the *1631 Depositions Project* to add scholarly annotations and fine-grained links between the text and images of the documents.

Alcala Ledger project (National University of Ireland, Maynooth)

The *Alcala Ledger Project* will create a digital version of an 18th century ledger consisting of 324 pages of income and expenditure from the Royal Irish College of Alcala in Spain. The digital version will include the original Spanish text as well as an English translation and images of the ledger. The project already has plans to segment the pages into areas that correspond with the ledger entries and to link those areas with the original text and translation, although they have not yet determined the standards or methods to be used. The *Alcala Ledger Project* will give us the opportunity to test TILE's interoperability with a project that may use standards and methods unlike those for which TILE was built.

Standards

The standard underlying TILE is the Text Encoding Initiative P5 (TEI) Guidelines, specifically the Digital Facsimile family of tags. Digital Facsimile is new as of November 2007, and is described in chapter 11 of the Guidelines, "Representation of Primary Sources" (TEI Consortium 2007).

Describing a Digital Facsimile

Digital Facsimile defines several new elements and attributes that enable the inclusion of digital images in TEI files. <facsimile> "contains a representation of some written source in the form of a set of images rather than as transcribed or encoded text." It does this by holding a series of <graphic> tags which in turn point to image files, which may be housed on a local machine or on a remote server. For example, a facsimile that points to low-quality JPEGs of folios 12r-13v of the Venetus A manuscript (from the Homer Multitext project) would look like this:

```
<facsimile>
  <graphic xml:id="id12r" url="http://chs75.harvard.edu/image_archive/JPEG_Low/VA012RN-0013.jpg"/>
  <graphic xml:id="id12v" url="http://chs75.harvard.edu/image_archive/JPEG_Low/VA012VN-0514.jpg"/>
  <graphic xml:id="id13r" url="http://chs75.harvard.edu/image_archive/JPEG_Low/VA013RN-0014.jpg"/>
  <graphic xml:id="id13v" url="http://chs75.harvard.edu/image_archive/JPEG_Low/VA013VN-0515.jpg"/>
</facsimile>
```

FIGURE 3: Basic Facsimile markup

The contents of facsimile are understood to be the complete facsimile representation of the physical object, and are to be read in sequence.

If there are multiple image files associated with a single object view (such as a folio or page), multiple

graphics may be provided, each one pointing to a different image file. This is done by using the <surface> tag, which indicates that each image file contained therein represents the same object surface. In the case of the Homer Multitext, three different image files of varying resolution are provided for each folio, but they were all processed from the same master image file and so represent the same exact surface area.

```
<facsimile>
  <surface xml:id="id12rHML">
    <graphic url="http://chs75.harvard.edu/image_archive/JPEG_High/VA012RN-0013.jpg"/>
    <graphic url="http://chs75.harvard.edu/image_archive/JPEG_Medium/VA012RN-0013.jpg"/>
    <graphic url="http://chs75.harvard.edu/image_archive/JPEG_Low/VA012RN-0013.jpg"/>
  </surface>
  <surface xml:id="id12vHML">
    <graphic url="http://chs75.harvard.edu/image_archive/JPEG_High/VA012VN-0514.jpg"/>
    <graphic url="http://chs75.harvard.edu/image_archive/JPEG_Medium/VA012VN-0514.jpg"/>
    <graphic url="http://chs75.harvard.edu/image_archive/JPEG_Low/VA012VN-0514.jpg"/>
  </surface>
</facsimile>
```

FIGURE 4: Multiple images of the same surface

To create links between encoded texts and the Digital Facsimile, TEI uses the @facs attribute. This is a global attribute, which means that it can be placed on any element in the entire TEI tagset. In the most simple case, @facs can point from a tag directly to an image file. This example points from the page break marking the start of Venetus A folio 12r to the low quality JPEG image of that folio located on the HMT server:

```
<pb n="12r" facs="http://chs75.harvard.edu/image_archive/JPEG_Low/VA012RN-0013.jpg"/>
```

