

The Model Editions Partnership: Putting TEI Theory into Scholarly Practice

C. M. Sperberg-McQueen and David R. Chesnutt

C. M. Sperberg-McQueen, University of Illinois at Chicago, Computer Center (M/C 135), 1940 W. Taylor St., Chicago IL 60612-7352

David R. Chesnutt, Editor, Laurens Papers Project, University of South Carolina, Department of History, Columbia, SC 29208

KEYWORDS: TEI, scholarly editions, history

AFFILIATION: University of Illinois at Chicago
and University of South Carolina

E-MAIL: tei@uic.edu
 david.chesnutt@sc.edu

FAX NUMBER: +1 (312) 996-6834

PHONE NUMBER: +1 (312) 413-0317

This paper describes recent work in the Model Editions Partnership, a cooperative effort of seven historical editing projects to develop adequate intellectual and practical models for electronic historical editions. In particular, we will report on the development of a TEI-conformant markup scheme for use in the samples of historical editions to be prepared in later stages of the project by the various partner projects. The paper will have five parts: an introduction to the partnership as a whole, a discussion of the document analysis process with which the project began, a summary of the requirements derived from that document analysis for markup of historical documents in the context of historical editions, a demonstration of the resulting modifications of the markup scheme developed by the Text Encoding Initiative (TEI), and a brief description of the future work of the partnership.

The paper is relevant to the areas of concern to attendees first because the Model Editions Partnership is a major effort of the historical documentary editing community in the United States to address the issues raised by the advent of electronic distribution methods and SGML. The paper will also provide an extended concrete example of the use of the TEI Guidelines in an important specialized application area, complete with local customization and extension of the DTDs.

1 The Model Editions Partnership

The Model Editions Partnership is a cooperative

project of seven major historical documentary editing projects in the United States:

- the Documentary History of the First Federal Congress of the United States of America (George Washington University, Washington, D.C.), a letterpress edition of all extant documents relating to the First Federal Congress
- the Documentary History of the Ratification of the Constitution and the Bill of Rights (University of Wisconsin, Madison), a selective letterpress edition of all important documents illustrating the debate, in each state, over the ratification of the Constitution, the passage of the Bill of Rights by the first Congress, and the ratification of the first ten amendments
- the Papers of General Nathanael Greene (Rhode Island Historical Society, Providence, Rhode Island), a letterpress edition of the extant correspondence and other papers of General Nathanael Greene, a prominent military figure in the U.S. Revolution
- the Papers of Henry Laurens (University of South Carolina, Columbia, S.C.), a selective letterpress edition of the papers of a prominent South Carolina planter and leader of the American Revolution
- the Lincoln Legal Papers (Illinois State Historical Library, Springfield, Illinois), a complete electronic facsimile edition of all extant documents relating to the legal practice of Abraham Lincoln and to cases in which he appeared
- the Margaret Sanger Papers (New York University, New York), a complete microfilm edition of all known papers of the birth control pioneer which were not included in the microfilm series issued by the Library of Congress, together with indices both to the new microfilm series and to the Library of Congress series
- the Papers of Elizabeth Cady Stanton and Susan B. Antony (Rutgers University, New Brunswick, N.J.), a selective letterpress edition of correspondence, speeches, and other writings of the two leaders of the U.S. woman suffrage movement.

The project is coordinated by David Chesnutt (South Carolina), Susan Hockey (Princeton and Rutgers), and C. M. Sperberg-McQueen (University of Illinois at Chicago).

2 The Prospectus for Historical Editions

In the first phase of its work, the Model Editions Partnership has attempted to clarify the implications of electronic editions for historical documentary editing as now practiced in the U.S.

Modern historical editing in the U.S. dates from the publication of Julian Boyd's first volume of *The Papers of Thomas Jefferson* in 1950. Although there had been earlier compilations of the papers of famous Americans, his carefully prepared texts of Jefferson's letters and other writings, "warts and all," set a new standard for accuracy and reliability. His equally careful selection of what to include and what not to include reflected the historian's thoughtful appraisal of what needed to be set before readers so they could begin to understand the essential Jefferson. And Boyd's incisive commentary provided the context needed to place Jefferson in the wider world of the American Revolution and the early national period of American history.

Boyd's greatest legacy, however, was the model he set for the generations of historians who followed in his footsteps – editing the letters and documents of a broad range of individuals who played roles, both large and small, in creating the new nation. The number of book and microfilm editions which followed in the wake of Boyd's Jefferson made available hundreds of thousands of historical documents gathered from repositories on both sides of the Atlantic and sometimes, both sides of the Pacific. As historical editors move toward the age of the digital library, however, they face the challenge of developing new kinds of editions in which to present the letters and documents which explain our past. The Model Editions Partnership is a first step toward meeting that challenge.

Principles for Designing Electronic Editions

Although markup is essential, the first task of the editors was to develop a set of principles to govern the creation of the markup scheme for the models. The partnership agreed on five principles of design for an electronic edition.

