



Journal of the Text Encoding Initiative

Issue 6 | 2013

Selected Papers from the 2012 TEI Conference

Encoding Financial Records for Historical Research

Kathryn Tomasek and Syd Bauman



Publisher
TEI Consortium

Electronic version

URL: <http://jtei.revues.org/895>

DOI: 10.4000/jtei.895

ISSN: 2162-5603

Electronic reference

Kathryn Tomasek and Syd Bauman, « Encoding Financial Records for Historical Research », *Journal of the Text Encoding Initiative* [Online], Issue 6 | December 2013, Online since 22 January 2014, connection on 30 September 2016. URL : <http://jtei.revues.org/895> ; DOI : 10.4000/jtei.895

The text is a facsimile of the print edition.

TEI Consortium 2013 (Creative Commons Attribution-NoDerivs 3.0 Unported License)

Encoding Financial Records for Historical Research

Kathryn Tomasek and Syd Bauman

AUTHOR'S NOTE

Initial discussions that led to the development of the model discussed in this paper were supported by a Level I Start-Up Grant from the Office of Digital Humanities at the National Endowment for the Humanities. The meeting funded by this award included Kurt Fendt, Julia Flanders, Scott P. Hamlin, Nancy Heywood, Ondine LeBlanc, Lauren Pfendner, E. Patrick Rashleigh, Jeff Ravel, Mary Beth Sievens, Jennifer Stertz, Zephorene L. Stickney, Jacqueline Wernimont, Ronald Zboray, and Mary Saracino Zboray, to whom the authors would like to extend their thanks and appreciation. Any views, findings, conclusions, or recommendations expressed in this paper do not necessarily reflect those of the National Endowment for the Humanities or the other meeting participants.

1. Introduction

[A]ccount books are among the most common but least accessible primary sources for historians...
(McGaw 1985)

- 1 Historical financial records (HFRs)—which can include such materials as bills, receipts, cashbooks, journals, and account ledgers—are abundant in analog archives and have generally been considered inaccessible and underutilized, though in the quarter century since the late Judith McGaw commented on the challenges such records pose, historians and other scholars have made use of these rich sources to produce a range of articles and monographs too numerous to list. At the same time, few digitization projects have tackled financial records directly until recently. Both the longstanding scholarly value of such

records and the current interest in their digitization suggest that the time has come to test potential models for markup and supporting software that might produce TEI-conformant files containing interpretive markup and thus available for comparative analysis across projects.

- 2 This paper describes a possible model for interpretive markup that attempts to express the relationships among individuals, goods and services, and cash or credit that are represented in double-entry bookkeeping. This model is presented as the result of a thought experiment that attempts to consider one possible way to take advantage of the expressive capabilities of TEI to produce contextual data about financial records in the interest of promoting humanities-centered research based on HFRs. Real-world markup examples are from the Wheaton College Digital History Project (WCDHP). A preliminary version of an ODD is currently available on the Encoding Historical Financial Records website (<http://www.encodinghfrs.org/>).
- 3 The authors invite readers to join the discussion in an online forum dedicated to the topic of markup for historical financial records, also available at Encoding Historical Financial Records.

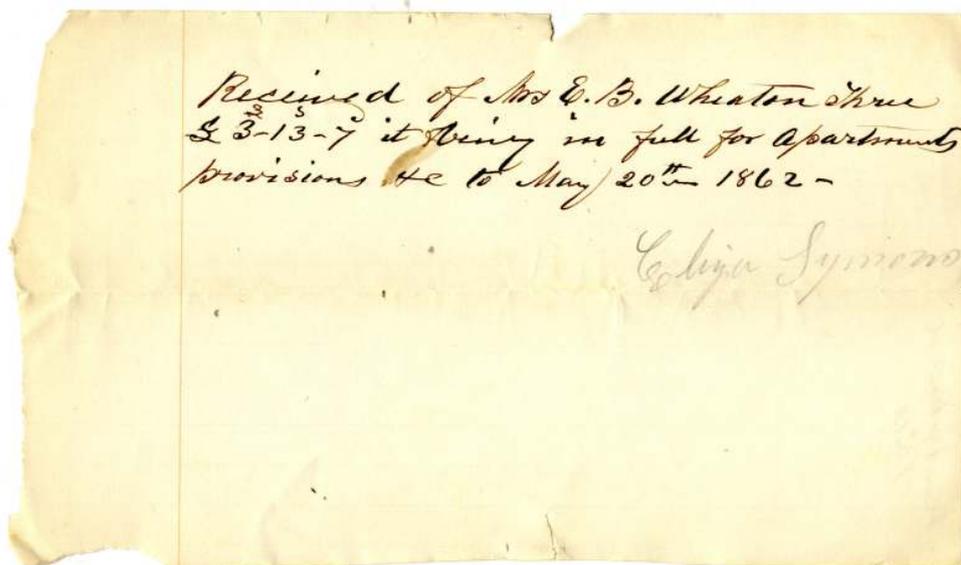
2. Historical Financial Records Defined

- 4 A genre of primary sources that includes but is not limited to such materials as bills, receipts, cashbooks, canceled checks, and the journals and ledgers associated with double-entry accounts, HFRs are part of a long history of records that humans have made to track their trades of goods and services, broadly defined. Archival collections also include bound volumes in which artisans, printers, or manufacturers recorded information about purchases of tools and materials of their trade as well as of wages they paid workers to help them produce such items as shoes and carriages, books and broadsides, or cotton and woolen textiles. Such records have long served as primary sources for historical studies of business and labor (see, for example, Hidy 1949, Dublin 1979, and McGaw 1987). Another form of HFRs, household accounts, refer to purchases of foodstuffs and clothing as well as to payments for services in the form of cooking, laundering, or other household labor. HFRs related to household management also include lists of wages paid for domestic labor, and such lists are sometimes found on the back pages of pocket diaries rather than in formal account books or ledgers. In regions where laborers were enslaved, HFRs can be found in the account books and personal records of slaveholders and of slave traders.¹ While a strong tradition of monographs and journals in economic history has resulted from investigation of archival financial records, no comparable project exists for their digitization. Over the past twenty years, however, HFRs have become as familiar a genre of primary sources to social historians as they are to historians of business and economics.
- 5 Merchants have kept financial records to track purchases and sales of goods since the commercial economies of ancient Mesopotamia, and the impulse to keep regular accounts produced standards for recording financial information. Through the centuries, various influential texts offered ordinary businessmen opportunities to learn how to do so. For early modern Europe, Fra Luca Pacioli's 1494 treatise *The Rules of Double-Entry Bookkeeping* outlined accounting principles that influenced commercial accounting throughout Europe. By the mid-eighteenth century, Scottish mathematician John Mair published an influential textbook entitled *Book-keeping methodiz'd; or, a methodical treatise of merchant-*