@facs can also point from a tag into a <facsimile>. In this example, the same page break points not directly to the file on the server, but to the <graphic> in the <facsimile> in Figure 3 - the value of @facs points to the @xml:id for the <graphic> that describes 12r:

```
<pb n="12r" facs="#id12r"/>
```

In this example, we want to make sure that our page break points to all of the images that we have for folio 12r, so the value of @facs points to the @xml:id for the <graphic> in Figure 4:

```
<pb n="12r" facs="#id12rHML"/>
```

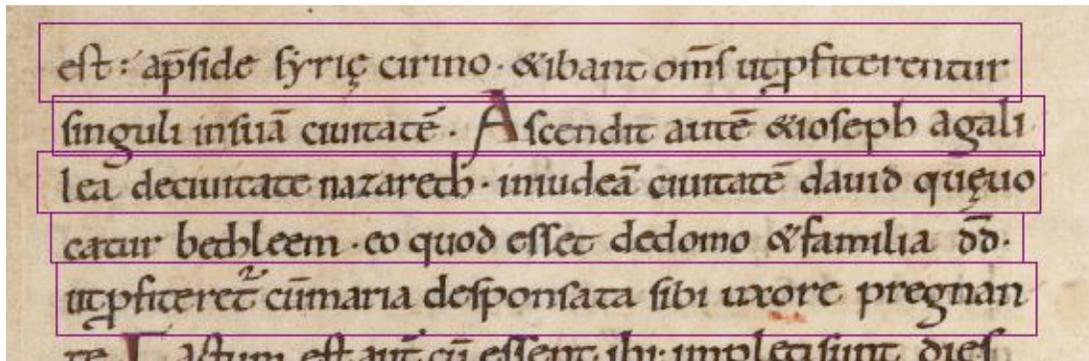
Linking Text to Image in TEI P5

TILE will provide methods for quickly and simply creating <facsimile> groups, and for building these simple links between surface images and high-level elements (<pb/> for manuscript folios or print books, or <div> for other materials such as letters). However, organizing the surface images is just the start of creating an integrated image-based project. True integration requires granular linking: transcribed text to image of text, illustration to a description of the illustration, and marginal note to the line of text described by the note - a link that exists image to image, text to text, and image to text. The Digital Facsimiles tagset provides a single method to satisfy all three of these concerns.

Granular linking in TEI is made by use of <zone> elements that are used to specify areas of the <surface>. Coordinates defining the four corners of the zone are provided in four attributes, as defined in the TEI Guidelines:

- @ulx gives the x coordinate value for the upper left corner of a rectangular space.
- @uly gives the y coordinate value for the upper left corner of a rectangular space.
- @lrx gives the x coordinate value for the lower right corner of a rectangular space.
- @lry gives the y coordinate value for the lower right corner of a rectangular space.

In the following example, zones are defined corresponding to the first five lines of Cambridge Pembroke College, MS 25, folio 12r:



```
<facsimile>
  <!--surfaces for fols. 1r-11v-->
  <surface xml:id="id12r"><!--surface for fol. 12r-->
    <graphic url="..\images\smaller\012r.jpg" width="1384px" height="2016px"/>
    <zone xml:id="id0" ulx="102" uly="184" lrx="993" lry="247"/>
    <zone xml:id="id1" ulx="90" uly="234" lrx="1011" lry="314"/>
    <zone xml:id="id2" ulx="91" uly="292" lrx="1038" lry="356"/>
    <zone xml:id="id3" ulx="104" uly="339" lrx="1017" lry="414"/>
    <zone xml:id="id4" ulx="106" uly="393" lrx="1015" lry="457"/>
  <!--continue 12r-->
</surface>
<!--surfaces through the end of the manuscript-->
</facsimile>
```

