

Digital Edition Publishing Cooperative for Historical Accounts

Abstract and Overview

The Digital Edition Publishing Cooperative for Historical Accounts will offer publication and access services to a wide range of editors and users interested in the information contained in historical accounting records. This hub will allow editors to upload their transcriptions from multiple formats—including Excel, Drupal, XML/TEI—without the need for encoding expertise. As outputs, the hub will offer visualizations of current interest to editors—price timelines, commodity pie charts and network diagrams that can be exported to editors’ own websites. In addition, the hosted data will be converted to Resource Definition Framework (RDF), which will make it discoverable for data mining for historians who seek to take advantage of the semantic web and for researchers in fields other than history.

This publishing cooperative will leverage relationships developed under Modeling semantically Enhanced Digital Edition of Accounts (MEDEA), a primarily European-based project funded through a 2015 Bilateral Digital Humanities award from the National Endowment for the Humanities and the German Research Foundation. The cooperative will share technical expertise and establish a platform for publication of digital editions that include textual and numerical representations; visualizations of accounting information; and data representation referencing a shared MEDEA bookkeeping ontology. The proposed platform will replicate a system developed at the Centre for Information Modeling (Zentrum für Informationsmodellierung—ZIM), Austrian Centre for Digital Humanities, at the University of Graz: the Humanities Asset Management System (Geisteswissenschaftliches Asset Management System—GAMS), a FEDORA Commons-based infrastructure for data enrichment, publication, and long-term preservation of digital humanities data.

Our team of experts shares a scholarly affinity for creating digital editions of accounts, with historian **Kathryn Tomasek** of Wheaton College serving as Principal Investigator. **Georg Vogeler** and **Christopher Pollin** join the team from Graz. MEDEA participants lead several U.S. projects focused on creating digital editions of accounts: **Jennifer Stertz** and **Worthy Martin** at the University of Virginia, **Anna Agbe-Davies** at the University of North Carolina at Chapel Hill, and **Jodi Eastberg** at Alverno College. Additional participants work in libraries that hold a substantial number of account books in their collections: **Molly Hardy** of the American Antiquarian Society and **Gregory Colati** of the University of Connecticut. **Ben Brumfield** and **Sara Brumfield**, of Brumfield Labs, LLC, will serve as technical consultants in the United States. **Kate Boylan**, Director of Archives and Digital Initiatives, contributes a leadership perspective from Library Services at Wheaton College, the proposed host institution. The project team will hold two in-person meetings and will meet via video conference 10 times, producing a plan for a digital edition publishing cooperative built on the GAMS technical model and the associated workflows. In-person meetings will coincide with the annual meeting of the American Historical Association (AHA) since most participants already attend the conference and can cover their own costs for travel and lodging. In two parallel meetings, groups of team members will focus on (1) technical concerns—evaluating needs of user communities and establishing programs and workflows for conversion of data and creating outputs, and (2) administrative matters—collecting and evaluating cost models and assessing hosting requirements, including feasibility of sharing technical data across editions and institutional agreements necessary for federation of editions. Together, the editors and technical experts will produce test files for demonstration purposes. The greatest portion of the budget will pay for technical experts, with a smaller portion budgeted for travel. The broad perspectives of the project team will ensure

outcomes that benefit the existing documentary editing community (Stertzner), the edition cluster (Stertzner, Agbe-Davies, and Tomasek), and future editions (Martin, Eastberg, Hardy, and Colati). The team represents stakeholders interested in digital edition of accounts, including members of the highly esteemed documentary edition of the Papers of George Washington, staff from libraries and archives with large holdings of account books (Hardy and Colati), and scholars from outside the well-established documentary community whose scholarly research focuses on the information contained in historical account books and who seek to produce reusable data from them (Tomasek, Agbe-Davies, Martin, Eastberg).

The Digital Edition Publishing Cooperative for Historical Accounts also will advance the digital editions initiative by employing MEDEA's connection to the international community of digital humanities scholars focused on Linked Open Data and on creating ontologies for sharing humanities data on the semantic web. The cooperative will further the transfer of knowledge piloted under MEDEA, sharing technical expertise developed at ZIM with editors in the United States, and thereby incorporating American scholars more fully into international communities of practices that involve sharing and harmonizing information resources that accommodate local variations and the distinctive features of particular projects. As a hub for publishing digital editions of accounts, the cooperative will offer an opportunity to bridge existing gaps between the world of traditional documentary editing and that of digital history/digital humanities. Through continuing exchanges among ZIM in Austria, the cooperative in the United States, and new editions, the cooperative will advance the state of the art in digital scholarship grounded in emerging technologies of the semantic web. Our cooperative will contribute one avenue of transnational transfer of knowledge and foster global collaborations for other cooperatives supported by this initiative.