accompts, according to the Italian form. Presenting Pacioli's system for English speakers, Mair's text went through numerous editions between 1736 and 1808.

- 6 Thus in manuscript and print documents dating from at least the early modern period, merchants and other businessmen often made some effort to follow the rules set out by Pacioli and Mair. The account books and ledgers they created share certain structural characteristics with such other genres of historical records as plague bills, probate records, census returns, and theater returns. Such lists of information are often represented as tables, and in many cases they include numerical sums. Pre-printed ledger books, paper spreadsheets, and later computer spreadsheet programs were created to aid the regular recording and representation of such information.
- 7 But the apparent regularity of these documents presents perhaps the most significant challenge for those who seek to encode them, as it often collapses in use. HFRs are not always or necessarily tabular. For example, the WCDHP includes some HFRs in ephemera associated with a trip to England and Europe undertaken by Eliza Baylies Wheaton, her husband Laban Morey Wheaton, and his cousin and business partner David Emory Holman in the spring and summer of 1862. Among the ephemera are several boarding receipts from the journey. The receipts reflect boarding accounts that appear in associated notebooks, and while the daily accounts in the notebooks appear in roughly tabular format, the receipts comprise a simple prose record of an exchange of money for goods and services. The record appears nowhere else in the collection. A prose document of this type offers an example of an HFR that might be transcribed and marked up using existing elements of the TEI. Here, we have used the `<measure>` element, which can identify parts of the kinds of transactions that appear in HFRs.

Figure 1. Receipt for room and board, May 20, 1862. Wheaton Family Papers.



```

<text facs="ebweur1862_009a.jpg">
  <body>
    <p>
      <handShift new="#pers_WCDH1"/>Received of Mrs E.B. Wheaton Three
      <measure type="currency"> £ 3<hi rend="superscript-above">£</hi>-13<hi
      rend="superscript-above">s</hi>-7<hi rend="superscript-above">d</hi>
      </measure> it <unclear>&amp;lt;/unclear> in full for <measure
      commodity="room">apartments</measure>
      <measure commodity="board">provisions</measure> &amp;c to <date
      notAfter="1862-05-20">May20<hi rend="superscript">tt</hi> 1862.</date>
    </p>
    <signed>
      <handShift new="#pers_WCDH716"/>Eliza Symons</signed>
    </body>
  </text>

```

- 8 Even HFRs that were written in tabular format frequently include information that cannot be represented through simple transcription of that layout. In fact, they tend to contain significant variations and idiosyncrasies, often within the same document or collection. For example, Maine midwife Martha Ballard tracked births she attended as well as other elements of her herbal healing practice during the late eighteenth and early nineteenth centuries, and she noted payments—both in kind and in cash—in a diary. Historian Laurel Thatcher Ulrich decoded the diary's hand-drawn tabular format for her Pulitzer Prize-winning monograph (Ulrich 1990). The website (<http://dohistory.org/>) associated with the documentary film based on the book includes downloadable images of the diary pages and a transcription presented in tabular format (for example, see the transcription for 17 to 24 April 1785, http://dohistory.org/diary/1785/04/17850417_txt.html). But this simple transcription of numbers and text on a grid layout would require considerable manipulation to yield machine-readable data that might be compared to that transcribed from other HFRs, and Ballard's diary offers only one example of the rich data available for transcription in this genre. In a more recent project from the Massachusetts Institute of Technology, images of the printed forms used by the Comédie Française (<http://web.mit.edu/hyperstudio/cfr/index.html>) demonstrate that tabular forms offered space for recording additional types of information beyond the theater receipts listed on the printed rectos. Cast lists, for example, appear on versos in some instances. The wide variety of contexts in which HFRs broadly defined have been preserved point to the need for expressive markup, both of tabular records in general and of HFRs that include but are not limited to double entry accounts.

3. Research Questions

- 9 Research questions specific to this paper focus on whether there might be methods of markup for HFRs that would take advantage of the longstanding existence of TEI standards for markup of text, both print and manuscript, to promote creation of large corpora of accessible digital versions of HFRs that might be used for historical research. Interest in developing such an approach seems to be growing among scholars in Europe

and North America, and a meeting held at Wheaton College in August 2011 affirmed as much.

