

## Examining Scholarly and Pedagogical Implications of the Text Encoding Initiative (TEI) for Small Liberal Arts Colleges

Kathleen Ebert-Zawasky, Director of Research and Instruction, and  
Kathryn Tomasek, Associate Professor of History, Wheaton College

Abstract

### *Encoding Text, Revealing Meaning*

*Faculty, students and librarians from six colleges are analyzing archival and literary texts by applying standards developed by the Text Encoding Initiative. One developing pedagogical model involves history students producing comparative editions of a document by transcribing and encoding the primary source material from the college archives.*

In the spring of 2004, a team of librarians and technology support colleagues from Mt Holyoke and Wheaton Colleges met to plan a three-day intensive summer workshop in explore the feasibility of implementing some of methods of the text encoding process using the Text Encoding Initiative (TEI) standard into the undergraduate liberal arts classroom. Simultaneously we sought learn more about creating digital finding aids for special collections using XML and the Encoded Archival Description (EAD) tag set. The close relationship between the two goals became quite clear as the finding aids provide a contextual framework and hyperlinks to encoded documents - letters, diaries and other original manuscripts. For example, a finding aid for the Lucy Larcom collection that provides access to specific encoded documents in the collection would satisfy both goals.

In June, teams from six MANE colleges converged at Wheaton for a three-day workshop.

Faculty, librarians and technologists shared fascinating project ideas. Participants cam from Dickinson College, Mount Holyoke, Muhlenberg, Union, Wheaton and Williams College.

The workshop enabled participants to learn from experts about TEI and its applications, the process of encoding and an introduction to the guidelines. In hands-on sessions, the teams learned about text editors how to use them to embed the text with tags that describe the structure of the text on the first pass and then interpret the text on a second pass. The second pass requires interpretation of the text and thoughtful embedding of meaningful descriptive tags following the syntax of the TEI standard.

Details and materials from the summer workshop are available at a link on Wheaton's TEI website: <http://tei.wheatonma.edu/>

The projects that were proposed came from diverse fields of study. Selected examples include:

Gavin Ferriby, Systems and Electronic Resources Librarian at Muhlenberg College, is working to digitize and encode 2nd-6th century C.E. Greek and Latin papyri for future use in the history curriculum.

Domingo Ledezma, a professor at Wheaton College, has students transcribing and encoding early modern Hispanic accounts of travels by sea and shipwrecks. This semester he and his students are working on *The Book of the Misfortunes and Shipwrecks* by Gonzalo Fernandez de Oviedo (1535) for a Spanish literature course.

Linda Patrik, a professor at Union College, is encoding selections from a Tibetan philosophy text (in translation) to emphasize the logical structure of the argument. This text is required in an upper level philosophy course in Zen and Tibetan Buddhism. Once encoded, it will also be used as an example of Tibetan debate for another course, Asian Philosophy.

The Schaffer Library TEI team of Union College have two projects from their special collections - encoding the Laws of Union College (1796 and 1802) and a collection of papers from the civil war. Once digitized these primary sources will be more accessible for scholarly work, Union history class projects, and student research projects.

Kathryn Tomasek, a professor at Wheaton College, plans to develop encoding projects for her Women's History courses that will enable students to work extensively with primary sources and that will help the Library Special Collections and Archives at Wheaton to make sources digitally available. This semester her class is transcribing and encoding a personal diary from the 19th century that reveals deep societal issues about women from the period.

---

Now I will introduce Kathryn Tomasek, Associate Professor of History from Wheaton. Kathryn will share the pedagogical model she is developing and currently using with one of her classes. We will finish with a description of the software tools we are using and plans for the TEI Seminar we are planning for January 21, 2005.

---

Incorporating the transcription and encoding of Maria Wood's journal into History 230, my survey of U.S. Women's History to 1869, is only the most recent of many projects I have undertaken over the past twelve years to explore the utility of technology as a pedagogical tool. This use of TEI (Text Encoding Initiative) in a course at a small liberal arts college is easily one of the most successful of these projects, and collaborations have been an integral part of the project and its success. Without the help of Wheaton's College Archivist Zephorene Stickney and Faculty Technology Liaison

Scott Hamlin, I could not have undertaken this project, and my students would have lost this important opportunity to learn about history and interpretation. The students and I have found the experience enormously enriching.

