Modeling semantically Enhanced Digital Edition of Accounts (MEDEA) for Discovery and Comparison on the Semantic Web

Kathryn Tomasek, Anna P. Orlowska, Kathrin Pindl, Cliff Anderson, Øyvind Eide, Georg Vogeler

Abstract

Historical accounting documents are a genre of texts that have considerable research potential if we treat them as humanities sources. MEDEA is a cooperative international project whose principal investigators recommend creating digital scholarly editions of accounts as a first step in a process that will open the information contained within them to the affordances of the Semantic Web. MEDEA researchers are at work on a bookkeeping ontology that can be used to intermediate between XML markup and exposing Linked Open Data as RDF. The information contained in the texts of accounts can then be used to explore humanities questions at levels from the granular or local to the regional or global. This paper reflects presentations from a multi-speaker session at DH2016 in which MEDEA participants discussed the kinds of humanities information found in accounts, the forms of electronic representation available for working with them, and an evolving bookkeeping ontology based on CIDOC-CRM.

Introduction

Historical accounting documents are a genre of texts that have considerable research potential if we treat them as humanities sources. We address questions about how to do so through a cooperative international project called MEDEA (Modeling semantically Enhanced Digital Edition of Accounts). Our principal investigators--Tomasek, Pindl, and Vogeler--recommend creating digital scholarly editions of accounts as a first step in a process that will open the information contained within them to the affordances of the Semantic Web. MEDEA researchers are at work on a bookkeeping ontology that can be used to intermediate between XML markup and exposing Linked Open Data as RDF.[1] The information contained in the texts of accounts can then be used to explore humanities questions at levels from the granular or local to the regional or global.

---

[1] Resource Description Framework, https://www.w3.org/RDF/ (checked 2016-10-30).

This paper reflects presentations from a multi-speaker session at the 2016 Digital Humanities conference in which MEDEA participants discussed the kinds of humanities information found in accounts, the forms of electronic representation available for working with them, and an evolving bookkeeping ontology based on CIDOC-CRM.[2]

MEDEA benefited from a Bilateral Digital Humanities award from the National Endowment for the Humanities and the German Research Foundation during the 2015-2016 academic year. Our website can be found at http://medea.hypotheses.org.[3]

The Many Forms and Formats of Historical Accounts

Historical accounting documents have taken many forms over the course of human history, and their formats have changed in ways that have not always been reflected in the historical literature. While a shorthand reference to the development of so-called double entry accounts based on the practices of early modern Italian merchants might prevail among historians of the modern era, considerable evidence demonstrates a more nuanced process of change over time. In fact, historians have used the term "accounts" for a wide range of primary sources. Long before the fifteenth century, people all over the world tracked exchanges of goods, services, and rights by writing on a variety of materials that included clay and wax tablets, slate, tallies, papyrus, parchment, and paper, and they used a range of tools that were appropriate to these different materials. At various times and places, the people making these records had literacy skills ranging from basic numeracy through full literacy; their purposes in creating the records extended from simple memory aids through complex professional evaluations. The common element among these highly differentiated sources was their focus on economic activities. Thus, we can deduce that the form and inner structure of the accounts evolved in relation to available materials and tools, in association with diverse purposes, and in connection to the development of economic activities.

Historians have demonstrated that those responsible for keeping track of the funds of cities, estates, businesses, charities, and other human organizations developed their accounting practices as they encountered new situations and problems. Medieval and early modern accounts from French- and German-speaking regions reveal some of the processes of this non-linear, locally influenced development (Arlinghaus 2000, Mersiovsky 2000). And research on nineteenth-century business accounts has shown that only in the final decades of the century did accounting become a professional pursuit in which such contemporary notions as cost accounting and depreciation of business machinery were standard management principles that reputable firms and organizations were expected to follow (McGaw 1985, Yates 2000). In the following four examples, MEDEA participants illustrate areas where digital scholarly editions of accounts can add to knowledge about the evolution of accounting practices as well as about daily life and the sorts of factors that have been the subject of social and economic history for the past eighty years.

---

[2] CIDOC's Conceptual Reference Model, http://www.cidoc-crm.org (checked 2016-10-30).

[3] Any views, findings, conclusions, or recommendations expressed in this presentation do not necessarily reflect those of the NEH or the DFG.

A Hanse Merchant of the Fifteenth Century: Johan Pyre's Methods

The Hanseatic League operated in Northern Europe between the mid-fourteenth and mid-seventeenth centuries. It began as an alliance of merchants who spoke Low German, and it expanded into a confederation of market towns that specialized in the trade between Eastern and Western Europe in the region along the coasts of the Baltic Sea and the North Sea. The League was at its height in the fifteenth century. Since these merchants were active during the same period that the so-called double-entry bookkeeping methods of Venetian merchants was codified by mathematician Luca Pacioli in 1494, digital scholarly editions of accounts from Hanse merchants add regional breadth to our understanding of late medieval and early modern European accounting methods.