- 10 At the most basic level—that of the value of standardization of markup—longstanding projects in addition to the TEI affirm the utility of standardized electronic data extracted from primary sources. Examples of such data include the Minnesota Population Center's Integrated Public Use Microdata Series (<http://ipums.org/>), which provides harmonized data for United States and international population studies. Developments in geographic information have led to the creation of similar historical data sets for the United States and Europe: the National Historic Geographic System (<https://www.nhgis.org/>) and the Great Britain Historical Geographic Information System (<http://www.gbhis.org/>). A current initiative of the European Science Foundation, the European Historical Population Samples Network (<http://www.esf.org/index.php?id=8361>), seeks "to create a common format for databases containing information on persons, families and households." More recently in the United States, large projects have focused on creating big data from digitized texts, most notably the Historic American Newspapers (<http://chroniclingamerica.loc.gov/>) project of the Library of Congress as well as the Hathi Trust (<http://www.hathitrust.org/>) and the Digital Public Library of America (<http://dp.la/>). For the most part, the complexity of HFRs has meant that few scholars have considered their transcription and markup at comparable scale.
- 11 Some projects have displayed HFRs on websites without extensive discussion of opportunities for creating interoperable data that might be analyzed across projects. The Bethlehem Digital History Project (http://bdhp.moravian.edu/community_records/business/busact.html) displays images of a few extracts from business ledgers. The Railroads in the Making of Modern America (<http://railroads.unl.edu/views/item/rrwork>) at the University of Nebraska includes a few transcriptions of payroll records for railroad employees, and this information is stored in a searchable database. In the Visible Prices (<http://staff.washington.edu/paigecm/vp/>) website, literary scholar Paige Morgan is constructing a database of prices mentioned in English literature.
- 12 The issue of transcription and representation of HFRs as documents has arisen within some of the major documentary projects in the United States, and in some cases editors have used existing TEI elements to encode manuscript collections that happened to include instances of such financial documentation as bills and receipts, daybooks, ledgers, cashbooks, and accounts. At the Massachusetts Historical Society, for example, editors have used TEI in the creation of digital editions of the Adams Family Papers (<http://www.masshist.org/digitaladams/aea/about/transdetail.html>) and of a collection of documents authored by Thomas Jefferson, including his Farm Book (<http://www.masshist.org/thomasjeffersonpapers/farm/index.html>). And editors at the Papers of George Washington (<http://gwpapers.virginia.edu/project/index.html>) at the University of Virginia (PGW) have begun to transcribe the thousands of HFRs attributed to Washington or his agents. Since this latter project is concerned with a transcription method to be used for a large number of HFRs, these editors are developing a relational database for capture of data specific to the documents as documents. Currently in development, the tool will be available at doctracker.org (<http://doctracker.org/>). We note this research to emphasize that neither transcription nor methods for capturing information about HFRs as manuscript documents is the focus of this essay.
- 13 Instead, we wish to explore methods that might facilitate a goal suggested by editors of the Alcalá Account Book Project (<http://archives.forasfeasa.ie/index.shtml>), a digital

edition of the account books of the Royal Irish College of Saint George the Martyr in Alcalá. Such records promise, these editors note, "insight into the day-to-day running of the college with valuable information on diet, discipline, and domestic matters" (Alcalá Account Book Project 2008). We believe that if multiple projects were to create digital versions of HFRs that were accessible in machine processable form, scholars might explore numerous questions based on the information about daily life found within them.

3.1 HFRs and Daily Life in the Wheaton College Digital History Project

- 14 A small project focused on local records held in the Wheaton College Archives and Special Collections, the Wheaton College Digital History Project (WCDHP) emerged from work on a set of digitization projects that began in 2004, when a confluence of events combined new interest in and experience with TEI at Wheaton College 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 Tomasek's section of the History Department's methods course for history majors began to transcribe and encode the daybook of Laban Morey Wheaton. At crucial stages, the project received funding through the college from both the Andrew W. Mellon Foundation and the Mars Family Foundation. It is part of the community of practice involved in the TAPAS project (<http://tapasproject.org/>), which is developing a site for online publication of TEI documents, particularly aimed at those produced at institutions without programming support.
- 15 Since the HFRs being digitized in the WCDHP offer granular information about daily life in a rural town in Southeastern Massachusetts from the late eighteenth century through the early twentieth century, the sorts of research questions arising from these data that might interest historians include those associated with the genres of community study and microhistory as well as larger questions related to histories of social and economic development. The project has focused on two sets of financial documents so far: a daybook and ledger kept between 1828 and 1859, and receipts and other ephemera from a European journey in 1862. The former speak especially to research questions associated with scholarship focused on the nineteenth-century transition to industrial capitalism. The town of Norton, Massachusetts, shared many characteristics with other rural towns in the northeastern United States in the first half of the nineteenth century. Located in the agricultural hinterland of four significant seaports—Boston, New Bedford, Newport, and Providence—the town featured an economy that mixed agricultural pursuits with various manufacturing interests. The Wheaton family themselves owned a dairy herd as well as factories that produced cotton batting and straw hats. Over the past twenty years, historians have described such local economies in monographs based on representative archival research completed by individual scholars (see for example Prude 1983, Clark 1990, and Kelly 1999).

Figure 4. Pages 217 and 218 of L. M. Wheaton Daybook. Wheaton Family Papers.

- 18 This transaction was recorded as a credit to the debit side of Braman's account in Wheaton's ledger. That is, Wheaton recorded that he had extended credit to Braman in the amount of \$14.75, to be paid at a later date, and we see that Braman settled his account with Wheaton on January 1, 1850.
- 19 In this way, Braman's account demonstrates Wheaton's adherence to the principles of double entry accounting as prescribed by Pacioli and Mair. Other accounts in Wheaton's daybook and ledger do not match up so neatly, and Wheaton is by no means the only merchant who kept less than perfect books. The popularity of double entry accounting could not guarantee perfect adherence to its principles. Human use of standards of any sort is subject to human error and idiosyncrasy. Nevertheless, use of double-entry accounting has persisted over several centuries, so developing guidelines for its markup seems prudent.

5. Approaches

- 20 Since this paper presents a thought experiment that one small project is in the process of exploring, we do not present here an exhaustive review of the many models for electronic collection and use of contemporary financial records that have been developed since the

mid-twentieth century. Our colleagues at the PGW have explored the literature on these models in depth, and in correspondence they have noted that the theory behind the models has not changed significantly over the past forty years (Everest and Weber 1977; Wigley n.d.; Ijiri 1975; McCarthy 1982; Wang, Du, and Lee 2002). In any case, contemporary personal financial accounting software is not designed for humanities-based enquiry into historical financial records.