I had three main pedagogical goals when I proposed this project:

- To give students an opportunity to learn and hone many of the skills important both to a student of history and to a participant in Wheaton's new curriculum.
  - Through transcription, students learn and practice their ability to read manuscripts.
  - Through encoding the text, a process analogous to creating a critical edition, students learn to think analytically about documents.
  - Through the entire process of transcription and encoding, students will improve their skills in reading and understanding primary sources and will learn the necessity of close reading for comprehending them.
- To provide students with the opportunity to experience the process of editing primary sources for broad distribution. This is an important process for History majors, in particular, to learn early in their careers at Wheaton, because it will prepare them for research projects that they will undertake in senior seminar. And this will make all students (regardless of their major) more sophisticated readers of such sources in print or on the Internet.
- To infuse technology into this course by using materials focused on "issues pertaining to... people" (in this case, American women) who have been traditionally "neglected by Western scholarship."

Maria Wood was an ordinary white woman of the lower middle class, a daughter of a minister, who lived and taught for a time in Alton, Illinois, in the 1870s. She moved back to New England with her family, perhaps because of her father's illness. He endured a long and painful illness and death, after which Maria Wood undertook various work to support herself and her mother and sister. She taught music in Skowhegan, Maine, before finally moving back to South Boston to live with her mother and sister.

Zeph, Scott, and I worked collaboratively over the summer of 2004 to prepare the assignments for the fall semester of 2004. Zeph and I practiced and copyedited transcriptions and made lists of names, places, organizations, and interpretive themes for encoding the document. Zeph, Scott, and I met several times to hone the steps students would go through in the process of transcription and encoding, deciding on the number of passes students would make through the manuscript and the dates when class time would be devoted to TEI instruction.

For our second class meeting of the semester, students met in the Woolley Electronic Classroom in the Madeleine Clark Wallace Library, where Zeph, Scott, and I presented an overview of the project. Zeph introduced students to Maria Wood and her

diary, Scott explained the technological aspects of the project, and I talked about the unique opportunity students would have to work with a primary source and to make it accessible to a broad audience in ways that were not possible before this semester. I explained the significance of written records of the lives of seemingly insignificant people and the vagaries of how historical documents are preserved. I emphasized the importance of the students' pioneering work in the use of TEI in a small liberal arts college and the great potential of such projects to make women's diaries accessible to researchers through links to the documents on the Archives and Special Collections webpage.

As part of her presentation that day, Zeph encouraged students to consider the questions that historians ask of all primary sources:

- What do we KNOW about the author and her life?
- What can we DEDUCE reasonably and with a fair degree of certainty?
- What can we INFER speculatively yet reasonably?
- What QUESTIONS are raised and/or unanswered?
  - What can we learn about the author?
  - What did this woman do?
  - Who and what made up her personal world?
  - What did she think about?
  - What did she care about?
  - What did she know about the world around her?
  - How did she view herself and her role in life?
  - How might we generalize about life in the nineteenth-century United States or about the lives of other women from what we read in this diary?

Zeph also prepared students for the themes that arise in the diary and for its incompleteness, encouraging them to think about the gaps in the diary, changes in ink color, and misspellings that should be transcribed as written.

We handed out an assignment sheet (available at: <http://www.wheatoncollege.edu/KACC/tei/teaching/tomasek04/assignment.html>), informing the students that they would make three passes through the document: transcribing the document; encoding names, dates, places, and organizations; and encoding interpretive themes.

On the whole, this series of assignments presents all of the collaborators with a series of significant opportunities:

- I am finding new ways to explore intersections between my teaching and research;
- Zeph, Scott, and I are exploring ways to better educate History majors, non-majors, and students majoring and minoring in Women's Studies;
- And students are preparing for possible continuing collaborations
  - With Zeph in the Archives and Special Collections, where they can transcribe and encode additional texts, thus increasing accessibility of such primary sources—this stands in stark contrast to practices at other liberal

- arts colleges, such as Hamilton, where libraries outsource transcription and encoding of documents;
- And with me as I conduct ongoing research in the Wheaton College Archives and Special Collections.

