



NATIONAL ENDOWMENT FOR THE

Humanities

OFFICE OF DIGITAL HUMANITIES

## **Narrative Section of a Successful Application**

The attached document contains the grant narrative and selected portions of a previously funded grant application. It is not intended to serve as a model, but to give you a sense of how a successful application may be crafted. Every successful application is different, and each applicant is urged to prepare a proposal that reflects its unique project and aspirations. Prospective applicants should consult the Office of Digital Humanities application guidelines at <http://www.neh.gov/grants/guidelines/digitalhumanitiesstartup.html> for instructions. Applicants are also strongly encouraged to consult with the NEH Office of Digital Humanities staff well before a grant deadline.

Note: The attachment only contains the grant narrative and selected portions, not the entire funded application. In addition, certain portions may have been redacted to protect the privacy interests of an individual and/or to protect confidential commercial and financial information and/or to protect copyrighted materials.

Project Title: Encoding Financial Records for Historical Research

Institution: Wheaton College

Project Director: Kathryn Tomasek

Grant Program: Digital Humanities Start-Up Grants Level I

### Abstract

The standard guidelines for scholarly markup of digitized sources, those of the Text Encoding Initiative (TEI), do not provide adequate models for representing the semantic value of financial records. Nevertheless, various digitization projects have used TEI-compliant XML to encode manuscript collections that include financial documentation. And now a handful of projects have begun to use TEI as they turn attention to financial records per se, revealing a need for extended markup guidelines to increase the accessibility of these resources. We will organize a meeting of historians, archivists, and technologists as a first step toward developing standards for markup of transcribed text and the application of metadata that will allow for searching across collections of manuscript financial records. Ultimately, the process begun with this meeting will lead to an extension of current TEI guidelines to include a module on financial records.

### Innovation

This project will begin the process of adding a new dimension to the Text Encoding Initiative, standards for markup and metadata for manuscript financial records. A genre of primary sources that includes such materials as bills and receipts, ledgers, and cashbooks, financial records are abundant in traditional archives, but current practices for digitization fail to capture their meanings as fully as possible. Our project is a next step in producing digital archives for the study of history.

### Statement of Humanities Significance

Over the long term, the process begun with the discussions proposed in this project will lead to standards for digitizing a rich yet currently inaccessible genre of manuscript historical records. By establishing standards for both markup of transcribed text and metadata that will facilitate searching across collections, this project will make possible the development of a rich digital archive that has the potential to open significant new lines of inquiry about economic and social history.

### Innovation in the Humanities

“[A]ccount books,” business historian Judith McGaw noted in 1985, “are among the most common but least accessible primary sources for historians....” In the intervening twenty-five years, historians have made use of these rich sources in a range of monographs, but few digitization projects have tackled financial records directly. The standard guidelines for scholarly markup of digitized sources, those of the Text Encoding Initiative (TEI), do not provide adequate models for representing the semantic value of such records. The activities to be funded by this Level I Start-Up Grant will initiate a community of practice focused on developing recommendations for encoding practices in order to improve access to digital collections of financial records from the eighteenth- and nineteenth-century United States. Drawing on the expertise developed by pioneering projects at the Massachusetts Historical Society, the University of Virginia, and Wheaton College (Massachusetts), we will describe the current state of affairs through discussion of problems and case studies, propose some paths forward, and identify model projects that can be fostered in order to test those paths. The final goal will be an extension of the current guidelines distributed by the Text Encoding Initiative for how digital humanities practitioners apply eXtensible Markup Language (XML) tagging to source materials. The proposed extension will include suggestions for both markup of transcribed text and the application of metadata that will allow for federated searching across collections.

Over the past twenty years, the TEI Guidelines have become the standard for detailed markup of the sorts of printed and manuscript texts—especially letters, journals, and diaries—that constitute the raw materials for literary, linguistic, and historical studies. Because the humanities community has adopted the TEI, various digitization projects have used TEI-compliant XML to encode manuscript collections that happened to include instances of such financial documentation as bills and receipts, ledgers, cashbooks, and accounts. To date, however, the percentage of such materials in the mass of material digitized has not been large enough to precipitate the creation of a TEI module designed to capture their unique semantic value. The most recent revision of the guidelines (TEI P5, 2007) contains only the barest of suggestions for encoding such significant historical sources (Appendix A). Luckily, a handful of projects have begun to use TEI as they turn attention to financial records per se, thus revealing an area where extended markup guidelines could increase the discoverability of these resources. The activities we propose will build on the natural constituencies that exist among archivists and historians who recognize the value of financial records for historical research to expand the community of scholars, archivists, and technologists who are prepared to stretch TEI Guidelines to broaden access to manuscript financial records.

Over the long term, we expect to contribute to the development of standards for both content tagging and metadata vocabularies. The latter will have the added value of improving federated searching, facilitating querying and analysis across collections. In the short term, this Level I Start-Up project will bring together historians, archivists, and technologists who will act as the nucleus of a growing community of practice that will expand access to rich sets of resources whose inaccessibility currently stands in the way of their use by scholars, students, and the general public. Facilitating digitization of such resources with attention to their particular semantic value has the potential to open significant new opportunities for research and teaching.