Elsewhere in the document is the transcription of these lines, and we can link each line to the corresponding zone by using the @facs attribute, as well as linking the page break to the surface (which, as we have seen, corresponds to the entire folio):

```
<pb n="12r" facs="#id12r"/>
<lb n="1" facs="#id0"/> est ; a preside syrie cirino . et ibant omnes ut profiterentur
<lb n="2" facs="#id1"/> singuli in suam ciuitatem . Ascendit autem et ioseph a gali
<lb n="3" facs="#id2"/> lea de ciuitate nazareth . in iudeam ciuitatem dauid que uo
<lb n="4" facs="#id3"/> catur bethleem . eo quod esset de domo et familia dauid .
<lb n="5" facs="#id4"/> ut profiteretur cum maria desponsata sibi uxore praegnan
```

Because the @facs attribute is global and can thus be placed on *any tag in the TEI*, this same method can be used to link any element in a text edition to a Digital Facsimile, or to elements within a Digital Facsimile. This could include linking a caption to an illustration or figure (or to zones within an illustration), linking an annotation to any zone on a surface, or linking a single textual element to multiple zones (@facs can contain pointers to multiple zones). TILE will give scholars a way to take advantage of the Digital Facsimiles markup by creating all these different types of links much more quickly and easily

than is currently possible.

As the TEI Digital Facsimiles recommendations are still in their infancy, it can be expected that there are limitations to its usability. For example, currently it is only possible to create horizontal, rectangular bounding boxes. If one needs to mark an area that is anything other than a rectangle he or she must use some other system, such as Scalable Vector Graphics (SVG), to encode this information. Also, if one is dealing with materials that have text or illustrations at angles other than straight horizontal (vertical, upside-down, or anywhere in between - such as the South American mapas) one would have to modify the TEI in an ad hoc way to express the angle in the encoding. During our development and especially during the testing process, we expect to come across these and other issues, and to incorporate solutions into the software as well as into our output. We will also work with the TEI Text and Graphics SIG and the TEI Technical Council to improve and extend the official TEI P5 Guidelines for image-based editions.

Other Methods to be Supported

In addition to general TEI digital facsimiles code, by the end of the second year our tools will output five other encoding structures, including both extensions of TEI and alternative methods, for image-text linking, and will also include a documented method, based on XSLT stylesheets, that users of the tool can use to create any XML-based markup. See **appendix** for examples of some of the methods described below.

Metadata Encoding and Transmission Standard (METS)

While TEI is designed specifically to encode texts, METS is a metadata wrapper, widely used in the digital libraries community to combine different metadata schemas and files, including potentially TEI and image files, into a single document. The most important section of a METS file is the Structural Map, which is where the various materials included in a METS project are organized. Within the structural map, users can organize links to image files and create links to areas of those image files using the attribute COORDS (coordinates). Enabling the creation of METS files through TILE will serve to make these tools more attractive for use by the digital libraries community.

Edition Production and Presentation Technology (EPPT)

The XML format used by the EPPT was created during the tool's initial development period 2003-2005, long before the release of TEI P5 and the introduction of the Digital Facsimiles tagset. The EPPT relies on an image catalog, a separate document that lists all image files associated with a project (several images of the same surface may be listed together, and thus linked to from the text encoding simultaneously). The linking system relies on a hierarchy within the text encoding: 1) A link to the entire image file (or list of image files) is created on a tag that is the designated "surface" tag; 2) Coordinates identifying areas on that surface are placed directly on the tags corresponding with the area. Any number of coordinates may be placed on a tag, which means that an area may be of any shape, not just a rectangular bounding box.

Image Markup Tool (IMT) TEI format

As described in the **Existing Tools** section, the IMT creates and accepts a tightly constrained format of TEI P5. The format reflects the functionality of the IMT as an annotation tool, placing the text that corresponds with each image area in its own separate <div> tag containing a header and annotation, and linking that <div> to the relevant zone defined in the facsimile section. Compare this with the more general example provided above under **Linking Text to Image in TEI P5**, which links zones directly to

line breaks within a transcription of a primary source text.