At this point in the semester (Nov. 8, 2004), students have completed the first two steps in the process of transcription and encoding. They have made use of jpeg images of the diary. Examples can be seen at:

<http://www.wheatoncollege.edu/KACC/tei/WooJour/woojour3.jpg>

and

<http://www.wheatoncollege.edu/KACC/tei/WooJour/woojour13.jpg>

Definition and examples are described at this site:

<http://www.wheatonma.edu/KACC/tei/teaching/teidef.html>

Entities and elements used in the project are here:

<http://www.wheatonma.edu/KACC/tei/teaching/HIST-230Aent.html>

Instructions for names and dates tagging follow at:

<http://www.wheatoncollege.edu/KACC/tei/teaching/namedate.html>

Students have also written short papers about the process. I quote from a few here. After the transcriptions were complete, students wrote:

“I have never actually dealt with a historical artifact like this. Being able to take part in the project has let me see a whole new side of history.”

“Initially, I thought that the transcription process was going to be difficult, but surprisingly it wasn’t too bad. I became engrossed in what Maria Wood was talking about and felt myself her interpreter as I transcribed her thoughts and emotions.”

“I enjoyed reading through her diary and learning about her life. This process exemplified the importance of being thorough and meticulous when dealing with the transcription of a historical document.”

And after tagging names, dates, places, and organizations, students wrote:

“Having to go through the diary in such detail really gives you a better understanding of the diary. The diary becomes much more clear because you have to really pay attention to what she is talking about.”

“I found this coding session to be bizarrely therapeutic. I may have achieved a state of XML Zen. ... I felt as though I was in a little vortex of codes and Maria Wood.”

“I think that coding things like this is a really great idea. I was explaining the process recently to a friend of mine and used their last name “Cook” as an example of how this tagging would help a historian search through journals. Without the coding, a historian searching for the

word “cook” would find people’s name, occupations, and probably a few towns with cook in the name. With the tagging, the search can be made very exact.”

“In the end, the process of tagging was actually fun... As the process of tagging Maria Wood’s journal progresses, it becomes increasingly interesting. The end of each stage of the transcribing process is like a cliffhanger, leaving me more interested in what will happen next.”

The work that students in History 231 have done this semester has been a truly successful example of the potential of technology in teaching liberal arts undergraduates some of the processes of historical interpretation. It would have been impossible without the collaboration of Zeph Stickney and Scott Hamlin.

Students will complete this project by writing a short paper on Maria Wood and how her diary adds to our understanding of women’s lives in the United States in the 1870s. One student wrote, “I found reading her journal to be a successful venture into humanizing the information presented in history books. I appreciated Maria Wood as a real-life example of some of the things I have been reading about all semester.” She is ready for her final exam.

---

## Tools

In addition to tackling the challenges of teaching how to encode text and how to carve out a subset of TEI tags that could meet our coding needs, we also faced the challenge of finding appropriate tools to use.

- We chose to use *<oxygen/> XML Editor* for authoring, editing and validating text.
- For transforming and displaying encoded documents, we adapted an in-house system – which we fondly call “Brian’s Transforming Machine.” (Thanks to Brian Gibson for this contribution.) Essentially, Brian built a front-end to the Saxon open source XSLT and Xquery Processor. XML documents plus xsl stylesheets are uploaded into the system and processed or transformed into html documents. This system, with permission, may be accessed at <http://webedit.wheatoncollege.edu/login.html>

Where are we now?

We are enthusiastically planning to bring back the teams who joined us in the summer workshop for Presentation and Exchange Seminar to be held on January 21, 2005 at Wheaton. The seminar has as its centerpiece a faculty panel. Four to six examples of pedagogical approaches will be presented along with the student an assessment of their first application when possible. Student responses and challenges encountered will be an important part of these sessions.

Concurrent Sessions:

1. Hands-on experience: engaging in one pedagogical approach
2. Archival Library Issues
  - Laws of Union College
  - Lucy Larcom Collection

Report: Highlights of the annual TEI conference

- Emerging Tools
- Advantages of the P5 standard

Discussion/Evaluation

- What have students learned?
- Does this type of undergraduate experience further our information fluency goals?
- What have we learned?
- How can we help each other to continue TEI/EAD work?
- Are these projects sustainable? Are the support issues manageable?

We wish to thank the Center for Educational Technology at Middlebury College for support for this project.

For further information, please visit <http://tei.wheatonma.edu/> or contact: [kzawasky@wheatoncollege.edu](mailto:kzawasky@wheatoncollege.edu) or [ktomasek@wheatoncollege.edu](mailto:ktomasek@wheatoncollege.edu)