Using maritime transport gave the merchants in Hanse ports a cost-effective way to convey such bulk goods as grain or salt. Cooperation through the League also enabled the merchants to develop a system in which Hanse agents in the ports acted both as paid employees of their peers in other ports and as independent entrepreneurs who helped in handling the merchandise of their partners within the League. These agents offered their services free of charge, as they understood their efforts to be repaid through reciprocal assistance within the League, whose rules guaranteed the quality of these services. This system saved the merchants the costs of maintaining their own establishments and employees in each port. Notably, Hanse merchants did not use the so-called "Italian method" of double-entry bookkeeping. They developed their own distinctive bookkeeping practices.

The accounts of Hanseatic merchant Johan Pyre (in older literature also Pisz) offer examples of traditional Hanse accounting practices as well of one individual's development of an idiosyncratic system of bookkeeping as an aid to tracking his own merchandise. As Orlowksa has demonstrated, Pyre's personal system evolved through eight stages over twenty-five years (2012). He was active in Danzig from 1421 to 1455. His accounts from the beginning of this period reflected Pyre's activities as an independent businessman. He set up his book using typical methods of traditional Hanse bookkeeping, including personal accounts and a register of purchases and sales. Between 1422 and 1424, Pyre divided complex trades from simpler records of purchase and sale by putting the two kinds of information in separate parts of his account book. At the same time he developed a method of tracking the progress of his own trades through the system of Hanse merchants and ports by writing out separate portions of information related to a single transaction on two facing pages in his account book, so that he could easily check the stages as his goods moved from merchant to merchant and from port to port. After 1425, Pyre further separated purchases from sales, and he started a second accounting book. In 1426, he started a new way to note the complex transactions, sorting each around the most important partner within the process. And in 1427, he further sorted out the information about the transactions in which he invested. Tracing these complex transactions had become for him an iterative process. Between 1434 and 1441, Pyre set apart the account of his main business partner with a still more detailed system in which he noted debits on one facing page and credits on the other. Pyre's business declined between 1451 and 1455, and by the end of the final year, he quit most of his connections entirely.

Late Medieval and Early Modern Urban Administration: Basel

The annual accounts of the city of Basel are typical of accounts produced in the course of late medieval and early modern urban administration (Harms 1909-1913, Vettori 1986, Burghartz et al. 2015). They share the story of account books in the German territorial states that has been told by Mark Mersiowsky (2000). In these regions, *rotuli* or rolls were replaced by booklets and full-sized books in the period between the thirteenth century and the early sixteenth century. Initially instead of individual entries for each transaction, these books included long paragraphs under each heading. Fixed columns to distinguish descriptive text and amounts booked were introduced only in the late fifteenth century. And even as the bookkeeping practice evolved, scribes tended to hold onto traditional forms: The Basel scribes continued to use roman numbers well in the sixteenth century even though arabic numbers had already become the default system for everyday calculations. This practice is demonstrated in the account for the year 1535, where the marginal total was written with arabic cyphers while the main body of the ledger used roman numerals (http://gams.uni-graz.at/o:srbas.1535/sdef:TEI/get#bs_DiegtenEptingenAusgaben-total-1). The categories under which the city administration organized their budget were based more on their history and administrative practice than on their economic function: Categories for income were maintained even if no transaction had been recorded for a long time. The Basel accounts demonstrate other peculiarities in historical accounting methods: The books record cash flow, thus loans granted or repaid are booked as "outgoing money" ("Ußgeben"), investments and everyday costs are booked together as general expenditure, while the 144,862 pounds the city borrowed from several people in 1587/88 (http://gams.uni-graz.at/o:srbas.1587/sdef:TEI/get#bs_KrediteAufgenommen-div-0) are "income" ("Einnehmen", in total 198,493 lb). The loans others repaid to the city were already recognized as a specific kind of income in 1543/44, but they formed a separate account only from 1572/3 onwards [Sagelsdorffer 2015, Section "Einnahmen der Stadt"].