- 21 Indeed, the work both of the late Judith McGaw and of literary historian Mary Poovey suggests that humanities scholars would do better to take a step back from software or accounting theory that claims simple description or transparent representation of the exchanges recorded in HFRs. McGaw argued for understanding accounting practices as historical, noting that while the accounting practices of the nineteenth-century paper manufacturers she studied hid capitalization, the manufacturers adopted new accounting practices as their needs for understanding the operations of their businesses changed. When they began to mechanize the production of paper, mercantile accounting allowed manufacturers to keep track of working capital. As machine technologies changed and manufacturers began to need to consider alternatives, they adopted historical and standard cost accounting. Historical interpretations that claimed accounting practices failed to keep pace with technological change misconstrued the manufacturers' adaptations and use of practices appropriate to their needs (McGaw 1985). Accounting itself was for McGaw a technology, a tool for interacting with the world, and manufacturers used it consciously. Thus projecting contemporary accounting practices back onto business practices in the past entails interpretive assumptions of which the scholar should be aware.
- 22 In a similar vein, Poovey has argued that the seeming descriptive transparency of double-entry bookkeeping resulted from cultural processes of the seventeenth and eighteenth centuries that separated so-called scientific description from humanistic interpretation. For Poovey, double-entry bookkeeping was "the prototype of the modern fact", a method that emphasized the system of recording accounts over the particulars of any specific transactions and thus allowed the merchant to conceive of his business as a whole, abstracted from the actual stock or cash that he might have on hand (Poovey 1998). If HFRs thus bear within them cultural assumptions about the human interactions they purport to describe, humanities-based research might well be better served by markup grounded in standards developed for humanities research. In this context, since TEI is the de facto international standard for markup of extant print and manuscript sources for humanities research, TEI-compatible markup of data from manuscript financial records makes sense.
- 23 If double-entry bookkeeping thus gave merchants a way to emphasize a system of accounting over the particulars of individual transactions, that system, we would argue, offered a method of abstraction that we might see as analogous to the sort of systematic abstraction that computer scientist Jeannette Wing has identified as the great advantage of computational thinking (Wing 2006). One considerable challenge for modeling a system of markup that will allow scholars to take advantage of the expressive capabilities of TEI-compatible XML lies in distinguishing between the variety of abstraction enabled by the system of double-entry bookkeeping and that allowed by expressive markup.
- 24 Thus, HFRs tend to include three levels of data to consider: layout, textual expression, and a third, more abstract level of financial semantics that are not as yet easily captured through TEI-conformant markup. We anticipate that decisions about representation of

layout will be made on a project-by-project basis. In cases where page images are included in online publication, for example, some projects may choose to omit digital representation of layout. Similarly, different projects place varying emphasis on particular textual features.

- 25 Presuming that the scholar or project involved is interested in features of the documents such as handwriting, pagination, orthography, and apparent errors, it is reasonable to believe that there will be a TEI encoding of the document either *as a logical text* (using `<text>`) or *as a document* (using `<sourceDoc>`), but in any case not as a database of information about the transactions themselves. Several different methods of adding the who, what, where, and how much of the financial transactions themselves to the TEI encoding are readily evident. These can be split broadly into those that incorporate the added information directly into the TEI-encoded file, and those that record the financial information externally to the TEI-encoded file that represents, for example, the manuscript that attests the transactions.
- 26 The authors make no claim that this list is exhaustive, and we welcome discussion of other options.

Incorporated with textual encoding

- **TEI**
 - Using in-line TEI markup to indicate the quantities of goods and currencies that changed hands (for example, with `<measure>`), which entries are related to one another (for example, with `@corresp`), and who the involved parties are (for example, with `<persName>`). This may or may not require extension of the current Guidelines (TEI Consortium 2012).
- **non-TEI**
 - It is possible that some other markup language, either pre-existing or newly invented, might be used for this same purpose.

Separate from textual encoding

- **contemporary accounting or personal finance software**
 - Keeping track of transactions and accounts in software designed for the purpose, for example GnuCash (<http://www.gnucash.org/>), iBank (<http://www.iggsoftware.com/ibank/>), or Quicken (<http://quicken.intuit.com/>).
 - **relational database or spreadsheet**
 - Keeping track of transactions and accounts in general-purpose relational or flat data management software, for example PostgreSQL (<http://www.postgresql.org/>) or LibreOffice Calc (<http://www.libreoffice.org/>).
 - **"transactionography"**
 - Keeping track of transactions and accounts in a (customized) TEI XML structure.
- 27 It is quite possible, of course, that a scholar or project is interested only in the financial transactions, and not at all in the documents that attest them. We do not consider these cases here, nor do we consider the in-line methods above, which Tomasek expects to address in a future publication with Ondine LeBlanc and Nancy Heywood. Rather, here we will briefly discuss the other stand-off methods above, and then describe the "transactionography" approach in much more detail.

5.1 Other Stand-Off Approaches: Contemporary Personal Financial Accounting Software

- 28 At first blush, contemporary accounting software or personal finance software, especially those programs explicitly designed to handle double-entry bookkeeping, would seem to be a natural fit for recording the transactions attested in historical documents. On a cursory examination, however, several reasons this is not a good fit include the following:
- **dates**
 - Typically, contemporary personal financial accounting software will reject dates from the nineteenth century, let alone earlier.
 - **units**
 - In a quick search, we did not find contemporary personal financial accounting software that handled pounds, shillings, and pence, let alone denarii.
 - **attestation**
 - No obvious mechanism exists to provide a link between a transaction as recorded in contemporary personal financial accounting software and the appropriate portion of the encoding of the document that attests it.
 - **certainty**
 - Contemporary personal financial accounting software does not provide the capability to formally express uncertainty, discrepancies, or alternate possibilities.
- 29 There are possible ways ("hacks") around some of these problems even without rewriting the software, and some software is open source. For example, some software may permit dates in the *future*, in which case a project might record eighteenth-century dates using values from the twenty-eighth century. Pounds, shillings, and pence could be converted to a modern standard unit (for example, US dollars; Measuring Worth provides calculators for determining values in some historical currencies). A project could consistently use the 'comments', 'memo', or 'notes' field to record pointers into the attesting document.
- 30 But even with these hacks, contemporary personal financial accounting software is not going to permit formal expression of "the amount in the ledger does not match the amount in the day book", or "this may have been paid to Tim Warren or to Liz Warren", or "amount is illegible".
- 31 It is possible that modern enterprise financial accounting software (e.g., SAP ERP, <http://www54.sap.com/pc/bp/erp.html>) does have these capabilities, but we think it unlikely, and a recent chat session with SAP support² would seem to confirm this.