Comic Book Markup Language (CBML)

As described above in the **Methods** section (under the description of projects at Indiana University Bloomington), CBML is a TEI extension for encoding comic books and graphic novels, and it can also be used to encode other diverse media types with tightly integrated text and graphics, e.g., illuminated manuscripts, illustrated children's books, art books, and others.

EpiDoc

EpiDoc is the standard for encoding epigraphic materials in TEI. It is an extension of TEI P4, an earlier version of TEI that does not include the Digital Facsimile family of tags. The EpiDoc development team will be updating to P5 by the end of 2008, however, and will then be including recommendations of their own for implementing the Digital Facsimile functions within their own Guidelines (<http://www.stoa.org/epidoc/gl/5/>).¹ EpiDoc is widely used in the classics community for encoding both inscriptions and papyrological documents, and this community will need simple tools for creating links between transcriptions, descriptions, and images of the stone inscriptions and papyrus scrolls (in many cases, damaged and difficult to read). TILE will be an attractive toolset for classicists if it has the ability to output valid EpiDoc.

Workplan

Project development is divided into two main parts. Year one is dedicated broadly to tools development, while year two is dedicated to tools refinement and testing.

Year One

Year one will begin with a two-day meeting at MITH with the project PIs, representatives from the project partners, and the external evaluator. During this meeting the group will discuss the broad aims of the project and each partner will have the opportunity to present his or her own project or projects and discuss its particular needs. Final decisions about the project plan will be made at this meeting, and quarterly benchmarks will be set.

Tools development will take place at MITH under the immediate supervision of Doug Reside. Porter will manage all the project partners. Walsh and Porter will test each iteration of the software and provide feedback to Reside and the programmer, communicating regularly via email and through weekly Skype chats. To further facilitate communication across the distance and to keep our project partners informed during development we will set up a project wiki (imagetool.pbwiki.com) to document the development process, including problems found and overcome, to develop tutorials on using the tools, and general information about the project. In addition, Porter and Walsh will visit MITH for a mid-year two-day worksprint, which will give the project team an opportunity to engage in design or redesign of components, data formats, and processes, optimize workflow, identify and manage unforeseen requirements, adjust internal scheduling, troubleshoot bugs, etc. Conducting these exercises in person, in an environment temporarily removed from day-to-day distractions, will effectively jumpstart the next stage of development and will be an indispensable part of the project work.² In January of 2010 the alpha

¹ co-PI Porter is part of the working group that is planning and will be implementing the conversion from P4 to P5.

² Worksprints have been used to great effect by the EpiDoc development team (<http://epidoc.sourceforge.net/>)

version of the tools will be released to the project partners and to external evaluator, Melissa Terras, who, at the end of the first year, will provide an evaluation with recommendations for the improvement of the tool and suggestions for the testing phase.

Schedule

May 2009: Initial meeting at MITH.

May 2009-December 2009: Software development at MITH.

January 2010: PI Meeting at MITH. Initial review of tools.

January-March 2010: Alpha version of tool released to project partners, Terras, and through the TEI Text and Graphics SIG. Alpha testing begins.

March 2010: Terras provides external evaluation report.

Year Two

Year two will also begin with a two-day meeting at MITH with the project PIs, representatives from the project partners, and the external evaluator. During this meeting our evaluator will have the opportunity to present her evaluation of the first year and we and the project partners will have the opportunity to respond to her concerns. We will look at the tool as a group, and the partners will be given individual lessons in using the tools. At this time, the tutorials on the wiki can be updated based on the partners' first experience with the software. Finally, projects will set benchmarks for themselves regarding how they will proceed with using the tools, and each project will be given its own wiki section for documentation and communication with the rest of the project team.