A Long-Lived and Well Documented Charitable Institution: St. Catherine's Hospital in Regensburg

The archives of St. Catherine's Hospital in Regensburg (below abbreviated as SpAR) hold a consistent series of books of account dating from 1354 to the present. These books of account and accompanying records allow economic historians to trace detailed information on wages, prices, and benefits in kind for a vast number of professions and goods over more than six hundred years, as well as to improve the understanding of how (pre-) modern persons in charge formed their economic expectations based on past experiences and economic learning. This can be done by a comparison of quantitative and qualitative data, because the Regensburg hospital archives contain additional narrative sources: Some of them in addition to or even within the ledgers as side-notes or inlay-notes, some of them as separate protocols and chronicles, which further illuminate continuities and changes in the hospital´s socio-economic situation from the mid-fourteenth century to the present.

This series of books of account comes in a variety of different sizes, formats, designs and hands, all depending on the state of the art of book-keeping at the time. The humanities information to be found in a collection covering such a long period of time offers both stability in the kinds of information recorded and diversity in the formats for recording that information. The serial nature of account books makes them a particularly stable genre of texts that are consequently

useful to linguists for tracing the evolution of terms (Ruge 2015). At the same time, books associated with grain for the year 1770/1 demonstrate divergent representations of economic thought, as is evident when the granary register (SpAR Kastenregister 1770/1) is compared to the corresponding grain selling book (SpAR Getraidkaufbuch).

Personal and Business Accounts from the Nineteenth-Century United States: Laban Morey Wheaton

By the early nineteenth century, businessmen in the United States were being trained to keep their accounts according to the so-called double entry accounting methods popularized by textbooks written by Scottish mathematician John Mair and others who promoted their manuals as translations of the "Italian system" found in the accounts of Venetian merchants and in the advice of Luca Pacioli (Mair 1765, Pacioli 1494). Massachusetts businessman Laban Morey Wheaton was typical in this regard. After graduating from Brown University in 1817, Wheaton ran a general store in the town of Norton where he held the office of postmaster. A daybook, ledger, and books of cash received and expended reflect shifts and consistencies in the commodities Wheaton stocked in his store and in the people who traded there. Rents, activities related to his dairy herd, and investments can be found in these books of account as well. After Wheaton inherited the family's land and fortune at his father's death in 1846, he extended his business activities, joining with other investors in running a cotton mill complex called the Wheaton Manufacturing Company, purchasing a cotton batting mill at another site called Center Mills, and joining with his cousin from nearby Attleboro in establishing a straw-hat manufactory in Norton. Although the surviving account books from Wheaton's enterprises are incomplete and refer to books that have not been preserved, those that do remain include not only economic information about commodities and prices but also the names of individuals and their families who worked for Wheaton, rented his farms, and traded at his store. These accounts offer granular level details about the daily lives of inhabitants of a New England town with a mixed agricultural and industrial economy during the second quarter of the nineteenth century. Like the account books of Hanse merchant Johann Pyre five hundred years earlier, Wheaton's accounts have significance both as sources of information about local interactions and as examples of larger patterns of recording economic activities.

Accounts as Social and Economic History, Cultural History, and Technologies

For decades, economic historians have been interested in reconstructing socio-economic interactions and, thus, manifestations of economic decisions by analyzing books of account, which have evolved from the Middle Ages to the present. Their structure and its development were the product of professional, educational, linguistic and administrative change over time, and these are reflected in their designs, logical structures, and formats. Certain aspects, however, have remained the same, and for the most part historians exploring social and economic aspects of the past have used these consistencies as justification for extracting data from historical accounting documents rather than producing full scholarly editions.

Much of this extracted data is stored either using spreadsheet software or slightly less flat relational databases. Since which data are extracted depends on the specific research interests of the scholars creating the datasets, these lack the scrutiny of historical source criticism and thus carry all the risks of being interpreted in misleading contexts, using undocumented data categories from alien research contexts, or including extrapolated data without documenting the processes of the extrapolation. What is more, the state of the art tends to focus too exclusively on cliometric aspects while the analysis of accounting is, in fact, inseparable from economic theories: for example if we think of theories of uncertainty from Knight to Taleb (Knight 1921, Taleb 2007), if we consider historiographical concepts concerning spaces of experiences and horizons of expectations, if we think of Systems Theory (Luhmann 1975), or even if we think of aspects of neuroeconomics in the light of how the design of information interacts biologically with the competences that decision-makers have to react to certain challenges (Glimcher/Fehr 2013).