### Environmental Scan

Digital scholarship remains an emerging field in the discipline of history in the United States. Since Robert Darnton’s 1989 digital article, the *American Historical Review* has engaged in little further experimentation with the format aside from the innovative “The Differences Slavery Made,” by William G. Thomas III and Edward L. Ayers (2003). The digital journal *Common-Place* has offered an online site for publication about the early American republic since 2000. The 2010 annual meeting of the American Historical Association (AHA) included a few sessions focused on digital history, and the 2010 annual

meeting of the Society for Historians of the Early American Republic (SHEAR) included a roundtable on the “Digital Early Republic.” The 2011 AHA will include a poster session, an innovation that characterizes such meetings as those of the annual digital humanities conference sponsored by the Alliance of Digital Humanities Organizations.

The best known, most ambitious, and most successful project in digital history to date remains the Valley of the Shadow website that was produced at the University of Virginia between 1991 and 2007. This site includes numerous letters, diaries, and newspaper accounts that were transcribed and marked up using XML and the TEI. Data from federal decennial censuses were compiled in databases that can be searched on the website. No financial records were included in the project.

In fact, only a handful of digital projects have been centered on tabular records of any kind. A theatre history project, the Comédie Française Registers Project (CFRP <http://ndakinna.mit.edu/cfrp>), offers an example of an application of TEI to expand the usability of a set of records that share the tabular characteristics of financial records. The CFRP website demonstrates how XML markup enhances the information that had previously been gathered in the databases of the CESAR website (<http://cesar.org.uk/cesar2/>). Few digital history projects have focused directly on financial records, though numerous projects include them. The Railroads in the Making of Modern America Project at the University of Nebraska includes transcriptions of payroll records for a small number of individual railroad employees who worked in Knoxville, Tennessee, in August 1865 ([http://railroads.unl.edu/documents/search.php?rends\[\]=payroll&yearStart=&yearStop=&keyword=&publication=](http://railroads.unl.edu/documents/search.php?rends[]=payroll&yearStart=&yearStop=&keyword=&publication=)). The Bethlehem Digital History Project includes a few extracts from business ledgers ([http://bdhp.moravian.edu/community\\_records/business/busact.html](http://bdhp.moravian.edu/community_records/business/busact.html)).

Editions at the center of the project proposed here offer significant opportunities to direct attention specifically to financial records. At the Massachusetts Historical Society (MHS), the Adams and Winthrop digital editions include a substantial amount of financial material, as does Thomas Jefferson’s farm book. Without tagging that is specific to the unique value of that content, however, its significance is too easily lost to the researcher. At the University of Virginia, editors at the Papers of George Washington (PGW) have begun to turn their attention to the copious financial records that are part of the collection. Working on a much smaller scale, the Wheaton College Digital History Project (WCDHP) recently began transcription and coding of a daybook kept by Massachusetts businessman Laban Morey Wheaton between 1828 and 1859. The meeting that will be funded by this Level I Start-Up Grant will draw on the expertise that has been developed at the MHS, PGW, and WCDHP in our discussions. Appendix B contains images from these collections.

#### History and Duration of the Project

At Wheaton College, our desire to develop a community of practice focused on using TEI to represent financial records emerges from our work on a set of digitization projects that began in 2004, when a confluence of events combined new interest in and experience with TEI on our campus and the acquisition of the pocket diaries of Eliza Baylies Wheaton. Wheaton College collaborated with Mount Holyoke College to host a two-part conference that explored many of the ways that TEI can be utilized in teaching and research at liberal arts colleges. Subsequently, students worked with the principals to create digital editions of the diaries of Eliza B. Wheaton. In the spring semesters of 2009 and 2010, students in the Project Director’s section of the History Department’s methods course for History majors began to transcribe and encode the daybook of Laban Morey Wheaton. The Wheaton College Digital History Project has also benefited from cross-institutional collaborations through an IMLS Planning Grant to create a tool for online publication of TEI documents produced at small liberal arts colleges. In the project proposed here, we hope to build upon and expand those collaborations to include historians and archivists specializing in the early American republic as a way to expand the use of TEI in this field.

We hope that the activities of this Level I Start-Up project will lead to a Level II project that will develop plans for a TEI Special Interest Group focused on creating a TEI module designed to capture the unique semantic value of financial records. This proposed extension of current TEI guidelines will

include models for both markup of transcribed text and the application of metadata that will allow for federated searching across collections. Through the development of standards for digitization of financial records, we will contribute significantly to transformations of scholarship, teaching, and publication in the emerging field of digital history.

#### Work Plan

The activities to be funded through this Level I Start-Up grant consist of an intensive two-day meeting of historians, archivists, and technologists with expertise in the eighteenth- and nineteenth-century United States and in TEI (Appendix C). The meeting will be held at Wheaton College in spring 2011. In the interest of economy, invitations have been extended to scholars and archivists working in the Northeast and Mid-Atlantic regions. Participation will be opened beyond these regions at subsequent stages of the project.