Project Team

Principal Investigator **Kathryn Tomasek**, Professor of History at Wheaton College, has been exploring models for transcription and markup of historical accounting records since 2009. Two previous awards have supported her methodological investigations: a Level 1 Start-Up Grant from the Office of Digital Humanities at the National Endowment for the Humanities for Encoding Historical Financial Records (2011) and a Bilateral Digital Humanities award from the National Endowment for the Humanities (NEH) and the German Research Foundation for MEDEA with the University of Regensburg (2015). She is a member of the Board of Directors of the Text Encoding Initiative (TEI) and has participated in numerous activities related to the Andrew W. Mellon Foundation's longstanding promotion of digital technologies in liberal education and scholarly communication. She has served on several National Institute for Technology and Liberal Education (NITLE) advisory boards and has participated in a Scholarly Communication Institute focused on history in 2005. She also has served on program committees for the annual meetings and conferences of the TEI, the Japanese Association for Digital Humanities, and the American Historical Association. Her work has been published in the *Journal of the TEI* and in the *Journal of Digital Humanities* as well in the proceedings of annual meetings of the TEI and DH2010, DH2011, DH2013, DH2014, and DH2016—annual conferences sponsored by the international Association of Digital Humanities Organizations. Tomasek's ontology based on account book transaction records developed in collaboration with co-author Syd Bauman has been translated into Japanese.

Georg Vogeler, University Professor and Chair for Digital Humanities at the Centre for Information Modelling, Austrian Centre for Digital Humanities at Graz University, has received several grants from the German Research Foundation (Deutsche Forschungsgemeinschaft—

DFG) and the Austrian Science Fund (Wissenschaftsfonds—FWF). He has been a member of Digital Scholarly Editions Initial Training Network (DiXiT) <<http://dixit.uni-koeln.de/>> and CO-OP, a project that fostered cross-border cooperations between archives and the general public <<http://coop-project.eu/>>. Both of these international projects were funded by the European Union. Vogeler has published numerous articles about digital scholarly edition of accounts optimized for the semantic web. His bookkeeping ontology is openly available on GitHub <<https://github.com/GVogeler/bookkeeping>>.

Jennifer Stertzer, Senior Editor of the Washington Papers and Director of the Center for Digital Editing (CDE) at the University of Virginia, has participated in MEDEA and kindly has supported Tomasek's work for many years. She will enhance the digital edition publishing cooperative's discussions of features that will appeal to documentary editors in the United States. CDE staff will produce XML/TEI files for use in exploring the possibility of automating addition of bookkeeping references to existing transcriptions. The proposed digital edition publishing cooperative will offer an opportunity for publishing a version of Washington accounts for discoverability and reuse through the semantic web.

Worthy N. Martin, Associate Professor of Computer Science and Acting Director of the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia, also is a MEDEA participant. At IATH, he has collaborated on numerous digital humanities projects, with many involving the transcription and analysis of manuscript materials. For example, the *Jefferson's University...the early life* (JUEL, see: <<http://juel.iath.virginia.edu/>>) Project has been making accessible documents from Special Collections at the University of Virginia, specifically documents from the first fifty years of the University. To date JUEL has not undertaken the full transcription of the accounts documents. Martin brings significant technical

knowledge on the design and development of digital humanities resources to the proposed planning process.

Anna Agbe-Davies is Associate Professor and Director of Undergraduate Studies in Archaeology at the University of North Carolina at Chapel Hill. She is an historical archaeologist whose analyses of account books from Stagville Plantation complement archaeological investigation and laboratory analysis of existing artifact collections. A MEDEA participant, she has been engaged in this work since 2014, supported by a fellowship from the Institute for Arts and Humanities, a Carolina Digital Humanities Initiative course redesign grant (both, UNC-Chapel Hill), and a grant from the Digital Archaeological Archive for Comparative Slavery Research Consortium (The Mellon Foundation and Monticello).

Jodi Eastberg is Professor of History and Director of the Center for Academic Excellence at Alverno College. A MEDEA participant, she serves as the family archivist for the Uihlein family, founders of the Schlitz brewing company. Her work with the Coutts & Co. historic financial accounts of Sir George Thomas Staunton was supported in part by a Franklin Research Grant from the American Philosophical Society. She currently serves on the advisory council for the Society for History Education and has published on Alverno's unique abilities-based education and her project-based approach to teaching history. Eastberg's edition of the Schlitz Papers will begin under a separate award.

Molly O'Hagan Hardy is the Director for Digital and Book History Initiatives at the American Antiquarian Society (AAS), where she held a fellowship from the American Council of Learned Societies between 2013 and 2017. Hardy has also received numerous awards from organizations that include the Andrew W. Mellon Foundation, the Society for the History of Authorship, and 18thConnect. She oversees a number of digital projects, including *The Printers'*

File, a prosopography in linked open data; *Isaiah Thomas Broadside Ballads Project: Verses in Vogue with the Vulgar*, a digital collection of 300 early-nineteenth century broadside ballads; and the *Just Teach One* and *Just Teach One: Early African American Print Initiatives*, an effort to make available lesser-known early American texts for pedagogical use. Her work has appeared in *The New Centennial Review*, *Book History*, *Debates in Digital Humanities*, and *American Literary History*. Her current project examines the history of rare book bibliographic records and their use in digital humanities projects. Hardy's inclusion as a member of the project team reflects the ongoing interest of the American Antiquarian Society in finding appropriate methods to represent the Mathew Carey account books online.

Gregory Colati is Assistant University Librarian for University Archives, Special Collections, and Digital Curation at the University of Connecticut. He brings to the team essential expertise in Fedora repositories as well as in OAIS and sustainability. Colati works with the Connecticut Digital Archive (CTDA), which includes images of numerous account books from the seventeenth, eighteenth, and nineteenth centuries. Originals of these sources are held by the Connecticut Historical Society.