MEDEA principal investigators agree that historical accounting records include information of historical interest in addition to data that can be extracted from them for social science research. Treating accounts as humanities sources and creating digital scholarly editions in which that information is represented in XML/TEI[4] instead of in a proprietary spreadsheet or relational database allows the researcher to enrich the information with insights gained from interpretation of the data and indicate the interpretation clearly. For example, the editor can add regional meanings of certain terms for commodities - "korn", e.g., can stand for a variety of different plants in early modern German-speaking records. Another example is repetitions of similar or identical personal names, where an additional layer of crucial information can help to distinguish the individual that is mentioned in the particular account. Thus, interpretations—or perhaps more appropriately, explanations and references—can be included in the TEI-based digital edition. The advantage of digital scholarly edition is that it allows adding interpretations to the transcriptions of the actual document as opposed to extracting bits of information out of their context. And the edition can be further enhanced by future scholars whose research interests differ from those of the initial editor.

TEI as a Step in Encoding Accounts

Thus we recommend digital scholarly edition of historical accounting documents using XML/TEI as an important step in encoding these rich humanities sources for web publication. Here, we offer a very brief overview: Since the various formats of historical accounting documents represent information sometimes as sentences, sometimes as paragraphs, and sometimes in apparently tabular formats, we recommend use of whichever elements might be most appropriate to the particular documents being edited: <p>, <ab>, or <table>. P5 of the TEI Guidelines allows embedded markup of many of the elements found in accounts. Both the <num> and <measure> elements offer ways to express information about amounts. Attributes can be added to these elements to indicate whether numbers are expressed as roman or arabic numerals as well as to identify commodities and currencies. To introduce a dedicated vocabulary for accounting concepts, standoff markup, and TEI extensions or customizations can be used to serialize information as has been demonstrated in the so-called transactionography proposed by

---

[4] Text Encoding Initiative, http://www.tei-c.org/ (checked 2016-10-30).

Tomasek and Bauman (Tomasek/Bauman 2013) or in the links to a basic ontology in digital scholarly edition projects as described by Vogeler (Vogeler 2014, Vogeler 2016). As a result of these experiences and considerations we recommend use of the @ana attribute to add references to a developing bookkeeping ontology that is described below.

XBRL: Contemporary Business Software for Historical Financial Data

An alternative XML based approach is offered by contemporary accounting formats represented in the "eXtensible Business Reporting Language" (XBRL), in particular since the TEI definition of tables as cells nested in rows runs counter to a fundamental mathematical function in accounts--the use of columns of numbers with sums at the bottoms of pages. XBRL has emerged during the past decade as the international standard for reporting financial information. The United States Securities and Exchange Commission (SEC) required publicly-traded companies to submit reports in XBRL by 2011 (Cohen 2009). Other financial markets in Asia and the European Union have either mandated or promoted the use of XBRL for financial reporting. Given that companies in these jurisdictions report their financial data according to different standards, XBRL does not impose any national reporting standard. Rather, it must remain flexible enough to accommodate the differences among standards internationally. As Vogeler has noted, XBRL should be considered for encoding historical financial data (2014, 2015: 322-324). But is it sufficiently pliable to accommodate the full range of data found in historical accounting documents?

A Brief Introduction to XBRL

As the name suggests, XBRL is an XML-based standard for reporting financial information (Engel et al. 2003). However, XBRL does not use a single schema per document as the TEI does. Rather, the XBRL standard incorporates a number of different kinds of specifications. At its core, XBRL consists of *instance* documents that record financial facts and *taxonomies* that define the concepts that describe those facts. *Linkbases* are another key component of the XBRL infrastructure, providing additional definitions and rules governing concepts. Encoding a financial report in XBRL requires connecting instance documents with (shared or sometimes mandated) taxonomies and associated linkbases.

XBRL instance documents look very flat to digital humanists who are accustomed to working with formats like the TEI. An XBRL instance document contains elements that establish the context for facts (generally, the period or instant of the reported information), the monetary units used to denominate those facts, and a series of facts. The information in these facts may also be amplified by the use of footnotes. These elements are mostly presented side-by-side at the same level of hierarchy and are interconnected through the use of attributes.

Understanding the XML Linking Language (XLink) sheds light on the nature of XBRL. The W3C XLink recommendation may be unfamiliar to many digital humanists (DeRose et al. 2010). XLink provides not only for outbound links from one document to another as in HTML and XML/TEI. It also provides for the semantics for more complex forms of links, such as inbound links from external documents. In essence, the XLink recommendation provides the semantics

for creating graph structures in XML. XBRL relies on XLink in large part because XBRL itself is fundamentally a graph model serialized in XML format.

