TEI Initiative at Mount Holyoke College

Text Encoding Initiative (TEI)
Today TEI has become a versatile technology tool for digital humanities scholarship, especially in the fields of history, philosophy, linguistics, literature, art, archaeology, music, and cultural studies. For example, at the recent Digital Humanities 2011 conference, hosted by the Stanford University Library, a great number of presentations, panels, workshops and papers were on TEI related topics. A few examples included “Possible Worlds: Authorial markup and digital scholarship,” “Semantically Rich Tools for Text Exploration: TEI and SEASR,” “The Tesseract Project: Intertextual Analysis of Latin Poetry,” “Charlotte’s Web: Encoding the Literary History of the Sentimental Novel,” and “The Time Machine: Capturing Worlds across Time in Texts.” There was also a tutorial on Integrating Digital Humanities Projects into the Undergraduate Curriculum.

Since Mount Holyoke’s approach in the curriculum is to treat our upper-level undergraduates like beginning researchers, it may be important to engage students in humanities research through this technology. TEI helps to engage the students in editorial practices that make apparent the textual history and culture of a print document by having to translate it or replicate it in digital format. In addition, the students can gain marketable skills by being exposed to up-to-date technologies. Furthermore, if our professors want to be on the cutting edge of research, this is a useful tool for combining scholarship and teaching in a way that allows them to advance their research projects.

TEI Consortium
http://www.tei-c.org/
The Text Encoding Initiative is a non-profit consortium composed of hundreds of academic institutions, research projects, and individual scholars from around the world. For the last three decades this consortium collectively has developed and maintained an international and interdisciplinary standard for the representation of texts in digital form.

http://www.tei-c.org/Guidelines/P5/
The main publication of the organization is a set of Guidelines for Electronic Text Encoding and Interchange, which specifies encoding methods for texts particularly in the humanities, social sciences, linguistics, etc. P5, the current version of the TEI Guidelines, based on XML, was officially released on November 1, 2007. It has been widely used by libraries, museums, publishers, and individual scholars worldwide to present texts for online research, teaching, and preservation, thus becoming a sustainable standard for digital humanities scholarship.

Why Encode Text with TEI?
The purpose of encoding text is twofold: TEI encoding employs SGML and (more recently) XML to enhance digital texts with structural and analytical metadata so they can be manipulated, represented, and queried in more complex ways than ordinary digital representations of texts. Many large research universities and institutes have utilized this kind of text encoding with TEI to develop large-scale projects like the World of Dante at UVA’s Institute for Advanced Technology in
the Humanities (IATH), the Newton Papers Project at the University of Sussex and the Women Writer’s Project at Brown University.

In addition to creating sizable digital products, the process of encoding requires decision-making by the encoder – choices which can amount to performing close readings and sophisticated analyses of the text. Both the act of encoding and the use of a TEI-encoded text, therefore, can lead to a deeper understanding of the text, reveal the creative process, and stimulate critical thinking.

Why Did We Choose XML/TEI Instead of HTML?
Derived from SGML, both HTML (Hypertext Markup Language) and XML (eXtensible Markup Language) are Web standards developed by the World Wide Web Consortium (W3C). The main difference between the two is that HTML has very limited tags and its content is difficult to reuse because of the mixture of content and format, while XML elements are defined by the users and its content can be repurposed in many formats, such as XHTML, text, Word, PDF, etc. with the aid of stylesheets (CSS or XSLT) and used on different mediums. While XML inserts meaning in the content, good for long-term preservation, HTML content has no longevity. Therefore, as one of the most innovative technologies in the 21st century, XML has laid the foundation for the next generation Web: the Semantic Web.

Based on XML rules, TEI elements are selected and defined by the scholarly community through user input, thus making TEI a versatile tool in encoding, analyzing and manipulating texts according to users’ needs. With the Semantic Web on the horizon it may be necessary for our students to get familiar with the TEI standard by learning how to insert meaningful metadata into the text.