Kate Boylan is Director of Archives and Digital Initiatives in Library Services at Wheaton College. She brings to the team an enthusiastic approach to the potential of digital initiatives in the context of a residential liberal arts college as well as an interest in shifting from Wheaton's current use of DSpace to a more flexible and robust asset management system for storage and publication of digital materials produced by students and faculty members at the college. Her experience involving design of an asset management system for her previous employer, Facing History and Ourselves, will enhance the team's discussion of hosting for the proposed publishing cooperative.

Christopher Pollin is a graduate student and developer at the Centre for Information Modelling, Austrian Centre for Digital Humanities at the University of Graz. He recently completed a joint master's degree in European Digital Media, Arts, and Cultural Heritage Studies (EuroMACHS) and is involved in the prototype of the MEDEA digital edition platform (alpha version at <http://glossa.uni-graz.at/context:medea>).

Ben Brumfield and **Sara Carlstead Brumfield**, of Brumfield Labs, LLC., have four decades of combined professional experience in software engineering with degrees in Computer Science from Rice University. Their extensive work on digital edition projects, including the Digital Austin Papers, the NHPRC-funded Civil War Governors of Kentucky Digital Documentary Edition, and the NEH-funded John Torrey Papers, will support the cooperative. Ben Brumfield is a leading expert on crowdsourced manuscript transcription and has presented on the subject at Digital Humanities, the American Historical Association, the American Library Association, and the Text Encoding Initiative as well as at international workshops including MEDEA, DiXiT, and Social Digital Scholarly Editing. Brumfield Labs's open-source FromThePage tool has been used for collaborative editing of scientific field notes, military diaries, literary drafts, punk rock fanzines and philosophers' notebooks since 2010.

Possible additional advisors include: Rhonda Barlow, Assistant Editor of the Adams Papers at the Massachusetts Historical Society; Øyvind Eide, Professor at the Institute for Historical-Cultural Information Processing at the University of Koeln, an expert in CIDOC-CRM; and Nancy Heywood, Digital Projects Coordinator at the Massachusetts Historical Society. Editors from European projects—Vera Schwarz-Ricci (Naples), Christelle Loubet (Nancy), Susanna Burghartz (Basel)—that have committed to using Vogeler's bookkeeping ontology and storing copies of their editions in GAMS may also be included in some of the

planning activities, as may Japanese digital humanities researchers working with the Tomasek/Bauman ontology. These ontologies were foundational to the origins of MEDEA.

The Edition Cluster and Its Target User Community

Leveraging International Ties to Advance Digital Edition of Accounts in the United States

A primary advantage this edition cluster brings to the initiative lies in its origins in a transnational cooperation that is grounded in shared interest in publishing digital scholarly editions of historical account books. Through MEDEA and extended outreach efforts, Vogeler has developed a growing European community of practice focused on creating editions of accounts referencing his bookkeeping ontology. MEDEA facilitated similar outreach in the United States. The data from digital editions of accounts can be made discoverable for data mining, which is a focus of research interest in digital history and in many other fields (Graham, Milligan & Weingart 2016). Thus, while the Digital Edition Publishing Cooperative for Historical Accounts is committed to offering the sorts of outputs desired by current editorial projects in the United States, it also will create a bridge between those projects and others around the world in which editors seek to combine the venerable practices of scholarly editing with the potentials of data aggregation through the web.

Data aggregation and harmonization has long been a feature of social science history, a field in which historians have been producing data sets, contributing them to repositories, and seeking ways to make them interoperable for at least the past half-century. Social science data sets are stored in such repositories as the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan in Ann Arbor <<https://www.icpsr.umich.edu>>. Efforts at open access and conversion for interoperability are underway through the DDI Alliance <<http://www.ddialliance.org/>> and OpenICPSR <<https://www.openicpsr.org/>>. And

American Historical Association Past President Patrick Manning is the lead researcher on the Collaborative for Historical Information and Analysis (CHIA) <<http://www.chia.pitt.edu/>>, a project that aims to compile historical data into a Human System Data Resource (Manning 2017). Thus, social science researchers have demonstrated significant interest in bringing together the considerable databases created for statistical analysis with SASS/SPSS in the second half of the twentieth century.

Social science historians also have raised questions about the reliability and reusability of such data. Our MEDEA colleagues from Regensburg, for example, are among a significant cohort of social and economic historians whose work was revitalized in the wake of an influential monograph published by former AHA president Kenneth Pomerantz, *The Great Divergence* (2000). Noting the similarity of European economic centers to comparable regions in Asia during the early modern period, Pomerantz explored the effects of demographic, technological, and ecological changes in such regions as the Yangzi Delta in China and Gujarat in India as well as in England and Western Europe. He sought to identify specific elements of similarity and difference that might contribute to a more empirically robust analysis of the development of modern industrial economies in the nineteenth century. While Pomerantz attempted to reframe our understanding of the early modern world as more polycentric than previous economic histories had recognized, Robert Allen urged assessment of the data sets available for purposes of comparison in economic history (2001). Following Allen, our Regensburg colleagues point to a local archive that holds account books covering a period of six centuries as an ideal source of social science data that can be used to test claims like those made by Pomerantz. To date, they are employing traditional sampling practices of economic history, while other MEDEA participants have adopted an alternate method focused on digital scholarly

edition of accounts. The notable example here is the edition of the accounts of the city of Basel, a project that was led by Susanna Burghartz and is published in GAMS <<http://gams.uni-graz.at/srbas/>>.