As we shall discuss in the following section, graph theory provides a solid theoretical basis for describing data of all types, especially at a high level of abstraction. There is significant interest in developing an RDF data model for XBRL that would expose and formalize its implicit graph structure (O'Raian 2012). The fact that XBRL shares the syntax of XML and the semantics of RDF graphs suggests, at least *prima facie*, that XBRL might function as a bridge between diplomatic transcriptions of financial data in TEI and semantic representations of that data in RDF or other graph-based data models.[5]

Encoding Historical Financial Data in XBRL

The inherent flexibility of XBRL seems to suggest that it could be used to encode historical financial data. The XBRL data model already accommodates widely varying forms of financial data. In principle, XBRL should be capable of encoding ledger books and other financial sources from previous centuries. The data from sources could be transcribed and encoded as facts in instance documents, and footnotes could be used to preserve contextual information. Financial historians would also need to develop appropriate taxonomies to provide the semantic information about the facts in instance documents. While contemporary XBRL users rely on existing taxonomies, overriding them with local taxonomies to handle specific kinds of facts, XBRL does not require the use of any particular taxonomy. Certainly, financial historians could develop their own taxonomies to indicate the categories and kinds of transactions at play in their sources.

An immediate problem arises, however, since the XBRL standards require monetary units of measure to be denoted with the ISO 4217 codes for international currencies (ISO 2015).[6] Obviously, this makes working with historical currencies problematic. This requirement reflects the implicit presupposition that XBRL will be used to encode contemporary financial reports. There is no reason in principle why this requirement could not be loosened or overridden for historical financial information, though it would be much more difficult to develop an analogue to ISO 4217 codes for historical currencies.

Solving this problem still leaves open the question of what gain would be achieved by recording historical information in XBRL. If every project were to develop its own taxonomy, it is doubtful that those independent taxonomies would be adaptable beyond the projects that gave rise to them. Any gains from using XBRL for encoding historical financial information would be severely diminished or obviated by incompatibilities between taxonomies. If every project were to develop highly-localized taxonomies, there would be no way to share or aggregate data.

A possible solution would be to make modular taxonomies that could be linked together to handle shared concepts from certain time periods or geographic localities, while allowing historians the freedom to create taxonomies whenever no analogues exist. While data sharing would be limited by incompatibilities between taxonomies, perhaps there would be sufficient commonalities to allow for partial aggregation of data. Exposing the network of these

---

[5] On the relationship between TEI and CIDOC-CRM see Ore and Eide (2009) and Ciula and Eide (2014).

[6] Syd Bauman of Northeastern University originally drew this problem to our attention.

taxonomies would constitute an interesting intellectual exercise in its own right as it would indicate vectors of dependency between different traditions of accounting and bookkeeping.

Identifying areas of overlap in historical accounting practices would require a great deal of communication among financial historians. The two international meetings of the MEDEA project indicated the promise of gathering together scholars working across widely varying time periods and geographical locations, but also underscored the difficulties.

There is also another option worth exploring. In recent years, the maintainers of the XBRL standards have created a framework called XBRL Global Ledger to encode the transactional information that becomes summarized in financial reports (XBRL 2015). Given that most historical financial sources are transactional rather than reports, the XBRL Global Ledger may prove a better path forward than the XBRL specification itself. However, validating that assumption would require extensive testing against heterogenous historical sources.

A CIDOC-CRM Based Ontology of Economic Activities

All these XML formats represent parts of what humans did when they exchanged goods and services and recorded these economic activities in written form. Each method has its advantages. But still it seems that behind all these forms is a common model which could be represented conceptually in the following way: There are *agents* agreeing on a *transfer* of something *measurable* and executing this transfer. Usually the transfer is mutual, i.e. at least two transfers aggregate to a *transaction*, in which usually *money* is exchanged for *commodities*, *services* or *rights*. Additionally one or the other party creates a *documentation* of this transfer, which can be an artifact like a tally, but usually is a written accounting document. It is important to notice that the financial document is not necessarily representing the transfer or the transaction itself; a record can even be fictitious. It also adds a level of interpretation: The transactions are organized in the document following a specific *scheme* meaningful to the agent commissioning the record, the "categories", "rubrics", or "accounts", which can be abstracted to *dimensions*; often the accountant aggregates the individual transactions into *totals*, *balancing* the financial state or demonstrating profits and losses; and some transactions are not recorded at all.

While the historical process of economic transactions can be modelled as seen above, the source material thus has a complex relationship to these historical events we assume to have happened. As in all manuscript analysis, a page is written by one or more hands (writers) in one or more sessions. One specific record can be found on (parts of) one or more pages. So we have an activity performed by one or more actors at a period of time, which may have been discontinuous, which resulted in a record. The traces of this activity can be found in the document and can be established as such by competent readers, as we have seen in the examples above.

Each record represents a claim about something that happened in the real world, an economic transaction of some sort. This is another activity, connected to but distinct from the record creation event. In an event-based modelling of historical accounts these two activities will form the core of a network of information, and information—about actors, time, location, the

historical account document, and other events connected to the transaction such as goods, money, currency information—become nodes attached to this core.