5.2 Other Stand-Off Approaches: Contemporary Tabular Data

- 32 One can well imagine trying to store the "tabular" historical financial data in a modern spreadsheet or MySQL table or similar. And in the case in which it is only a simple representation of this financial information that is of interest, such an approach is likely appropriate, and may even be preferable to a TEI encoding. But recording uncertainty, ambiguity, and references to attesting documents will likely prove difficult (although not impossible).

- 33 TEI, of course, already provides mechanisms for representing uncertainty and ambiguity, and a mechanism we adopt here for referencing encodings of the attesting documents. It is our belief that consistent use of such mechanisms across a project (rather than using TEI mechanisms for the primary source documents and a different ad-hoc mechanism for the table that represents the transactions) will prove beneficial.

6. Transactionography

- 34 In some sense, the financial records that attest a transaction refer to said transaction in a manner parallel to the way a name of a person or place that appears in running prose refers to said person or place. That is, there exists a real-world thing or action that is being referred to by the words (or other marks) in a document we are encoding.
- 35 For people and places, the TEI provides a data structure for recording information about the real-world person or place, and a mechanism to associate a reference in the encoded source document to that data structure, thus asserting that the real-world person or place described by the data structure is the one that was referred to in the source document.
- 36 Why not do something similar for transactions? Develop a data structure for recording transactions, and link that data structure with the segments of running prose or apparently tabular information that attest it or refer to it. This seems like a novel and labor-intensive approach, but nonetheless one worth serious consideration. In order to develop a suitable data structure, one needs to model *transactions*.

6.1 Transaction Model

- 37 Our model of a transaction is perhaps somewhat more inclusive than some dictionary definitions. We think of a transaction as a coherent set of one or more transfers of something of perceived value from one entity to another. A transfer has three main components, which may be summarized as 'what', 'from whom', and 'to whom'.³ Furthermore, the 'what' is likely to be divisible into a certain amount (that is, a quantity and a unit) of a given commodity. It is worth noting that transfers take place at a certain point in time, even if we do not know when that time was. Or, at least, transfers are completed at a certain point in time. A transfer may occupy a significant duration of time from start to finish. For example, a transaction may be conducted by ground postal service.
- 38 Some common transaction categories can be identified.
- A standard exchange of money for goods is a purchase, consisting of two transfers: for instance, I transfer \$2 to the convenience store, and a tiny little bottle of apple juice is transferred from the store to me.
 - A similar trade that does not involve currency is a barter, for example, trading a large red paper clip for a fish-shaped pen.⁴
 - A gift is a single uni-directional *voluntary* transfer; a single uni-directional *involuntary* transfer is called theft or embezzlement.⁵
 - A set of transfers among more than two entities may be referred to as a multilateral trade.

- 39 To model the 'what', the TEI `<measure>` (or `<measureGrp>`) element seems tailor-made for the purpose. To model the 'when', the attributes from the TEI `att.datable` class seem to be quite reasonable: they can handle specific dates, times, date ranges, and particular dates that took place sometime within a range.⁶ However, to model the 'from' and 'to' in the TEI universe is a bit problematic, as `@from` and `@to` already exist in the `att.datable.w3c` class, which we would like to use for modeling the 'when' (via the `att.datable` class). It may make sense to use the words 'debtor' and 'creditor' as the names of attributes; however, these words mean something somewhat different in the context of double-entry bookkeeping when accounts, not people or organizations, are debited and credited. So, at least initially, the authors have shied away from these words. Mary Beth Sievens suggested the Norwegian words for 'from' and 'to', namely 'fra' and 'til'. These were used in the first draft ODD for transactionographies, and are still our current choice.
- 40 The `@fra` and `@til` attributes represent the person, organization, or account from which or to which something is being transferred by providing one or more pointers (via the TEI `data.pointer` datatype) to an XML structure describing said person, organization, or account. Typically that XML structure will be an entry in a TEI prosopography, "orgography", or "transactionography".

6.2 Sample Transfer

- 41 With assumptions listed below in place, we can discern from the following encoding that Abigail Webber obtained five pounds of lard from Laban Morey Wheaton at Wheaton's general store.

```
<hfr:transfer xmlns:hfr="http://www.wheatoncollege.edu/tei-extensions/
financialRecords/1.0" fra="persons.xml#WCDH2" placeRef="places.xml#W2GS"
source="daybook.xml#d1e363 daybook.xml#d1e393" til="persons.xml#WCDH633">
  <measure commodity="lard" quantity="2.268" unit="kg"/>
</hfr:transfer>
```

- 42 Assumptions:

- The prefix `hfr` is bound to `http://www.wheatoncollege.edu/tei-extensions/financialRecords/1.0`.
- The file `persons.xml` is in the same directory as the file in which this transfer appears, and it has a `<person>`, `<org>`, or `<hfr:account>` element with an `@xml:id` of "WCDH2" (ostensibly a `<person>` describing Laban Morey Wheaton).
- The file `persons.xml` has a `<person>`, `<org>`, or `<hfr:account>` element with an `@xml:id` of "WCDH633" (ostensibly a `<person>` describing Abigail Webber).
- The file `places.xml` is in the same directory as the file in which this transfer appears, and it has a `<place>` element with an `@xml:id` of "W2GS" (ostensibly describing the place where Laban Morey Wheaton's general store was based; for example, Norton, MA).
- The file `daybook.xml` is in the same directory as the file in which this transfer appears, and it has XML elements with `@xml:ids` of "d1e363" and "d1e393" (the

daybook typically lists transactions chronologically; each might be encoded as an `<item>`, `<row>`, `<p>`, `<ab>`, or other element).