During the year, project partners will provide feedback to the programming team for refinements to be made to the functionality of TILE, and the programmer will respond to their requests. At the same time, the programmer will create XSLT for output using the EPPT linking method, in the IMT TEI format, in CBML, and in EpiDoc TEI. As in the first year, the project PIs will communicate regularly via email and weekly Skype chats. In addition, as the project partners will be really the most active participants in the second year, Reside, Walsh and Porter will also have twice-monthly chats with each project representative, in addition to wiki communication. We will also hold a mid-year worksprint in the second year at Indiana University Bloomington.

At the end of year two, TILE will be thoroughly tested and ready for final public release. We will make the code available through SourceForge on an open source license. In addition, our project partners will have a year's worth of encoded texts and images linked by our software.

Schedule:

May 2010: Second meeting at MITH.

May 2010-December 2010: Software testing and refinement at partner institutions and MITH.

and in the recent Mellon-funded Integrating Digital Papyrology project (<http://www.cch.kcl.ac.uk/research/projects/proj-epiduke.html>). In our view they are vital to the success of the project; there is nothing that can replace regular face-to-face meetings.

January 2011: PI worksprint at IUB. Final decisions about tool refinements made.

January-March 2011: Final refinements completed at MITH.

April 2011: TILE is released.

Staff

Principal Investigators

Since contributing to the development of the EPT (now EPPT), **Ms. Dot Porter** has worked on many image-based digital humanities projects. Porter's experience ranges from editions of medieval manuscripts (Szarmach and Hall, Kleist) to larger collections of classical texts (Dué) to multimedia collections (Rouhier-Willoughby) and projects concerned with advanced imaging of artifacts (Seales). She also consulted on the Mapas project at the University of Oregon (January 2008), one of our project partners. During her tenure on the TEI Technical Council (2006-2007) she spearheaded the development of the image-linking recommendations that were included in the TEI P5 release (along with Conal Tuohy from the New Zealand Electronic Text Center). As part of that work she compiled a document including all XML-based image-text linking methods that were then available, and this document is still available at the TEI Facsimile wiki (<http://www.tei-c.org/wiki/index.php/LegacyFacsimileMarkup>). She taught a course on image-based editing at the Digital Humanities Summer Institute at the University of Victoria in 2007, and will be teaching another course on expressing physical materiality in digital projects at the same venue in 2009. Her familiarity with the needs of various image-based projects, coupled with her knowledge of different linking methods, make her uniquely capable to lead the development of software that can be usable across projects. Porter will test the tools in the context of three active projects, participate in development calls and twice annual meetings, and in the second year supervise a full-time RA, partially funded by the project, who will work intensively with the tools. She will provide feedback to the project team and illustrate examples for other potential testers and users. 10%

After earning his PhD from the University of Kentucky in 2006, **Dr. Doug Reside** moved to MITH where he assumed the role of assistant director. At MITH, Reside manages all of the technical work at the center and has helped to design many projects requiring links between text and image including the Our Americas project (an IMLS-funded multi-institutional archive of texts and images documenting the cultural history of the Americas), Soweto '76 (a 3-dimensional museum of artifacts related to a student uprising in South Africa), and the Shakespeare Quartos Archive (an NEH/JISC funded series of image-based electronic editions of the earliest printings of Shakespeare's works). In August of 2007, Reside also received a Digital Humanities startup grant for the Ajax XML Encoder (AXE), which serves as a prototype for the current project. Reside will supervise the full-time programmer, participate in development calls and twice annual meetings, and provide leadership in tools development. 10%.

Much of the TILE testing will be directed by **Dr. John Walsh** at Indiana University in Bloomington. Walsh is currently working on three relevant digital humanities projects: The Algernon Charles Swinburne Project, The Chymistry of Isaac Newton, and the Comic Book Markup Language. All three projects involve the editing and encoding of texts and images. In the printed volumes and autograph literary manuscripts of the Swinburne Project; the laboratory notebooks and alchemical texts, with their illustrations and special symbols, of the Newton Project; and the tightly integrated text and graphics of the comic book and graphic novel, these three projects provide examples of different levels of integration of text and image. Walsh has a general research interest in the digital representation of graphically rich and