Formalizing the conceptual model resulting from these considerations would allow us to map the formalizations described above for their procession as complete set. This goal seems reasonable as of course the records of an economic activity in the ledger of one company should find their counterpart in the ledger of its exchange partner. A meta search on digital representations of various Hanse records could therefore allow extracting networks and flows of trade activities. The research on long term price series or their regional diversion would profit from a common access to the records. The recent solutions historians at the University of Tokyo (Kokaze et al, 2016) have found to represent chronology in historical accounts confirm this: They study the relationship between agents mentioned in the accounts of the British Lay-Osborn Flotilla and Chinese administration from the 1860s in time.

The conceptual model covers completely, what a TEI based transactionography (hsfr) and XBRL encoding describe: The hfrs:transaction/*medea:transaction* consists of hfrs:transfers, each of them involving *medea:agents*/@hfrs:fra and @hfrs:til. The hfrs:transfer is defined as 'A transfer of something of value', which is a tei:measure/*medea:measurable*. The *medea:documentation* of the transfer is encoded with hfrs:attested as a reference to a TEI text structure. The hfrs:listAccount and hfrs:account express the interpretative organization of transactions in *medea:scheme*s and *medea:dimension*s. hfrs:listTransaction and hfrs:transferGrp are generic grouping functionalities which are reflected in the *medea:dimension*s of documentation and the *medea:total* and *medea:balance* concepts. The relationship between the suggested conceptual model and the abstract level in XBRL is even more obvious: The XBRL 'financial facts' are a superclass to *medea:transfer*, *medea:transaction*, or *medea:total*. They are recorded in 'instances' (*medea:document*) and the 'taxonomies' describe their interpretative dimensions (*medea:dimension*, *medea:schema*).

The currently most adaptable formal representation for this kind of data is the Resource Description Framework (RDF) of the W3C. Its graph-based model is flexible but structured enough to express complex data structures. The basic conceptualization similar to simple natural language sentences as subject-predicate-object triples makes it fairly easy to understand. RDFs and OWL[7] offer rich possibilities for formal modelling. A preliminary RDFs schema for the conceptual model has therefore been created and published on github (https://github.com/GVogeler/bookkeeping).

The CIDOC-CRM, as an event-based ontology, is a good structure for encoding such information, as shown through the use of the standard for several types of cultural-historical evidence. As a matter of fact, it serves well to cover the main parts of the conceptual model described above, with its two event types:

1. The creation of the accounting document
2. The transfer as the recorded economic activity

While the first is clearly a subclass of the E65_Creation class, the latter can be considered as a subtype of E7_Activity and of E28_Conceptual_Object. Both are connected by the document

---

[7] Web Ontology Language, https://www.w3.org/OWL/ (checked 2016-10-30).

written by the accountant and attesting a transfer. The example from Kokaze et al. (2016) demonstrates that the general properties of CIDOC-CRM events have practical research relevance.

Given the specific need in modelling these documents and the activities documented in them, some extensions to CIDOC-CRM should be made. This is also in line with what we have seen in other areas including archaeological excavations[8] and buildings,[9] and scientific observations[10] establishing a CIDOC-CRM based family of standards. In the development of this CIDOC-CRM based extended model it will be an important question what the target for the modelling is, that is, what time periods, geographical areas, and specific document types will function as evidence for the modelling. This will determine how specific the model can be: the larger the scope, the looser and more flexible the model must be.

The two event types described above have different relationships to truth. As for the records in the historical account we know they exist and were written, at least as long as the original document exists. What is questionable is who wrote it, when, and also the level of unity, that is, what makes one record separate from other records. For the economic transaction documented in the record the question of existence is somewhat different than from the record. We only know there is a claim for an economic transaction, not that any transaction actually took place, for which one often needs additional evidence, knowledge, and analysis to determine.

This creates an interesting question of modelling principles: should one model the worldview expressed by the documents or should one express the worldview seen as most likely to be correct by the historians (Ciula, Spence, Vieira 2008)? The choice of modelling standards should not determine that choice. However, as CIDOC-CRM is a tool for information integration it is more common to use it for models aiming towards historical truthfulness. That does not mean that the worldview expressed in the documents should be hidden, just that the activities expressed in the model will be the ones the historians believe to have taken place. Modelling and encoding the transactions recorded in historical accounting documents as stand-off reference between textual evidence and description of the economic and social 'facts' even makes possible multiple interpretations of the textual evidence, which would lead to considering the addition of a factoid-based approach (Bradley and Pasin 2015) to the model.