6.3 Formal Model of a Transfer

- 43 Using RELAX NG compact syntax we can express the model of a TEI transfer as follows.

```

element transfer = {
  att.global,
  att.dateable,
  att.transferable,
  ( model.glossLike | model.measureLike | model.global | attested )+
}
att.transferable =
  att.editLike.attributes,
  att.transferable.attribute.fra,
  att.transferable.attribute.til,
  att.transferable.attribute.placeRef

```

6.4 Sample Transaction

- 44 The following encoding asserts that a person (ostensibly named 'Calvin') purchased a 2-oz. candy bar (a Snickers®) for \$1.25 on Monday, October 17, 1988.

```

<hfr:transaction xmlns:hfr="http://www.wheatoncollege.edu/tei-extensions/
financialRecords/1.0" when="1988-10-17">
  <hfr:transfer fra="orgs.xml#candy-counter" til="persons.xml#calvin">
    <measure commodity="candy" quantity="58.7" unit="g">1 Snickers bar</
measure>
  </hfr:transfer>
  <hfr:transfer fra="persons.xml#calvin" til="orgs.xml#candy-counter">
    <measure commodity="currency" quantity="1.25" unit="USD"/>
  </hfr:transfer>
</hfr:transaction>

```

6.5 Ordering or Alternates: `<transferGrp>`

- 45 Some common situations that would be hard to handle with the encoding so far described jump quickly to mind. Several of these can be easily handled with the addition of a grouping element to hold a set of `<transfer>`s.
- 46 For some transactions, the order of transfers is known and interesting.⁷ If we wish to represent the chronological order of transfers, this could be handled by using the `@when` (or other dating attributes) on each transfer. However, it is likely to be the case that although the order is known, specific dates are not.

- 47 Since we are describing historical (rather than contemporary) events for which we may have insufficient or contradictory documentary evidence, it is quite possible that some aspects of a transaction will be known and others will not. One common case, for example, occurs when a transaction is recorded with only a partial name of one of the involved parties, leaving it to the scholar to determine which of several possible people was the one involved. Another is when a particular detail of a transfer cannot be reconstructed, for example because the manuscript is illegible or a page is missing. A third similar problem occurs when a single transfer is recorded in different ways in different sources. For example, if a store's records indicate an item was sold for one price, but the purchaser's records indicate he paid a different price. Or, for instance, if a merchant's ledger indicates one price, but his daybook another.
- 48 Many, if not most, of these problems could be handled through the application of existing TEI technologies. For example, if a Washington, DC, vendor sold some wax to 'Bell', we might not know whether it was purchased by Alexander or Chichester. We could indicate this ambiguity by referring from the `@fra` or `@til` attribute to an `<alt>` element, which itself points to each of the two possible involved parties. Alternatively, two separate `<transfer>` elements could be used, the same except for referring to a different person involved. Each of these could bear an `@exclude` attribute indicating the other.
- 49 However, `<alt>` and `@exclude` are seldom used in the TEI community, and some mechanism needs to be developed for handling relative chronological ordering when absolute chronological information is not available. We think a good way to solve all of these problems is to introduce a grouping element that contains a set of `<transfer>`s and asserts whether the transfers are mutually exclusive possibilities, chronologically ordered, or neither. We believe that if this grouping element is allowed to self-nest, arbitrary combinations of chronologically ordered, partial, or ambiguous transfers can be represented.⁸
- 50 Thus we introduce the `<transferGrp>` element, a grouping element somewhat like `<surfaceGrp>` but, like `<altGrp>` or `<attList>`, with the ability to assert the relationship among its children rather than just common features. This is accomplished with an `@org` attribute whose suggested values include:
- **"group"**
 - an unordered set
 - **"sequence"**
 - a (chronologically) ordered set
 - **"choice"**
 - only one of the child `<transfer>`s obtained
- 51 The following example demonstrates the use of `<transferGrp>` to indicate a transaction which is attested in (at least) two source documents which disagree as to the amount of money involved.

```

<hfr:transaction xmlns:hfr="http://www.wheatoncollege.edu/tei-extensions/
financialRecords/1.0" when="1982-07-09">
  <hfr:transfer fra="persons.xml#A" til="persons.xml#B">
    <measure commodity="beer" quantity="99" unit="bottle"/></hfr:transfer>
  <hfr:transferGrp fra="persons.xml#B" org="choice" til="persons.xml#A">
    <hfr:transfer source="daybook.xml#P17L14">
      <measure commodity="currency" quantity="125" unit="USD"/></
hfr:transfer>
    <hfr:transfer source="ledger.xml#P08L12">
      <measure commodity="currency" quantity="150" unit="USD"/>
    </hfr:transfer>
  </hfr:transferGrp>
</hfr:transaction>