People may wonder how expensive the startup costs would be and how difficult it would be for a faculty member to incorporate TEI in their courses. TEI guidelines are complicated and extensive, composed of hundreds of elements and attributes. However if a faculty member just wants to bring a letter, a manuscript or a poem to the class for textual analysis it would not be too hard since we only need to select a small number of elements and a few attributes from the guidelines to teach the students how to use them in their projects. We also need to teach the students to use Oxygen, an XML editor, to encode the text. We have often hired a tech mentor to support the student group projects; that has been our only expense so far. Of course the support staff should know more about TEI in order to select the appropriate TEI tags to help faculty achieve their pedagogical objectives, and learn how to design stylesheets with CSS or XSLT to transform the encoded texts.

How Did the TEI Initiative Start at MHC?
At Mount Holyoke College (MHC), a small liberal arts college with limited resources, we have tried to bring TEI into the curriculum to enhance teaching, learning and research in several disciplines. As early as 2003 Don O’Shea, Dean of Faculty, and Susanne Woods, former Provost of Wheaton College, initiated the TEI initiative at MHC when they decided to collaborate on writing a grant proposal to NITLE for organizing a one-week workshop for small library arts colleges with solicitation of faculty participation. Nicole Vaget,
Professor in French, Donald Cotter, Associate Professor in Chemistry, participated in that workshop in January 2004.

**Who Has Participated?**
Over the years Nicole Vaget (French), Donald Cotter (Chemistry), Kate Singer (English), and Bob Schwartz (History), have utilized TEI technology in a number of courses in several disciplines. Nicole Vaget, James Hartley (Economics), Mohammed Jiyad (Asian Studies), and Fang-chun Kuo (Asian Studies), have developed several small-scale TEI research projects through collaboration with RIS, or with RIS and Archives and Special Collections.

**Selected TEI Courses and Project**
*Mémoires de madame la comtesse de L... -- an 18th century manuscript*
Nicole Vaget, Professor of French
A collective and collaborative research project, French 331, fall 2005

Professor Vaget was the pioneer in bringing the TEI technology into the MHC curriculum. In fall 2005 Professor Vaget taught an advanced seminar in French literature on the condition of women of the aristocracy in 18th century France. In her class, students first read and analyzed the works of major authors. Then, they studied relevant art works of the 18th century. Students wrote two traditional papers, but their main research project was the encoding of the unpublished 18th century manuscript. The purpose of teaching students how to mark up was to invite them to do a very close reading of the text, and integrate the cultural material learned in class.

The students’ assignment was to mark up structural, contextual, and analytical elements of the text, step by step by following the TEI Guidelines. Students were assigned the following tasks:

1. Analyzing themes
2. Finding appropriate images
3. Incorporating cultural notes
4. Identifying people and places

Students’ positive responses showed that, with this method, they got a close reading of the text. They gained insight into the historical, political and social issues of the original material. It was clear that the intellectual challenges posed by the interpretation of the manuscript, along with the extensive study of major works of the period, led students to a better understanding of life before the French Revolution, and of the condition of noble women in the French provinces in the 18th century. In addition, the students felt fortunate to be exposed to the more current technologies, as one student commented in the assessment “Learning XML puts you one step ahead of HTML, and gives an extra hedge.”

Revised and completed by Sealia Thévenau as an independent study project, 2009-2010
Melesina Trench Edition: Laura’s Dream, or the Moonlanders  
Kate Singer Assistant Professor in English  
http://www.mtholyoke.edu/courses/smoss/moon_project/

In the spring 2011 Professor Kate Singer worked with her upper-level class, English 377 to create a teaching edition – built by and for students – of Melesina Chevenix Trench’s poem “Laura’s Dream, or the Moonlanders. This poem serves as the starting point of a scholarly edition and archive of Melesina Chevenix Trench Collection, which includes all of her publications, such as poetry, articles, letters, etc.

Members of the English 377/Gender Studies 333 seminar “Feminist Poetics” worked in groups to analyze substantial passages from Melesina Trench’s lunar voyage poem entitled Laura’s Dream; or the Moonlanders. After reading and discussing the entire poem in class, groups were taught TEI in a short workshop and then asked to encode the following items in their group projects: lines and stanzas, people and places (both real and imaginary), and then three figurative tropes of their choosing (e.g., “affect,” “spirituality,” and “morbidity”). After this, students color-coded these tropes, using CSS. At the semester’s end, groups marking up the same passage compared their encoding practices, discussing what kinds of interpretations each mark-up fostered. A website of the full-text alongside student color-coded passages will be online by the end of the Fall 2011 semester.