Like the Basel project, MEDEA explores opportunities for the production of richer and potentially more accurate data with recommendations that emerge from the practices of scholarly editing within the humanities. In this model, we begin from the scholarly edition, transcribing and marking up text and numbers using XML and the Guidelines of the Text Encoding Initiative because they offer a well-established international standard for production of archival-quality digital editions of historical sources. Not coincidentally, they are also original tools of the semantic web, and they can be converted to the component parts of Linked Open Data: Resource Definition Framework (RDF)/Simple Knowledge Organization System (SKOS)/Web Ontology Language (OWL)/SPARQL Protocol and RDF Query Language (SPARQL). Some of the advantages of these tools are described below.

MEDEA participant Andrew Wareham has noted that digital scholarly edition of historical tax records has the potential to improve the quality of data for social science history by diminishing some of the pitfalls of collecting data using social science tools. The historiography of late-seventeenth-century Hearth Tax records from England, Scotland, and Wales points to challenges that result from the complexity of the records and the interrelated documentation contained within them. These records included the returns themselves, which listed both numbers of hearths and those exempt from the tax, as well as exemption certificates, which described the forms of hardship that justified exemptions among the very poor, the poor, working families, and middling households. When historians used the returns to create pure structured data for statistical analysis, they omitted the richer information found in the exemption certificates.

Resulting historical interpretations incorrectly linked numbers of hearths or exemptions to levels of poverty or produced false geographical patterns of poverty (Arkel, 2003). In more than one case, historians presented and analyzed the exemption certificates as entirely separate sets of sources better suited to local than national analysis (Seaman et al. 2001, Neave et al. 2015). The inflexibility of structured database tools like SASS/SPSS leaves no room for inclusion of the sort of “non-standard” information found in more narrative records like the exemption certificates, information that does not lend itself to incorporation into rigid database fields. Thus for Wareham, digital scholarly edition of the kinds of records that once seemed best suited to social science data collection “has the potential to transform the study of the hearth tax” (Wareham 2015).

Similarly, treating account books as humanities data and trying to capture both the standard and non-standard information found within them can open these rich sources to analysis suitable to social and economic history as well as to less obvious uses. As Colati notes in his letter of commitment for this publishing cooperative, the kind of information contained in account books is of interest not only to historians but also to biologists and climate scientists through information about crop sales, for example. As MEDEA participant Nikolaus Ruge pointed out, account books can also be of particular use to linguists who appreciate the reliability of dating and location these sources offer. In addition, the serial characteristics of account books can “facilitate a step-by-step reconstruction of language change phenomena,” according to Ruge (2015). And, as Hardy notes in her letter of support, accounts of publishers and booksellers contain rich data about print culture and book history that libraries struggle to present online. Book historians have long relied on this rich data to understand, for example, how the industrialization of printing, a development that coincided with the development of railroads and

other forms of mass transportation of people and goods in the 1830s, led to massive upswings in the literary marketplace (Tryon and Charvat 1949). Applying semantic web technologies to the name indexes found in publishers' accounts can lead to new discoveries about the role of the book trade in the development of American culture (Zboray 1993).

Rationale for Choosing Editions

For the planning phase of the initiative, we have intentionally chosen three MEDEA editions that use a variety of transcription models so that we can explore the feasibility of creating a publishing hub based on transatlantic sharing of technical expertise. One of these editions lies fully within the realm of traditional documentary editions that have adapted to new digital contexts. Our colleagues from the George Washington Financial Papers Project bring to the cooperative the benefits of their position within the current documentary editing community. Another edition has more in common with current digital humanities projects that focus on African-American Studies than it does with more traditional documentary editing projects. Through her edition of Stagville store accounts, historical archaeologist Anna Agbe-Davies demonstrates the significance of the information contained in historical account books to researchers who study African-American experiences from outside the discipline of history. And Tomasek's edition of Wheaton Family account books is fully immersed in the model of scholarly digital editions represented in GAMS to date.

Current Editions

Washington Financial Papers: The George Washington Financial Papers Project (GWFP; published 2017) makes accessible Washington's three ledger books of accounts, which contain the basic business accounts of Washington's estates for forty-nine years (1750-1799), documenting transactions with individuals and commercial entities. Mostly in Washington's own

hand, these records show the acquisition of land, the sales of farm products, the work of servants, the operation of mills, and the purchases and sales of slaves. The GWFP allows users to read transcriptions of GW's three ledger books of accounts; perform simple and advanced searches on the documents and data; explore documents by people, places, ships, occupations and titles, services, food and beverages, agriculture, and place types; download search results, transcriptions, and data; and follow links to related correspondence in The Papers of George Washington Digital Editions (Rotunda and Founders Online).