- The design must accommodate current scholarly editorial practice.
- The design must accommodate changes in editorial practice.
- The design must accommodate post-publication enhancements.
- The design must accommodate multiple forms of publication.
- The design must be based on relevant, non-proprietary standards.

These principles are discussed at some length in a "Prospectus for Electronic Historical Editions" posted on the Web site of the partnership (<http://mep.cla.sc.edu>). The basic point to be made here is that the editors wanted a set of clear statements which emphasize that scholarly, not technical, criteria should govern the development of editions in an electronic environment.

The first three principles address the editors' scholarly concerns. "Current scholarly editorial practice" encompasses the basic practice of providing reliable texts of the documents, adequate commentary to explain their historical context, and tools like good indices to provide intellectual access – all hallmarks of the Boyd tradition. Well-designed markup must accommodate those practices. "Changes in editorial practice" will certainly emerge as editors learn to take advantage of the electronic environment. Even in this early stage of the partnership, the editors have begun to realize that new forms of annotation and commentary are both possible and desirable in electronic editions. Other changes in the way in which editions are organized or made accessible to readers are also likely. "Post-publication enhancements" refer to the ability to integrate newly-found documents, to correct misreadings of an existing text, or perhaps to create a subset of the larger edition which can be used as a classroom reader. Although we cannot anticipate all of the needs of future generations of scholars, we can design our editions so that the texts are both durable and reusable.

The last two principles address more practical matters. Because we are in an age of transition in which historical editions may continue to be published in book and microfilm editions, markup which accommodates "multiple forms of publication" is important. Well-designed markup will enable projects that begin as microfilm editions or projects that begin as book editions to migrate smoothly to image editions or live-text editions on CD-ROM or the Web. "Non-proprietary standards" are essential if long-term resources are to survive in midst of rapidly changing technology. De facto proprietary standards are simply too volatile. WordStar 3.3 became a kind of de facto standard for text files in the mid-1980s. Almost any word processing software could read those files, a situation which is no longer true. For text files, the partnership will use the SGML standard inherent in the TEI Guidelines. Relevant standards for images and other digital resources have yet to be determined.

Currently, discussions of editing practice and types of editions use the nature of the printed product as a touchstone, and it can be hard to disentangle characteristics of our current editions which are imposed by the nature of the material and the intellectual requirements of users from those which are imposed by the nature of current typesetting.

A Typology for Electronic Editions

In addition to articulating principles to guide the design for editorial markup, the editors defined a typology for electronic historical editions. Editors have a well-developed shorthand for describing

current editions. "Microfilm editions" contain images of documents and usually have indices and other access tools, but limited commentary. "Selected Letterpress Editions" are usually based on microfilm editions and tend to present a small sample of the documents. They include transcriptions of selected documents, extensive commentary, indices and other editorial apparatus. "Comprehensive Letterpress Editions" are generally understood to be more exhaustive and may or may not be based on previous microfilm editions. (The latter descriptions were developed in the 1950s when volumes were typeset in hot metal and printed on letterpresses.) A somewhat analogous typology is set forth in the partnership's Prospectus: Image Editions, Live-Text Editions, Combined Editions and Transitional Editions.

Image editions are envisioned as editions which present images or pictures of historical documents linked to control files and other types of scholarly apparatus. Control files usually identify each document, the date it was created, and the repository which holds the original. Other bits of information may also be included (the author and recipient of a letter, copyright information for modern materials, etc.).

Two of the partner projects, one addressing papers of Abraham Lincoln and the second the papers of Margaret Sanger, fall into this category. These projects are creating "silicon microfilms" and both editions will go far beyond current microfilm editions by providing greater supplementary information; by allowing users to define subsets of the documents to suit their particular interests; and by eliminating the necessary tedium of cranking through reel after reel of film to reach the documents. The Lincoln Legal Papers is creating a CD-ROM edition of letters and documents relating to Lincoln's career as a lawyer before his election as president. The editors have amassed more than 250,000 photocopies from which they are creating digital images. Extensive database files are used to provide item-level control of the collection and to provide information about the individuals involved, the types of cases at issue, written summaries of the cases, and many other kinds of information. These database files will be used to retrieve documents and to provide supplementary information about the documents.

The editors of the Margaret Sanger Papers are planning an equally interesting image edition to bring together three discrete microfilm collections totaling more than 300,000 pages. The Sanger databases provide item-level access as well as links to four related research files created by the project: a chronology of Sanger's day-to-day activities, biographical sketches of prominent correspondents, copyright information regarding individual correspondents, and repository holdings for

the documents. Hypertext links will also bring together letters, enclosures, and referenced documents which are now separated in the three microfilm collections.

Live-text editions will contain searchable ASCII transcriptions of documents. They are seen as being somewhat like letterpress editions, since transcriptions of documents, commentary, indices and other scholarly apparatus would form the core of this kind of edition. The editors of the Papers of Elizabeth Cady Stanton and Susan B. Anthony are in the initial phases of creating a six-volume, selected letterpress edition. Because they are using computers to prepare the texts and supplementary materials for print publication, the creation of a "live-text" edition seems eminently feasible.

