

The Text-Image Linking Environment

[opening slide]

[Peter Stokes: First, let me be clear again that I am reading this out for Dot Porter who can't be here today. I'm not involved in the project at all and certainly don't deserve any credit for it; unfortunately I probably won't be able to answer any questions about the project itself either, though I am familiar with the general context and some of the other projects and standards which are mentioned, so I might be able to help there. Otherwise I'm sure the project members would be happy for you to contact them directly.]

[SLIDE] The Text-Image Linking Environment (or TILE) is a project recently funded through the National Endowment for the Humanities (Preservation and Access program) to develop a web-based, modular set of tools for various types of linking between texts and images. TILE is a collaboration amongst Doug Reside at the Maryland Institute for Technology in the Humanities, Dot Porter at the Digital Humanities Observatory, Royal Irish Academy, Dublin, and John Walsh at the School of Library and Information Science, Indiana University Bloomington. Project partners include the Homer Multitext Project at Harvard's Center for Hellenic Studies and the Mapas Project at the University of Oregon.

For a few years, the project PIs have been aware of the need for a cross-platform, web-based, modular set of tools for ((working with and editing images) (with and alongside texts)) in various different ways. Although there are of course several tools in existence, they felt that something was not being done right and felt that now is the time to bring their particular vision to light.

Why now? First the Ajax XML Encoder, built through an NEH Digital Humanities Start-Up Grant, reached the beta development stage and so now provides a reliable basis for building this next-generation set of tools. **[SLIDE]** Second, in the past few years there has been a noticeable increase in the interest of project developers to incorporate images into their projects in increasingly complex ways. **[SLIDE]** Third, related to the second issue, was the release in 2007 of the Text Encoding Initiative Guidelines P5, a vast change and improvement over P4 and which included for the first time a substantial set of tags designed specifically for incorporating image-related information into TEI files. **[SLIDE]** What this provides is a standard method for linking text and image, which was likely to be adopted by the text encoding community, and so we finally have a method for TIE to be built around.

Tools

As the community of scholars developing image-based projects has grown over the past decade (and with more speed over the past five years or so), more tools have been created that are actively used for project development and that have been successfully maintained past initial creation. The project PIs know of no fewer than ten tools or collections of tools that have been designed to 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, such as Juxta or the Versioning Machine **[SLIDE]** to very robust software suites which support the development of complete image-based projects with substantial functionality beyond simple text-to-image mapping, such as the IBX (previously EPPT) and the Image Markup Tool (or IMT) **[SLIDE]**.

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). 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. Similar but more simple tools along the same lines have been developed by Hugh Cayless at New York University (who has recently received a DH Start-Up grant to complete development of the img2text tool, and who is himself a TILE collaborator) and Doug Reside, who originally developed his Word Linking tool for the Shakespeare Quartos project (and who is one of the PIs on the TILE project). **[SLIDE]** At the moment, the project PIs are planning to incorporate aspects of both tools into TILE.

Weaknesses of Existing Tools

TILE was proposed to answer 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 and Hugh Cayless's img2xml) are manual, depending on a human user to create the relationships between the image and the text. The project PIs have used these tools in their own work and know first-hand how time consuming and tiring the manual tagging can be. A major goal of TILE is therefore to automate wherever possible, in order to make the process of incorporating images much easier than is possible now.

The second major weakness of current tools is that they are not, in general, interoperable. Individually, each of these tools allows a user to do interesting things, but—because of variations in output—interoperability is simply not possible.

TILE will help enable the interoperation of existing tools by providing a system for

converting between the outputs of frequently used tools (in the first instance, between the IMT and the IBX). TILE seeks not to replace existing tools so much as to create a system that enables these tools to work well together, and accounting for different types of output in the first year will allow us 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.

How is TILE different?

How is TILE different from these existing tools, then? The philosophy of the project PIs is that, from the start of development, the individual needs of multiple users and many different types of materials need to be taken into account. These users need to be a part of the team, and TILE programmers need to be mindful of their needs at all stages of development.

TILE represents an innovative approach to tools development. In fact, the project team like to think of the TILE development as a *metaproject* rather than as a project in and of itself: rather than creating collections, they 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 from the start and will contribute recommendations for refining the tool during an intensive testing period during the project's second year.

Evaluation

The development process and TILE itself represent serious innovation in the practice of tools development for the humanities, and the project has arranged for both internal and

external evaluation to measure our success. Partners will work closely with the PIs during the first year as they develop the tool, and will also test the tool during the second year and provide suggestions for further refinement. The PIs have also involved an external evaluator, Melissa Terras, who is senior Lecturer in Electronic Communication at University College London's School of Library, Archive and Information Studies. She will attend annual meetings with the development team and project partners, and will provide expert recommendations during the development and testing process. With extensive experience both in software development and in usability testing, Terras is a valuable member of the team, and her knowledge and experience will be invaluable to them as they develop TILE.

Conclusion

[SLIDE] TILE is being built with the expectation that it will play a central role in the next generation of technical infrastructure supporting image-based editions and electronic archives of humanities content. 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 will dramatically increase the ease and efficiency of this work. At the end of the two-year project, which started this May and will finish in May 2011, the project team will have produced software interoperable with other popular tools and capable of producing TEI-compliant XML for linking image to text. It will also put the image linking features of the newest version of the Text Encoding Standard (TEI P5) through one of its first rigorous, "real world" tests. TILE will be both modular and extensible; projects in the future will be able to add their own tools to the mix in addition to using those created as part of TILE in the project's first instance. TILE will be developed and thoroughly tested with the assistance of project partners as well as volunteers from other

projects, 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, and with the expectation that, unlike so many other tools, it will be *used by* the community.