```

7. Expressive Capabilities

- 52 Although no full real-world examples yet exist, we believe that a transactionography is inherently flexible, and thus can be used to represent situations beyond simple transactions like purchases, barter, and gift-giving.
- 53 For example, the extension of credit can be represented in a transactionography by the transfer of a commodity of indebtedness. We have been using the keyword "iou" as the value of @commodity of <measure> for this purpose.⁹
- 54 But in many situations, including transactions recorded using most double-entry accounting systems, transfers are not made directly to or from a person or organization, but rather are buffered through accounts which are themselves associated with a person or organization. For example, when your bank pays you interest, it does not send money to you, rather it puts the money into *your account*.
- 55 Transaction entries, as conceived here, represent the entity from which and to which something (for example, currency or a candy bar) was transferred by pointing to an XML object that gives us information about the giving or receiving entity. Correspondingly, in order for a transactionography to represent an account, we would like an (XML) object to refer to, one that would give us information about the account. For this we have developed the <hfr:account> element, and the <hfr:listAccount> element with which to hold a logical set of accounts. (For example, all those belonging to the same person or institution, or all those measured in EUR.) An <hfr:account> provides a mechanism for describing a "registry of pecuniary transactions; a written or printed statement of business dealings or debts and credits,"¹⁰ including bank accounts. We have not yet worked out all the details of how metadata pertaining to a given account would be recorded, however. Rather, for the moment, the content permits members of both the TEI model.labelLike and model.glossLike classes, as well as the <idno> element, for providing loosely structured information about the account.
- 56 Furthermore, it is worth noting that the usual TEI mechanisms for expressing doubt and uncertainty, for providing analysis, or for describing manuscript revisions are applicable. For example, the element might be used to indicate that a series of entries in a ledger occurred in a particular context:

```
<span from="#entry_1842-05-25.01" to="#entry_1842-05-27.34"
type="interpretive"> These
  represent a flurry of activity that is probably due to the upcoming
county fair which
  was held on Saturday 28 May 1842. </span>
```

8. Limitations

- 57 While we are confident that our system of representing the transactions attested by historical financial records is robust, we do not wish to claim that it is necessarily the best way to record these. There are some obvious problems. First and foremost, the encoding is dense, abstract, and full of pointers. Thus, significant software support is required if humans are to accurately record and proofread any more than a handful of transactions.
- 58 Moreover, significant software support will also be required to process these records in any meaningful way. And while certain kinds of activities are obvious and would seem relatively easy to support (for example, "show me all transactions involving gunpowder"), we are quite frankly not sure how much the research capabilities provided by this sort of encoding will align with the kinds of questions historians might wish to ask of this sort of data. The need for such software support (in both creating and processing transactionographies) and limited resources to hire programmers to provide it may make transactionographies too unwieldy for widespread use.
- 59 That said, we think this representation cleanly separates information about the transactions from information about the documents that attest them, while allowing the two to be explicitly linked. And we believe that this system is very robust and flexible, and will allow a useful representation of the vast majority of historical financial records. However, we have discovered one shortcoming of the representative system we are employing. In many cases one of the "items" being transferred is a service, not a commodity. Our current system, being based on the TEI `<measure>` element, seems a clumsy way to handle this. For example,

```
<measure commodity="babysitting" quantity="2" unit="hours"/>
```

may be reasonable, but when the service being provided is recorded either by things *on* which it is performed or the people *for* whom it is provided, rather than the amount *of the service* that is provided, it becomes difficult to express formally using the current system. Take, for example, this laundry list from 1862.

Figure 5. Receipt for Mrs. Wheaton's Wash, May 19, 1862. Wheaton Family Papers.

May 19 - Mrs. Wheaton's	
Return by Saturday P.M.	
or before	
2 wool skirts	6
1 " vest	3
2 " hose	2
1 cot. hose	1
2 Bosom shirts	4
1 night shirt	3
1 chemise	2
1 waist	2
1 cotton drawers	2
1 Dinner Napkin	1
2 Towels	2
1 Little do	1 ¹ / ₂
2 hkerfs	1 ¹ / ₂
1 half hkerf	1
1 pr. gloves	1
5 gent's collars	5
1 Ladies' do.	1
2 pr. cuffs	2
5 Laces	4
33 piers	39
Old Bill	2-5
	6-2

- 60 This is a list of various items of clothing, each associated with a price. But the price is not the purchase price, but rather the price Mrs. Fearn charged Mrs. Wheaton for laundering the listed articles of clothing. Here is a sample TEI transcription of a portion of this receipt.

```

<item>
  <measure commodity="skirt" quantity="2" unit="count">2 wool skirts</measure>
measure>
  <measure commodity="currency" quantity="6" unit="pence">6</measure>
</item>
<item>
  <measure commodity="vest" quantity="1" unit="count">1 " vest</measure>
  <measure commodity="currency" quantity="3" unit="pence">3</measure>
</item>
<item>
  <measure commodity="hose" quantity="2" unit="pair">2 " hose</measure>
  <measure commodity="currency" quantity="2" unit="pence">2</measure>
</item>

```

- 61 The fact that the 6 pence is the price of *laundering* the 2 wool skirts is not clear, here. And because we use the `<measure>` element similarly in the transactionography the same confusion obtains. We have yet to seriously address this question, but hope to do so in the future.

9. Conclusion

- 62 HFRs contain a range of information that is of interest to scholars in a variety of fields. Therefore, we see significant scholarly potential in their markup. For small collections or so-called craft projects (like the WCDHP), local records can form the foundation of a case study or microhistory that can illuminate larger historical questions. Of course such studies could contribute to scholarship as the raw data for traditional monographs. But beyond such uses, these encoded HFRs could form the basis of a variety of digital projects based on the data, including analysis and visualizations. Less likely, but even more exciting, is the prospect that multiple such projects might create harvestable, shareable data that could be aggregated and analyzed across multiple projects. Even within a single large project, the possibility of making interesting connections between or inferences from information about historical transactions that might open new lines of inquiry for economic, social, and cultural history is intriguing.
- 63 We have developed and presented here the basis for one possible encoding system for the semantics of HFRs, the so-called "transactionography". While we realize that the system is complex, we are comforted by the realization that HFRs are inherently complex, and no simple system is likely to reflect their intrinsic nuances. We are not confident that this system is the right mechanism for TEI and the digital historical community to adopt, but we are sure it deserves serious consideration.
-

BIBLIOGRAPHY

- Clark, Christopher. 1990. *The Roots of Rural Capitalism: Western Massachusetts, 1780-1860*. Ithaca, NY: Cornell University Press.
- Dublin, Thomas. 1979. *Women at Work: The Transformation of Work and Community in Lowell, Massachusetts, 1826-1860*. New York: Columbia University Press.
- Everest, Gordon C., and Ron Weber. 1977. "A Relational Approach to Accounting Models." *The Accounting Review* 52, no. 2: 340-59.
- Hidy, Ralph W. 1949. *The House of Baring in American Trade and Finance: English Merchant Bankers at Work, 1763-1861*. Cambridge, MA: Harvard University Press.
- Ijiri, Yuji. 1975. *Theory of Accounting Measurement*. Studies in Accounting Research 10. Sarasota, FL: American Accounting Association.