Combined editions would include both images of the documents as well as live-text transcriptions of the documents. This type of edition would be well suited for a small classroom "reader" which provided students with both images and transcriptions of seminal documents like the Declaration of Independence. This is a concept which has been incorporated in some of the Library of Congress exhibits like the Walt Whitman journals and Lincoln's Gettysburg Address.

Transitional editions are seen as a way of bringing together existing letterpress volumes and subsequent volumes or supplements for which live-text exists. Editions like the Ratification project, the First Congress project, the Laurens Papers, the Greene Papers and almost every major edition began many years ago before computers were used to prepare material for the printed volume. Most of these projects also began using word processing software in the early 1980s and thus have electronic files for their more recent volumes. Because creating live-text for the early volumes does not seem financially feasible at this time, we have proposed that this type of edition combine images of the printed pages for the early volumes with live-texts for the later volumes or supplements. Printed indices for the earlier volumes would be combined with the electronic indices of later volumes to produce a comprehensive index for the entire edition. Page numbers in the comprehensive index would point either to page images or to live-text as appropriate. Our preliminary experiments indicate that this kind of edition would be both a technologically feasible and cost-effective approach to making entire editions available in an electronic form deliverable on CD-ROM or over networks.

3 Requirements for a MEP Markup Scheme

The markup language to be used by the Model Editions Partnership must meet several requirements:

- It must conform to the Guidelines issued by the Text Encoding Initiative.
- It must select among the tag sets and tags of the TEI those which are to be used for MEP materials.
- Where the Guidelines offer several possible methods of encoding the same phenomena, it must (where feasible) choose among those methods.
- It must extend the TEI markup language to handle document types common in historical editions but not fully treated in the current version of the Guidelines, or to handle specialized types of annotation or analysis important for historical editions.

The necessity of selection, choice among alternatives, and extension of the existing tag sets are common to all users of the TEI Guidelines, and so the account of how the Partnership has gone about specifying its particular selection and extension of the Guidelines should be of interest to other potential users of the TEI. In the case of MEP, the customization of the TEI Guidelines is driven by the types of documents included in the editions, the types of interests which editions must serve, and the variations among editions in transcription practice.

Historical editions may contain virtually any type of document, but in practice at least half, and often ninety percent, of the documents in a conventional historical edition are letters and their enclosures. Official documents (e.g. legislation or, for legal editions, court papers) and newspaper accounts are also relatively common. The TEI Guidelines have only minimal advice for those encoding historical letters, and virtually no advice for those encoding official documentary records. New tags, and refinements of existing tags, will be useful in ensuring that the sender(s), recipient(s), date(s), and similar critical attributes of letters are readily identifiable from the markup.

Because the letters in historical editions are typically not published elsewhere, it is essential that the edition systematically provide relevant information about each document, including the repository where the original is held, the length and physical state of the original, etc. Specialized tags for the description of the source document (in the TEI header), and for providing repository information (associated with each transcription), will be very useful.

Electronic image editions will need to document the creation of their electronic images in the same way that electronic transcriptions of text document their encoding principles and editorial practice; the TEI header will need extensions to deal with non-textual, non-SGML data, in order that it can be used to document images, and not just transcriptions, of documents.

The paper will identify in some detail the most important textual features requiring extension of the TEI markup scheme, and provide examples both of documents exhibiting these features and of the proposed markup.

4 Modifying the TEI Markup Scheme

The paper will then illustrate in very concrete terms the work needed to incorporate extensions and modifications of the sort needed by MEP into the SGML document type definition provided by the TEI. Several files are needed for this purpose; each file, or portions of it, will be shown and discussed.

- * a file with redefinitions of some TEI element classes and other parameter entities
- * a file with the new and modified SGML element declarations
- * a driver file for the MEP document type definition
- * a tag-set documentation file, providing detailed description of the elements added or modified by MEP, in the form familiar to readers of TEI P3 from the second volume of the Guidelines

In addition to these files, which work together with the standard TEI DTD, a free-standing version of the MEP DTD, in a single file, will be generated. Such free-standing versions of the TEI DTD, including all selections and modifications of the TEI markup scheme, are useful because they are less confusing to read, and may be simpler to process, than the full TEI DTDs together with the local modification files.

5 What Next?

The paper will conclude with a look forward to the remaining work of the Model Editions Partnership. After the completion of the MEP markup scheme, the partnership will encode sizable samples (50 printed pages or more) from each of the partner editions, using the markup scheme described here. In later phases, the samples will be distributed together with search and display software in a CD-ROM edition of the MEP sampler. The CD-ROM edition should demonstrate various approaches to CD-ROM-based publication of electronic editions. A companion effort will involve the creation of an Internet-based version of the sampler, which will demonstrate approaches to network distribution. The final phase of the project will involve the completion of extensive documentation of the work of the partnership, in order that the lessons of this project's experimentation can be readily accessible to scholars in the field.