Professor Kate Singer represents the new generation of college professors who have taken the lead in bringing the digital humanity scholarship into their disciplines. She is fluent with TEI text encoding, fully understanding its pedagogical use. She adopted this new approach in teaching by asking her students to insert descriptive metadata into the text and color-coding meaningful words and phrases in the poem. In this way the students were able to get inside the material through group interpretation of “visualizations” and reflection on descriptive vocabularies. As a result Kate originated an excellent way to teach “analogue” close reading, which is much effective than the traditional close reading practices using a pencil and paper.

Letter from Emma P. Carr to Professor Victor Henri [ca July 5, 1924]  
Donald Cotter, Associate Professor in Chemistry  
Student group projects in Chem250: History of Chemistry, fall 2004  

Mary Lyon: Documents and Writings - an electronic version of James E. Hartley’s book. Professor Mary Renda in Gender Studies has integrated this resource into her course on Mary Lyon to aid her students’ research projects this fall semester.  
http://www.mtholyoke.edu/lits/library/arch/marylyon/home.html

This project investigates XML/TEI to create a comprehensive digital version of the Mary Lyon papers in Archives and Special Collections. It could be used to apply for grants and other funding for its development.
**Women's Deceit** (work in progress)
Mohammed Jiyad in Asian Studies
http://www.mtholyoke.edu/courses/smoss/women_deceit_v3/home.html

A Scholarly Online Edition of Translation of a 15th-century Arabic manuscript "Women’s Deceits,” translated by Professor Mohammed Jiyad. Currently his independent study students are incorporating various references to the texts, including videos, maps, pictures, and articles to make it a useful Web resource for his students to do research on attitudes towards women in the Islamic world.

**Emily Dickinson Collection**
http://www.mtholyoke.edu/lits/library/arch/emily_dickinson/Emily_Dickinson.html

Starr Florio, Simmons Intern in Archives, completed this project last summer by using TEI to encode a personal letter written by Dickinson while a student at the Seminary. The letter links to a variety of primary source materials, including classmate correspondence, autograph albums, photograph of classmates and the seminary building, a course catalog, curricular material, textbooks, student account book essays, menus, and other material. It is a collection of primary sources documenting Emily Dickinson’s student days at Mount Holyoke Female Seminary, 1847-1848.

Mellon Library Intern Oomaru encoded summaries of translated French correspondence in the Betty Wells papers.
http://asteria.fivecolleges.edu/findaids/mountholyoke/mshm290.html

**Grants:**
The TAPAS group have applied for four grants since 2009 and received two so far:

1. The IMLS funded National Leadership grant, “Publishing TEI Documents for Small Liberal Arts Colleges: Planning a Service, Building a Community,” 2009 (received)

2. The NEH Office of Digital Humanities Start-up Grant, "Developing a User Experience for TAPAS (The TEI Archiving, Publishing and Access Service)," 2011 (received)

3. An IMLS Leadership Implementation Grant, entitled "Developing TAPAS: The TEI Archiving, Publishing, and Access Service" (pending)

4. An NEH Preservation and Access Grant, entitled "TEI for Real Scholarship: The TEI Archiving, Preservation, and Access Service" (pending)

**Why Did We Apply for the Grants?**
It was discovered that while faculty and students at many liberal arts colleges are working on or have produced TEI encoded materials, they are often unable to effectively and efficiently make them available to the broader academic community because they lack the tools and services needed to store and display their work. In addition, they are often
unable to take full advantage of the considerable structure they have put into their marked-up texts for lack of tools to process transform and manipulated them via native XML databases and other advanced tools, like visualization, GIS, etc.

A single institution may play host to many TEI related projects, therefore, but it cannot justify the expense involved in investing in a comprehensive service. However, several smaller colleges with similar situations and limited means may do so under a consortium arrangement. Therefore we have applied to several grant agencies, seeking funding to find a solution to this important issue – a solution that includes a centralized, shared service that a community of schools will maintain and use to store, display and manipulate their encoded texts.