Stagville Accounts: Historical archaeologist Anna Agbe-Davies focuses her edition on selected account books from the extensive Cameron Family Papers (Southern Historical Collection, UNC-Chapel Hill). Members of the extended Cameron family owned Stagville ca 1785-1950. Accounts pertaining to the plantation's store commence in 1785 and extend to 1886. This edition is currently in progress, and the text currently undergoing transcription is Folder 3617 (Slave Ledger 1792-1812). Within the next three years, Agbe-Davies seeks to accomplish the complete transcription and markup of two daybook volumes sv133/61 (Blotter Commencing August 1808) and sv133/66 (Daybook 31 March 1885-13 November 1886). The former volume complements the one currently in process. Whereas the "Slave Ledger" includes only the purchases of and debts accruing to enslaved patrons, sv133/61 includes transactions by free customers. The post-Emancipation daybook will provide an important basis for comparison with earlier material.

Wheaton Accounts: Tomasek is at work on a full edition of the set of surviving daybooks, ledgers, cashbooks, and mill accounts associated with the personal and business interests of Laban Morey Wheaton, a member of the family that founded Wheaton Female Seminary (now Wheaton College) in 1834. Transcription and markup of a daybook dating from

1828 to 1859 was completed in summer 2016, and half of the eight associated cash books were transcribed in summer 2017. The associated ledger and remaining cashbooks will constitute the next set of documents to be transcribed, to be followed by a volume of mill accounts from 1847-48. A file based on a sample page from the daybook is part of the alpha version of the MEDEA site on GAMS; the full daybook edition will be ready for upload to GAMS in September 2017.

Future Editions

UVA Accounts: Martin is one of the co-founders (with Maury McInnis, Kirt von Daacke and Lewis Nelson) of the project, *Jefferson's University .. the early life* (JUEL, see: <http://juel.iath.virginia.edu/>). Over the last four years, JUEL has been making accessible documents from Special Collections at the University of Virginia, specifically documents from the first fifty years of the University. To date, JUEL has not undertaken the full transcription of the accounts documents. Future work will focus on transcription and markup of informal faculty accounts from the second quarter of the nineteenth century. These accounts were kept by teaching faculty at the University of Virginia in their additional roles managing construction, provisions, and personnel. Initial transcription will focus on the Proctor's Ledgers (1836-1837) and related material, with the resulting transcripts and identified people and places integrated into JUEL resources.

Schlitz Accounts: Eastberg will focus on account books from the Schlitz manuscripts. The initial ten account records that will serve as the basis for this project include representative and historically significant pieces from a much larger collection. Each account book represents a different aspect of the brewing company and together will provide vital information on its changing nature. The first three ledgers that will be completed by 2020 are the Schlitz Brewing Company General Ledger, 1887-89: 637 pages of accounts covering two years of income and

expenses for the company and its principal owners including everything from cooperage to donations and personal family expenses; the Schlitz Brewing Company Report of Materials Purchased for Brewing Beer (Malt, Barley, Hops, Maizone, Rice), 1913-1926, a 200 page summary ledger of materials purchased for brewing beer during those years; and the Schlitz brewing company's hired hands ledger, 1881-83, which includes the names and wages of over five hundred brewery employees.

Matthew Carey Accounts: These records include receipts, bills, memoranda, invoices, bills of lading, and other records of Carey's publishing house—arguably the most influential in the early Republic—and its successors: Carey, Lea, and Company; and Lea and Blanchard. Because of the heterogeneous nature and considerable volume of these accounts, when the American Antiquarian Society first made all 16,000 scans browsable and downloadable in a public-facing digital asset management system, they were of little use. The semantic markup offered by MEDEA, as well as the best practices that the cooperative offers, could change all of this. Because the AAS has done considerable work with the Carey papers already, they will not be starting from scratch when editing begins.

Synergies

The proposed Digital Edition Publishing Collaborative for Historical Accounts will leverage synergies that emerged from the editors' participation in MEDEA activities. Tomasek, Vogeler, and their colleagues have already advanced research on the data modeling that will help create scholarly, verifiable, and interoperable data taken from account books. The MEDEA ontology is compatible with the CIDOC Conceptual Reference Model <<http://www.cidoc-crm.org/>>. Vogeler currently teaches workshops in which he recommends use of the TEI @ana attribute to add references to the MEDEA bookkeeping ontology to the TEI markup of digital

scholarly editions of accounts. The addition of such references enables the extraction of an RDF serialization that then allows for comparison of the data on the semantic web. The references used with the @ana attribute include the following:

- #bk_entry
- #bk_amount
- #bk_quantity
- #bk_what
- #bk_to
- #bk_from
- #bk_debit
- #bk_credit
- #bk_between
- #bk_transfer

Far more powerful than simple aggregation, common use of the MEDEA ontology in participating editions will enable transformation to RDF on Vogeler’s model and will yield genuinely verifiable and interoperable data based on scholarly digital editions of accounts.

Vogeler has demonstrated the utility of semantic web technologies—RDF, RDFs/OWL, SKOS, and SPARQL—for comparing data from several projects that have presented economic content from the medieval and early modern periods on the web (Vogeler 2016).