complex texts and documents, including also illuminated manuscripts, art books, and illustrated children's books. Along with Porter, Walsh convenes the Text and Graphics SIG of the Text Encoding Initiative Consortium. The SIG is sponsoring two sessions at the TEI Member's Meeting in 2008 and the SIG (and the TEI community as a whole) will be the primary audience for the tools during the testing phase of the project. Walsh will test the tools in the context of three active projects, participate in development calls and twice annual meetings, and in the second year supervise a full-time RA, partially funded by the project, who will work intensively with the tools. He will provide feedback to the project team and illustrate examples for other potential testers and users. 10%

Consultant

Melissa Terras is a senior Lecturer in Electronic Communication at University College London School of Library, Archive and Information Studies. Terras' research interests include Humanities Computing, Digitization and Digital Imaging, Artificial Intelligence, Palaeography, Knowledge Elicitation, and Internet Technologies. She is the Associate Director of the JISC funded Virtual Environments for Research in Archaeology (VERA) Project (with the University of Reading), and co-Investigator of the AHRC-EPSRC-JISC funded Image, Text, Interpretation: e-Science, Technology and Documents Project (with the University of Oxford). Terras' dissertation involved the creation of a software system to help classical scholars read damaged texts (Terras 2006). Terras will participate in annual meetings, providing feedback and official evaluation reports. Four days.

Programmer

The programmer, to be hired at the beginning of the funded period, will be responsible for tools development in JavaScript, PHP, XML, XSLT, and MySQL. 100%

Directors of the project partners

Mapas Project: **Stephanie Wood** and **Judith Musick**, University of Oregon;

Homer Multitext: **Casey Dué**, University of Houston and **Mary Ebbott**, College of the Holy Cross (Editors); **Christopher Blackwell**, Furman University and **Neel Smith**, College of the Holy Cross (Information Architects)

Royal Irish Academy: **Susan Schreibman**

The project directors (one from each project) will attend annual meetings with the project PIs to advise on the initial tool development (year 1) and to receive intensive tutoring on the tools (year 2). In the second year they will hire and supervise a student or students who will work intensively with the tools, provide feedback to the project team, and document usage and illustrate examples for other potential testers and users.

Partner Project Student Assistants

In year two, the project partners will hire student assistants using funds provided by the project. These students will work intensively with the tools, using materials provided by the projects themselves, provide feedback to the project team, and document usage and illustrate examples for other potential testers and users.

Dissemination

Porter will present a paper on the planning for these tools during the Text Encoding Initiative Member's Meeting in London in November, 2008, as part of the sessions organized through the Text and Image SIG, and at that time we will publicize the wiki for image tool development (<http://imagetool.pbwiki.com>). Members will be invited to contribute their project descriptions and use cases, as the project partners and others have already done. During the first year of development, quarterly updates on development will be sent to the SIG listserv as well as to the main TEI listserv. During the second year of the project the tools will be released in an alpha version and given to our project partners for intensive testing. At the same time, it will be released under an open source license and made available through the Texts and Graphics SIG for general use. At the close of the project, the "final" version of the tools will be released more widely, made available through SourceForge and publicized by the project PIs through presentations at conferences both digital-humanities and subject oriented. Porter will present papers at the International Congress on Medieval Studies, Kalamazoo, MI, May 2011 and at the International Medieval Congress in Leeds, UK, July 2011. Reside and Walsh will also present at conferences relevant to their own subject specialties, TBD.

Sustainability

The tools developed through this project will be made sustainable through use and continued development in the community. Individual scholars and projects will be able to download the tools and the source code, made available under an open source license through a page on SourceForge. The tools will be incorporated into the Son of Suda Online, a collaborative editing suite being developed at the University of Kentucky under a grant from the Mellon Foundation, and will also become a part of the Backstop project for the preservation and dissemination of classical materials at the Institute for the Study of the Ancient World at NYU (see appendix for letters of commitment). In addition, there is the potential for official TEI support.