Conclusions

The current results of the work done by the MEDEA collaborative are therefore fivefold:
1. The textual genre of accounts shows a great variety in form in its history with a substantial common ground in concepts and human activities represented in these texts.
2. These concepts and human activities are of high interest in the humanities and historical social science research.
3. Based on the work of the DH community in the TEI and of contemporary business reporting standardization activities in XBRL, it is possible to develop a conceptual model for the documentary representation of human activities in the field of economics.

---

[8] CRMdig, http://new.cidoc-crm.org/crmarchaeo/home-3 (checked 2016-10-30).

[9] CRMba, http://new.cidoc-crm.org/crmba/home-7 (checked 2016-10-30).

[10] CRMsci, http://new.cidoc-crm.org/crmsci/home-1 (checked 2016-10-30).

4. This model can be expressed as a graph structure, and it can be integrated into the larger conceptual framework of the CIDOC-CRM family of standards.
5. It is worthwhile to continue the development of the formal model and make suggestions about how to integrate the formal model into encoding practice and data publication of historical accounting records.

Certainly, the solutions discussed here are only a few of all the possible ways that web technologies can enhance research with historical accounting records. The MEDEA scholars look forward to suggestions for modifications of the ontology, additions of other possible methods for encoding economic facts represented in accounting documents, more efficient setups for execution of projects, and a thorough scrutiny of the conceptual consistency of our proposal. The MEDEA scholars will continue their work on the ontology and will try to set up a collection of examples showing how we think the integration of the representation of content expressed in RDF and scholarly edition of accounting documents can be realized.

Funding

References

Arlinghaus, Franz-Josef (2000). *Zwischen Notiz und Bilanz: Zur Eigendynamik des Schriftgebrauchs in der kaufmännischen Buchführung am Beispiel der Datini/di Berto-Handelsgesellschaft in Avignon (1367-1373).* Gesellschaft, Kultur und Schrift mediävistische Beiträge, ed. by Hagen Keller, Band 8. Frankfurt: Peter Lang.

Bradley, John; Michele Pasin (2015). Factoid-based Prosopography and Computer Ontologies. Towards an integrated approach. *DSH* 30,1: 86-97.

Ciula, A. and Eide, Ø. (2014). Reflections on Cultural Heritage and Digital Humanities: Modelling in Practice and Theory. In *First International Conference on Digital Access to Textual Cultural Heritage (DATeCH),* Madrid, Spain, pp. 35–41. http://dl.acm.org/citation.cfm?id=2595207 (checked 2016-10-30).

Ciula, Arianna; Paul Spence; José Miguel Vieira (2008). Expressing Complex Associations in Medieval Historical Documents: The Henry III Fine Rolls Project. *Literary and Linguistic Computing* 23, 3: 311-25.

Cohen, Eric E. (2009). "XBRL's Global Ledger Framework: Exploring the Standardised Missing Link to ERP Integration," *International Journal of Disclosure and Governance* 6, no. 3: 129, doi:http://dx.doi.org/10.1057/jdg.2009.5.

Crofts, N.; Doerr, M.; Gill, T.; Stead, S. and Stiff, M. (eds.). (2011). *Definition of the CIDOC Conceptual Reference Model*. ISO 21127:2014.

DeRose, Steve et al., eds. (2010). "XML Linking Language (XLink) Version 1.1" , https://www.w3.org/TR/xlink11/.

Engel, Phillip, et al., eds. (2003). "Extensible Business Reporting Language (XBRL) 2.1" http://www.xbrl.org/Specification/XBRL-2.1/REC-2003-12-31/XBRL-2.1-REC-2003-12-31+corrected-errata-2013-02-20.html.

Glimcher, Paul W., and Ernst Fehr, eds (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Harms, Bernhard (1909-1913). *Der Stadthaushalt Basels im ausgehenden Mittelalter. Quellen und Studien zur Basler Finanzgeschichte*. 3 vol. Tübingen: Laupp in Komm.

International Organization for Standardization, (2015). "ISO 4217:2015(en), Codes for the Representation of Currencies," https://www.iso.org/obp/ui/#iso:std:iso:4217:ed-8:v1:en.

Knight, F. H. (1921): *Risk, uncertainty and profit.* Boston: Hart, Schaffner, and Marx.

Kokaze, N.; Nagasaki, K.; Shimoda, M.; Muller, A.C. (2016). Modeling New TEI/XML Attributes for the Semantic Markup of Historical Transactions, based on 'Transactionography.' *Proceedings of the 6th Conference of Japanese Association for Digital Humanities: "Digital Scholarship in History and the Humanities.* Tokyo: University of Tokyo, 2016. http://conf2016.jadh.org. http://conf2016.jadh.org/abstracts/s5-2/

Luhman, N. (1975), "Systemtheorie, Evolutionstheorie und Kommunikationstheorie". *Soziologische Gids* 22, 3: 154–168.