Prior to the meeting, readings and sample materials will be made available to participants on a wiki to be set up by the Project Director and Staff at Wheaton College. All participants will be encouraged to contribute materials to the wiki, including additional readings, sample materials, and topics for discussion. The meeting will begin with discussions about uses and challenges of financial records in historical research, with examples brought forward from the wiki. Subsequent discussions will feature questions, challenges, and proposed solutions based on past experiences and current projects. We expect to talk about the benefits and drawbacks of the databases that have featured in many historical digitization projects in the past, with attention to the difficulties of searching across databases. Among the topics under consideration will be questions of when TEI is and is not appropriate for digitization of financial records. We will focus on articulating potential benefits and challenges in digitizing financial records through transcription and markup with TEI conformable XML. Our short-term goal will be to identify model projects for a next step that will move toward producing documentation and metadata standards, and examples from work underway with the Laban Morey Wheaton Account Books and at the Papers of George Washington will offer two possible models.

#### Staff

As conveners and hosts, the Wheaton College Digital History Project will undertake planning and arrangements under the direction of Associate Professor of History Kathryn Tomasek, with technical and archival advice provided by staff participants in the Wheaton College Digital History Project. Administrative services will be provided by two student assistants. Key participants in the proposed meeting include archivists from the American Antiquarian Society and the Massachusetts Historical Society; historians from the Papers of George Washington; TEI experts; and historians who specialize in the early nineteenth-century United States and use financial records in their research.

#### Final Product and Dissemination

Plans to draft the required white paper are built in to plans for the meeting, its preparation and follow-up activities (Appendix D). In addition, participants will propose papers to present the lessons learned from the grant activities at the following professional meetings: American Historical Association, Organization of American Historians, Society for Historians of the Early American Republic, Association of Documentary Editors, Digital Humanities, and TEI Consortium Members' Meeting. Papers will be proposed to additional conferences as appropriate. Articles will be proposed to the journals of the organizations that hold these meetings. A long term goal is to develop a TEI module on financial records. Once this module has been produced, workshops will be proposed to such venues as the Digital Humanities Summer Institute, the Digital Humanities Conference, and the TEI Consortium Members' Meeting to expand the community of practice through instruction.

Appendix C—Proposed Agenda for Spring 2011 Meeting at Wheaton College

First Day (Friday)

- 8:30 Coffee and pastries
- 9:00 Welcome, Opening Comments, Introductions  
Guiding Question for Introductions: How do historians/institutions hope to contribute to and benefit from the conversations?
- 9:30 Financial Records as Historical Sources  
Guiding Questions: How have you used financial records in your research? What do we know or imagine to be the uses of financial records in historical scholarship?  
What sorts of questions do we seek to ask of such sources?
- 10:15 Break
- 10:30 Technologies for Transcription and Analysis of Financial Records  
Guiding Questions: What technologies have you used for transcription and analysis?  
What challenges or successes have you encountered? Have you used databases and/or statistical programs? What features are (un)helpful? Is it possible to use XML/TEI like a database? What are the (dis)advantages?
- 12:00 Working lunch  
Short presentations from Massachusetts Historical Society, Papers of George Washington, Wheaton College Digital History Project
- 2:00 TEI Guidelines and Numbers  
Guiding Questions: What do the TEI guidelines say about representing numerical data? What are the unique semantic values of financial records? Can the TEI help us represent those values? In what sorts of cases is TEI an (in)appropriate choice for digitizing financial records?
- 2:45 Break
- 3:00 Markup  
Guiding Questions: What do we imagine markup for financial records will look like? What would we add to the TEI Guidelines in order to capture the unique semantic values of such sources? What features of financial records would we like to facilitate highlighting, describing, analyzing?
- 4:30 Close for the day

Second Day (Saturday)

- 8:30 Coffee and pastries
- 8:45 Metadata  
Guiding Questions: What sort of metadata would optimize searchability across collections? What would facilitate querying and analysis across collections?
- 10:15 Break
- 10:30 Model Projects for Next Steps  
Goal: Identify at least three model projects for next steps  
Points to consider:  
Models for transcription and coding  
Support necessary for such projects  
Ways to provide such support across institutions
- 12:00 Working lunch  
Discussion of process for drafting white paper and Level II proposal
- 2:00 Close

Appendix D—Work Plan

February 2011—

- Confer with participants on plans for meetings: KT/student assistants
- Set up wiki: KT, EPR
- Participants post pertinent readings, examples to Wiki
- Create assessment form using Survey Monkey: KT, EPR
- Participants plan details of presenting their experiences at meeting
- Confirm meeting space reservations: KT/student assistants
- Plan meals: KT/student assistants
- Confirm attendance, travel, and accommodation details: KT/student assistants

March 2011—

- Meeting with historians, archivists, and technological experts at Wheaton College
- All participants take part in writing up and circulating draft results using wiki
- One week after meeting, all participants invited to respond to assessment form created on Survey Monkey: KT/student assistants

April 2011—

- All participants use wiki to compose white paper and paper proposals for upcoming conferences

May 2011—

- All participants use wiki to comment on draft of white paper
- All participants use wiki to approve final draft of white paper

June 2011—

- KT submits completed white paper
- Appropriate participants complete planning for Level II grant application (projected due date Oct 2011)