Vogeler is the leading European scholar working on digital representation of historical account books. He combines expertise as an historian of medieval and early modern Europe with deep understanding of the underlying technologies of the semantic web. The web uses standards managed by the WorldWideWeb Consortium <<https://www.w3.org/>> to publish structured data in a way that allows machines to aggregate its “meaning” through use of common vocabularies or “ontologies” and by references between single data points across different web sites. The main tool for making these references is RDF, with which we can assign to any abstract concept or concrete object a URI and make statements about it by assigning properties with literal information or by linking it to other concepts. These statements have the form “triples,” which

can be read as “subject predicate object” sentences. The Web Ontology Language (OWL) adds functionalities of first order logic to these triples. With the help of the Semantic Web, scholars can publish controlled vocabularies, shared data description models, and formal knowledge representations. We can identify concepts and objects over the whole web, expand information stored in a database with knowledge from other resources, query aggregated datasets, and foster logical reasoning based on information published in this way.

Since the web is a network of information that can be connected through multiple reference points, representing historical accounting data using these technologies has the potential to create far more robust systems of information about the past than was possible using twentieth-century database technologies. Digital humanities scholars like Vogeler point out that the web is an appropriate space for historians to apply scholarly expertise and ensure that information on the web is accurate and reliable for scholarly research.

Names are one example of the kind of information that is commonly found in the many kinds of historical accounting records. A pioneering twenty-first century digital edition project demonstrated some of the advantages of adding an extra layer of interpretive markup about relationships among people to an edition of a significant source in the history of late medieval England. Through the addition of interpretive metadata about the relationships between individuals mentioned in the Henry III Fine Rolls, scholars working on the digital edition enhanced not only searchability of the source but also the quality of the information available in the online version (Ciula et al. 2008). Similarly, transcriptions of George Washington’s ledgers, the Stagville store blotters, and the daybook, cashbooks, and ledgers of Laban Morey Wheaton yield a plethora of names to which editors can add information about the people who bore those names, including relationships that are documented within the account books or associated

sources. Such names and additional information represent only a small portion of the information that might be of interest to scholars and the general public and will be made available as a result of publication through the Digital Edition Publishing Collaborative for Historical Accounts.

Alternative Solutions

Several institutions around the world are working on integrated solutions for digital scholarly edition frameworks: several are built upon eXist-db or other XML based technologies (XTF or Philologic); others employ dedicated solutions like kiln (King's College London), Collex (Charlottesville), eLaborate (Amsterdam), Salsah (Basel), FUD (Trier), EVT (Pisa), AustESE (Brisbane), and EVI (Madrid); and there are several adoptions of generic content management systems (like Drupal, Typo3). They all offer similar basic functionalities, but at the time of this writing we are not aware that any of the generic solutions have dedicated parts to support publication of historical accounts, as they usually treat editions as text, sometimes with elaborate indices. But none that we know of treat the numbers in editions as calculable values for economic analysis like GAMS does.

GAMS resembles Social Networks in Archival Context (SNAC), the research tool referenced in the Call for Proposals for the Digital Edition Publishing Cooperatives Initiative, in its application of emerging technologies to facilitate humanities research. It is compliant with Open Archival Information System (OAIS), which has become a library standard for archival storage of information (Lee 2010, Schumann and Recker 2012). OAIS employs the concept of a “package” of information in its recommendations for how to manage data stored in a repository. The project team for the Digital Edition Publishing Cooperative for Historical Accounts considers the consistent OAIS compliant approach of GAMS an important asset for three main reasons:

1. Scholarly editions are resources meant to be used over an extended period of time, so a solution that pays attention to long-term preservation has real advantages over those that do not.
2. The separation of submission and archival packages lends itself to setting up a distributed system in which several data creation solutions can contribute to the hub.
3. The separation of submission, archival, and dissemination packages allows representation of scholarly editions for interpretation of their multiple characteristics as texts as well as economic data by storing different data representations in one object and by adding dedicated dissemination tools.

The GAMS Workflow

In a series of blog posts (<http://dayofdh2017.linhd.es/GVogeler/>), Vogeler sketched the basic workflow of digital scholarly editing: data creation as imaging, transcription and enrichment; data testing by technical validation and visual check; and data publication and archiving as creation of a reliable publicly available resource. The GAMS infrastructure is built as part of this workflow, with publication and archiving at the core. The model implements the OAIS model for digital archives by separating the data creation technologies from the archival packages and the dissemination methods. This allows for multiple forms of data creation (e.g. transcription platforms, local databases, customized TEI etc.) to be processed during the ingest into the archive in order to separate the display software from the data. The workflow has substantial advantages for a digital edition platform for accounts:

1. It offers the opportunity to integrate complex analysis of the transcription during the ingest process. This includes expansion with authority files and transformations of customized data into standard data. GAMS supports this process with an open source Java client (“cirilo,”

<https://github.com/acdh/cirilo>), which offers data ingest from various sources (e.g.

spreadsheets, word processing files, customized TEI files, eXist-db-databases) and several processing pipelines (e.g. extract RDF from TEI, expand geo-references, extract DC—Dublin Core or METS—Metadata Encoding and Transmission Standard—metadata from TEI files).

2. It allows users to build dissemination methods around datatypes encapsulated in the archival packages. GAMS currently offers a wide range of disseminators for images, statistical datasets, METS, TEI, RDF, data streams: XSLT-stylesheets for TEI representations converting the source into different displays (HTML, PDF, ...) via Apache-Cocoon pipelines integrating LaTeX processing, R analysis of datasets, IIIF access to image data and image collections described via TEI and/or METS, and SPARQL queries on RDF data (see <http://gams.uni-graz.at/docs>). These dissemination methods include temporary user spaces in which users can build their own data sets from the data in the hub and download them for further processing.
3. It allows users to build generic discoverability tools, like making the data available via Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), offering the whole data on a SPARQL Protocol and RDF Query Language (SPARQL) endpoint, making it accessible via SOLR search engine.