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).

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.

Mersiowsky, Mark (2000). *Die Anfänge territorialer Rechnungslegung im deutschen Nordwesten. Spätmittelalterliche Rechnungen, Verwaltungspraxis, Hof und Territorium* (zugl. Diss. phil. Münster 1992) , Sigmaringen: Thorbecke(Residenzenforschung 9).

O'Riain, Seán; Edward Curry; and Andreas Harth. (2012) "XBRL and Open Data for Global Financial Ecosystems: A Linked Data Approach." *International Journal of Accounting Information Systems, XBRL: Research Implications and Future Directions*, 13, no. 2: 141–62.

Ore, Christian-Emil Smith; Eide, Øyvind (2009). TEI and Cultural Heritage Ontologies: Exchange of Information? *Literary & Linguistic Computing* 24, no. 2: 161--72.

Orlowska, Anna Paulina (2012). "Handel in einem Kaufmannsnetz. Der Danziger Johan Pyre". *Vertraute Ferne. Kommunikation und Mobilität im Hanseraum* / ed. By Mähnert, Joachim : Husum. 32-39.

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).

RDF Working Group. *Resource Description Framework* (RDF). http://www.w3.org/RDF/.

Ruge, N. (2015). The *comptes de la baumaîtrie* of the City of Luxembourg (1388-1500). Abstract of presentation at MEDEA conference in Regensburg. http://medea.hypotheses.org/304.

Taleb, N.N./**Pilpel**, A. (2007): Epistemology and Risk Management, *Risk and Regulation* 13:6—7.

TEI Consortium, eds. *Guidelines for Electronic Text Encoding and Interchange* . [Last updated on 26th March 2016]. http://www.teic.org/P5/.

Tomasek, K. and Bauman, S. (2013). Encoding Financial Records for Historical Research, *Journal of the Text Encoding Initiativ*e, 6. 10.4000/jtei.895.

Vettori, Arthur (1984). *Finanzhaushalt und Wirtschaftsverwaltung Basels (1689-1798). Wirtschafts- und Lebensverhältnisse einer Gesellschaft zwischen Tradition und Umbruch*. Basel, Frankfurt a. M.: Helbing & Lichtenhahn.

Vogeler, G. (2014). Modelling digital edition of medieval and early modern accounting documents. *Digital Humanities 2014*: Conference Abstracts. Lausanne: EPFL and UNIL, July 2012, pp. 398-400.

Vogeler, Georg (2015). Digitale Edition von Wirtschafts- und Rechnungsbüchern. Gudrun Gleba / Niels Petersen (eds.): *Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit. Formen und Methoden der Rechnungslegung: Städte, Klöster, Kaufleute.* Göttingen. Universitätsverlag Göttingen: 307-328.

Vogeler, Georg (2016). "The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies" In: Sarnowsky, Jürgen (Hg.): Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters. Göttingen. V&R unipress. 2016. 13-41.

W3C. OWL Web Ontology Language. http://www.w3.org/standards/techs/owl#w3c_all.

XBRL Global Ledger. https://www.xbrl.org/the-standard/.

XBRL (2015). "XBRL Global Ledger Taxonomy Framework 2015." http://www.xbrl.org/int/gl/2015-03-25/gl-framework-REC-2015-03-25.html.

Yates, JoAnne (2000). Business Use of Information and Technology during the Industrial Age. *A Nation Transformed by Information: How Information Has Shaped the United States from Colonial Times to the Present*, ed. Alfred W. Chandler, Jr. and James W. Cortada. New York: Oxford University Press: 107-135.

Authors

Kathryn Tomasek is Professor of History at Wheaton College in Norton, Massachusetts.

Anna P. Orlowska works in the Department of Historical Atlas at the Tadeusz Manteuffla Institute of History, Polish Academy of Sciences in Warsaw.

Kathrin Pindl is Scientific Coordinator for Research Training Group 2337/1 "Metropolitics in the Premodern" at the University of Regensburg.

Clifford Anderson is Associate University Librarian for Research and Learning at Vanderbilt University Libraries in Nashville, Tennessee.

Øyvind Eide is a deputy professor in Digital Humanities (Historisch-Kulturwissenschaftliche Informationsverarbeitung) at the University of Cologne.

Georg Vogeler is University Professor for Digital Humanities at the Zentrum für Informationsmodellierung (ZIM), Austrian Center for Digital Humanities, at the University of Graz.