- Kelly, Catherine E. 1999. *In the New England Fashion: Reshaping Women's Lives in the Nineteenth Century*. Ithaca, NY: Cornell University Press.
- Mair, John. 1765. *Book-keeping methodiz'd; or, a methodical treatise of merchant-accompts, according to the Italian form. Wherein the theory of the art is fully explained, . . . To which is added, a large appendix. . . .* 8th ed. Edinburgh: W. Sands, A. Kincaid & J. Bell, and A. Donaldson. Reprint, Gale ECCO Print Edition (June 10, 2010).
- McCarthy, William E. 1982. "The REA Accounting Model: A Generalized Framework for Accounting Systems in a Shared Data Environment." *The Accounting Review* 57, no. 3: 554–78.
- McGaw, Judith A. 1985. "Accounting for Innovation: Technological Change and Business Practice in the Berkshire County Paper Industry." *Technology and Culture* 26, no. 4 (October): 703–25.
- McGaw, Judith A. 1987. *Most Wonderful Machine: Mechanization and Social Change in Berkshire Paper Making, 1801–1885*. Princeton, NJ: Princeton University Press.
- Pacioli, Luca. 1494. *Particularis de computis et scripturis, Summa de arithmetica, geometria, proportioni et proportionalita*. Venice, Italy. Translated by Michael Schemmann as *The Rules of Double-Entry Bookkeeping* (IICPA [International Institute of Certified Public Accountants Incorporated] Publications, 2010).
- Poovey, Mary. 1998. *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago: University of Chicago Press.
- Prude, Jonathan. 1983. *The Coming of Industrial Order: Town and Factory Life in Rural Massachusetts, 1810–1860*. New York: Cambridge University Press.
- TEI Consortium. 2012. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.2.0. Last updated October 25. N.p.: TEI Consortium. <http://www.tei-c.org/Vault/P5/2.2.0/doc/tei-p5-doc/en/html/index.html>.
- Thomas, William G., III. 2004. "Computing and the Historical Imagination": *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell Publishing.
- Ulrich, Laurel Thatcher. 1990. *A Midwife's Tale: The Life of Martha Ballard, Based on Her Diary, 1785–1812*. New York: Knopf.
- Wang, Ting J., Hui Du, and Hur-Li Lee. 2002. "A User-Oriented Approach to Data Modeling: A Blueprint for Generating Financial Statements and Other Accounting-Related Documents and Reports." *The Review of Business Information Systems* 6(4): 17–32.
- Wheaton Family Papers*. Wheaton College Archives and Special Collections, M089. Wheaton College, Norton, MA.
- Wigley, Michael. *Double Entry Accounting in a Relational Database*. http://homepages.tcp.co.uk/~m-wigley/gc_wp_ded.html. Accessed August 15, 2013.
- Wing, Jeannette M. 2006. "Computational Thinking." *Communications of the ACM* 49 (3): 33–35.

NOTES

1. Use of computational methods for exploring the history of slavery in the United States led to considerable controversy over the work of Robert Fogel and Stanley Engerman in *Time on the Cross: The Economics of American Negro Slavery* (1974). As the historian William G. Thomas III has noted, the negative impact of that controversy persisted into the mid-1990s, when several authors took pains to distinguish the use of computers in the history from quantification and so-called "social science methods". Thomas notes numerous additional developments in the use of computers in historical study, both in the United States and elsewhere between the 1940s and the early twenty-first century (Thomas 2004).
2. Chat session between co-author and SAP initiated on August 12, 2013, by choosing Chat Now menu item from https://store.sap.com/sap/cpa/ui/resources/store/html/Solutions.html?pcntry=US&sap-language=EN&catID=PACALL&_cp_id=id-1376333722684-0#All
3. It has been suggested that CDOC CRM is likely to offer a helpful model of transfer of an object from one entity to another. The authors hope to investigate this possibility at a future date.
4. Brian Bergstein, "Blogger's Quest Ends with Keys to House", *NBC News*, <http://www.msnbc.msn.com/id/13804920/>, accessed October 29, 2012.
5. We leave it to the reader to determine into which category taxation falls.
6. We anticipate that some users will want more flexibility in formal representation of approximate dates, for which the LOC recommendation (<http://www.loc.gov/standards/datetime/pre-submission.html>) could be employed using the TEI `att.datable.custom` class.
7. Ask any forensic accountant.
8. Proof is left for future work. So far all we have done is quickly tried, and failed, to find an obvious counter-example.
9. In English, indebtedness is often acknowledged using an informal commitment or document called an 'IOU', which is derived from the phrase "I owe you".
10. *Webster's Revised Unabridged Dictionary*, 1913, s.v. "account," accessed October 29, 2013, <http://machaut.uchicago.edu/?action=search&resource=Webster%27s&word=Account&quicksearch=on>.

ABSTRACTS

This paper focuses on a thought experiment in which the authors propose an encoding system not unlike the contextual markup of prosopographies or gazetteers using TEI P5 that could be used to capture the particular financial semantics involved in HFRs. This possible system is expressed as a TEI "extension" customization, and is flexible enough to encode the subgenre of financial records called "double-entry accounts", a bookkeeping method for recording transfers of goods or services in exchange for currency or credit.

Real-world markup examples are from the Wheaton College Digital History Project. A preliminary version of the ODD is currently available on Encoding Historical Financial Records (<http://www.customization.encodinghfrs.org/>).

INDEX

Keywords: historical financial records, double-entry accounts, transactionography, TEI guidelines, TEI ODD, TEI customization