These technical specifications mean that GAMS will offer the proposed Digital Edition Publishing Cooperative for Historical Accounts the ability to support a wide set of local data creation environments fitting the needs of each project. It will provide a preview area to support quality control technically, by validating against XML schemas and OWL ontologies, and manually, by displaying the data for the human reader. It will facilitate a set of local publication sites as well as a hub aggregating data from those sites and offer common search and analysis

tools, maintaining basic ontologies and common authority files and securing long-term preservation of the data published.

Planning Issues

In order to create a successful and sustainable digital publishing cooperative for historical accounts in the United States on the GAMS model, the edition cluster will need to consider a number of issues, including relationships among participating institutions and linkages to distributed collections of materials as well as identifying, developing, and testing financial/cost models that provide free online access to components and finding appropriate and cost-effective methods for sharing technical expertise across editions. Planning issues are treated in detail in the work plan, which includes a division of labor involving working groups devoted to technical and administrative matters, with the opportunity for individual team members to float among the groups. After two plenary meetings in January and February 2018, the administrative group—Vogeler, Colati, Hardy, Stertz, Boylan, and Tomasek—begins separate meetings to focus on questions related to sharing technical expertise across editions. Issues include replicability of the GAMS model as well as costs—both initial and ongoing, institutional relationships, and business models. The technical group—including staff from ZIM, UVA, and Brumfield Labs—will meet separately to focus on prototyping programs to convert data from existing editions for GAMS. Editors of existing and emerging editions will also meet with developers to discuss needs.

Since we take GAMS as our model, we must take into consideration the differences between a distributed set of edition projects that lack technical expertise in RDF/SKOS/OWL/SPARQL and a digital humanities center with a focus on information modeling (ZIM). GAMS is an asset management system built around the open source repository software FEDORA commons. It integrates several other open source solutions in a common

workflow as web services (including IIF, REST interfaces, and SPARQL federation). The technical infrastructure of GAMS is built with long-term sustainability in mind and is situated in an expanding scholarly environment at Graz University, but it is currently not operating as a commercially calculated host. Thus, the long-term sustainability of the services currently offered by ZIM has to be studied. Do the institutions represented in the collaborative have access to the skills and resources necessary to adopt the concepts and technologies of GAMS? Are the web services included in GAMS appropriate to the technical environment of the other partners? What kind of contractual framework will the collaborative need? How will the transnational nature of the proposed partnership affect that framework? How will future editions be incorporated into the collaborative? How will this cooperative leverage the collaboration with other cooperatives funded through this initiative?

We propose Wheaton College as the U.S. host for the proposed Digital Edition Publishing Cooperative for Historical Accounts. Library Services at the college has undergone a thorough restructuring in the past two years, and the Interim Dean of Library Services works closely with the Provost. Staffing models are under active review, and the necessary systems administration skills are available at the college. Current Library Services staff should be able to provide 10% FTE, and we are cautiously optimistic that we will be able to meet the approximately 25% FTE needed to maintain the services of the publishing cooperative through the implementation process. Since GAMS is a well designed system structured specifically for storage of humanities data, its installation is a matter of a day's work for a systems administrator who is familiar with Java and associated technologies. We will use the planning phase of the initiative as an opportunity to explore how Wheaton can ensure its success as a host institution. The administrative group will examine business models that will make a systems administration

position for the cooperative self-sustaining. At ZIM, a resident community of graduate level technologists trained through coursework at the university is available for employment on an academic term basis. ZIM charges a fee that covers the graduate employees' stipends for the term, offering storage of images of original documents for an additional fee. The administrative group will consider whether such a business model is feasible at Wheaton College. Might a tiered model with free basic services and different fee levels based on time required by new editions prove a practical solution? What kinds of agreements will be needed between the host institution and participating editions? What kinds of technical requirements will facilitate linking to image files located remotely? Where will the collaborative find the population of technically adept support staff that will be needed? Can the college train student employees to provide some of the services? In addition, we will need to work closely with ZIM to identify and develop appropriate protocols for incorporating linked open data resources, full-text materials, metadata, annotations, and other core informational components such as images of original documents. Files created during the award period will be permanently stored at ZIM as part of the MEDEA data set.

Since MEDEA methods are relatively new to the potential user community of scholars engaged in book history, linguistics, and social and economic history in the United States, we will need to develop sample uses of the editions we produce as a way to explain the utility of the model. The GWFPP has created one excellent example of a website that makes the information from account books accessible and legible to the general public. Tomasek plans to use GAMS features to produce an interactive interface for the Wheaton accounts tentatively entitled Digital Norton. This online publication through GAMS will incorporate images of historical maps with the data from the digital scholarly edition of Wheaton accounts to allow users to follow

individuals referenced in the accounts and connect their stories. Digital Norton will be only one example of a possible use of digital editions of accounts. Agbe-Davies's publications about the enslaved and freed populations at Stagville will offer another model. Vogeler will continue to incorporate future digital scholarly editions of accounts into the MEDEA data model, in the hope of inspiring new scholarly uses of the editions.