Objectives to achieve with the grants:

- Plan a TEI document publishing service for small library arts colleges
- Develop and implementing a Fedora repository (based on the Brown digital repository) at Brown University for the TAPAS service
- Develop a front-end user interface to a Fedora repository in Drupal, which will help users upload, manage, publish, and analyze their encoded texts

During the process of the grant application and implementation we obtained participation and support from faculty and staff in over ten institutions. At this point while Brown and others might be doing the technical work to design the infrastructure, the success of the TAPAS project really depends on the consortium—the participation of faculty and staff, especially in contributing projects that will use this service. Faculty from MHC and other similar institutions like Wheaton, Dickinson, Vassar, etc. are very much interested in using the project’s infrastructure once it is operable.

We believe that the grants will make the startup costs much less for faculty who wish to use TEI texts in dynamic ways – enabling them to move beyond static text. Although MHC isn’t providing technical work at this stage, it is extremely important for MHC staff and faculty to continue to play an advisory role, including data testing, because we will be the ones writing grants and using the service for our texts.

Next Steps

We will:

- continue the TEI initiative by supporting Kate, Nicole, Mohammed and other faculty with their TEI courses and projects to meet their pedagogical goals;
- reach out to other faculty by inviting them to attend TEI workshops and faculty seminars, encouraging them to adopt TEI and/or other appropriate technology tools for their teaching and research;
- continue our involvement with the TAPAS project by participating with data testing and grant writing, pushing to materialize the vision of building a TEI document publishing service for small liberal arts colleges;
- expand our technology tool kit by exploring tools like SEASR, MAXQDA, Voyeur and Mandala for quantitative and qualitative text analysis, manipulation and visualization.
Selected Presentations

- *Trench Edition*, panel "Digital Pedagogy," MLA, January 6, 2012 (Kate Singer)
- "Feminist Poetics," presentation on English 377/Gender Studies 333, Innovations in Pedagogy, with a focus on the TEI project, Faculty Retreat, MHC, Sept. 1, 2011 (Kate Singer)
- *The TAPAS Project: Creating a Shared Service for Digital Humanities Projects using TEI*, NERCOMP 2011 Annual Conference (EDUCAUSE), Providence, RI, March 29, 2011 (Shaoping Moss and her colleagues at Wheaton and Brown)
- *Melesina Trench, 'Laura’s Dream,’ and a Vision of a Teaching Edition*, ICT Summit at University of Massachusetts, Amherst, MA, March 30, 2011. (Kate Singer)
- *Mémoires de madame la comtesse de L... -- an 18th century manuscript*, “ ICT Summit at University of Massachusetts, Amherst, MA, March 30, 2011. (Nicole Vaget)
- *Pedagogical Use of XML/TEI Technologies in the Classroom*, NERCOMP 2007 Annual Conference (EDUCAUSE), Worcester, MA, March 20, 2007 (Nicole Vaget and Shaoping Moss)

Acknowledgements
I would like to acknowledge contributions from:

- Don O’Shea, Dean of Faculty, and Susanne Woods, former Provost of Wheaton College, who initiated the TEI initiative at MHC;
- Mount Holyoke faculty -- Nicole Vaget who pioneered the TEI efforts, and Donald Cotter, Kate Singer, Bob Schwartz, Mohammed Jihad, James Hartley, Fang-chun Kuo, and Mary Renda, who have participated in the TEI Initiative;
- LITS staff – Mary Stettner, Jennifer King, Patricia Albright, Chris Woods, Jason Proctor, Sven Aas, who have participated in and supported the TEI Initiative;
- Jennifer King, Director of Archives and Special Collections, Alex Wirth-Cauchon, Director of RIS, Professor Nicole Vaget and Professor Kate Singer, who have participated in and supported the grant applications;
- Charlotte Slocum Patriquin, CIO, and Executive Director of LITS for her continued support of the TEI Initiative at MHC.

Shaoping Moss
Research and Instructional Support, LITS
10/10/11