Conclusion

One of the most exciting elements of digital scholarly editing is the fact that we cannot really know how future researchers will use our editions. The proposed cooperative's project team is committed to developing a user-friendly publishing environment that allows ingestion of data without extensive coding expertise and that produces the kinds of outputs desired by current editors. We seek also to use the publishing cooperative as a bridge between the worlds of documentary editing in the United States and those of digital history and digital humanities around the world. The proposed Digital Edition Publishing Cooperative for Historical Accounts seeks to ensure the accuracy and reliability of information from historical accounting records on the web by promoting historians' intentional use of semantic web technologies. Many scholars in digital humanities express a hope that future researchers will encounter the results of our work and create something original that we cannot currently begin to imagine. Many librarians and archivists in digital humanities communities are convinced that researchers in fields beyond history stand ready to reuse data from historical accounts as soon as we produce it. Developing stable and sustainable publishing models is an important step as we continue to build mechanisms to create historically accurate "big data" for use by scholars, students, and the general public through the publication of scholarly digital editions of accounts.

Selected Bibliography--Digital Edition Publishing Cooperative for Historical Accounts

- Allen, R. C. 2001: The Great Divergence in European Wages and Prices from the Middle Ages to the First World War. In: *Explorations in Economic History* 38, 411–447.
- Allen, R.C. 2013. The High wage Economy and the Industrial Revolution. A Restatement. In: *Oxford Economic and Social History Working Papers* (Ref: Number 115), <http://www.economics.ox.ac.uk/index.php/Oxford-Economic-and-Social-History-Working-Papers/the-high-wage-economy-and-the-industrial-revolution-a-restatement> (accessed August 30, 2013).
- Allen, R.C., Clark, G., Devereux, J., Hellie, R., Hoffman, P.T., Jacks, D. S., Lindert, P. H., Ma, D., Mironov, B. N., Pamuk, S., Van Zanden, J. L., Ward, M. 2004. Preliminary Global Price Comparisons 1500-1870. In: Lindert, P. H. et. al.: *Towards a Global History of Prices and Wages*, 19-21 Aug. 2004, <http://www.iisg.nl/hpw/conference.html>
- Arkell, T. 2003. Identifying regional variations from the hearth tax. *The Local Historian* 33:148-74.
- 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.
- Graham, Shawn, Ian Milligan, and Scott Weingart. 2016. *Exploring Big Historical Data: The Historian's Macroscope*. London: Imperial College Press.
- Guidelines for Electronic Text Encoding and Interchange, originally edited by C.M. Sperberg-McQueen, Lou Burnard, version 3.1.0. 2016 <http://www.tei-c.org/Guidelines/P5/>.
- 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. <http://conf2016.jadh.org/abstracts/s5-2/>
- Lee, Christopher A. 2010. Open Archival Information System (OAIS) Reference Model. *Encyclopedia of Library and Information Sciences*, Third Edition. Taylor and Francis. DOI: 10.1081/E-ELIS3-120044377.
- Manning, Patrick. 2017. Inequality: Historical and Disciplinary Approaches. *American Historical Review*, 122 (1): 1-22. DOI: <https://doi.org/10.1093/ahr/122.1.1> Published: 31 January 2017. Accessed: 17 February 2017.
- Neave, D., S. Neave, C. Ferguson, and E. Parkinson. 2015. *Yorkshire East Riding Hearth Tax, 1672-73*. London: British Record Society.

Pomerantz, Kenneth. 2000. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*. Princeton, N.J.: Princeton University Press.

Ruge, Nikolaus. 2015. Linguistic Perspectives on Medieval Account Books – The comptes de la baumaîtrie of the City of Luxembourg (1388-1500). <<http://medea.hypotheses.org/304>>

Schuman, Natascha, and Astrid Recker. 2012. Benefits and Challenges of Mapping the OAIS Reference Model to the GESIS Data Archive. *IASSIST Quarterly*. International Association for Social Science Information Service and Technology.

Seaman, P.J., John Pound, Robert Smith. 2001. *Norfolk hearth tax exemption certificates 1670-1674 : Norwich, Great Yarmouth, King's Lynn and Thetford*. London: British Record Society.

Steiner, Elisabeth, and Johannes Stigler. Cirilo Client. Software Reference and Tutorial, Graz 2011–2015, <http://gams.uni-graz.at/doku>

Stertzer, Jennifer. 2014. Working with the Financial Records of George Washington. Document vs. Data, in: *Digital Studies* 4 <http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/247/308>.

Tomasek, Kathryn, and Syd Bauman. 2013. Encoding Financial Records for Historical Research, in: *jTEI* 6 <<http://jtei.revues.org/895>>, DOI : 10.4000/jtei.895. Japanese translation by Naoki Kokaze: <<http://hdl.handle.net/2261/56940>>.

Tryon, Warren S. and William Charvat. 1949. *The Cost Books of Ticknor and Fields and Their Predecessors, 1832-1858*. New York: Bibliographical Society.

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

Wareham, Andrew. 2015. Using the Digital Humanities to Analyse Non-Standard Information in the Late Seventeenth-Century Hearth Tax Returns. <<http://medea.hypotheses.org/279>>.

Zboray, Ronald J. 1993. *A Fictive People: Antebellum Economic Development and the American Reading Public*. Oxford University Press.