

## Letters and Lacunae: Editing an Electronic Scholarly Edition of Correspondence

---

**Susan Schreibman** (*sschreib@umd.edu*)

*University of Maryland*

**Gretchen Gueguen** (*ggueguen@wam.umd.edu*)

*University of Maryland*

**Amit Kumar** (*amitku@uiuc.edu*)

*University of Illinois at Urbana Champaign*

**Ann Saddlemeyer** (*sadlemey@uvic.ca*)

*University of Victoria*

---

Encoding editions of documentary texts, particularly editions of correspondence, within the *Text Encoding Initiative (TEI) Guidelines* raises special challenges not encountered when editing previously published works. The challenges fall into three broad categories: 1) difficulties in capturing bibliographic meta-information describing the physical object and its transmission history; 2) challenges in developing a controlled vocabulary suitable to the informal nature of texts which were never intended for publication; and 3) difficulties in encoding both physical characteristics of the documentary texts, as well as their intellectual content, i.e. adopting a principle of encoding the text either as a physical artifact or as a conceptual work. These challenges, particularly as they relate to encoding letters, will be explored by through an edition currently being edited entitled *Thomas MacGreevy and George Yeats: A Friendship in Letters*.

During the next two years members of *The Thomas MacGreevy Archive* team will be creating for online publication an edition of the correspondence between George Yeats (1893-1968), wife of the Irish poet W.B. Yeats, and Thomas MacGreevy (1893-1967), Irish poet, art and literary critic, and Director of the National Gallery of Ireland (1950-63). It is a collection spanning 41 years, comprising 148 letters. The letters are fascinating documentary records which provide a window not only into the personal lives of the authors, but into the artistic and political circles in which they moved, providing a unique insight into the new Irish Free State and the cultural climate of Europe during the first half of the twentieth century. The letters are being encoded using Extensible Markup Language (XML) according to newly released *P5 TEI Guidelines* to take advantage of the TEI's new chapter on Manuscript Description.

Although the *TEI Guidelines* were not developed specifically to encode previously published texts, many of the rules built into the syntax of the Document Type Definitions (DTDs) favor this document type. To cite but one example, the content model of `tei.divbot` does not allow for a paragraph `<p>` element after the closer element `<closer>`. While the need for additional paragraphs after closing material in published texts may be uncommon, letters frequently have a closing salutation, followed by a postscript. Moreover, it has proved difficult within the TEI header to detail the type of descriptive information that editors, scholars, and bibliographers require when engaging with handwritten documents.

Individual projects (such as *DALF: Digital Archive of Letters in Flanders Project*) and subject-area consortiums (such as *The Model Editions Partnership*) have developed their own extensions to the *TEI Guidelines* to accommodate the needs of electronic editions of correspondence. After a brief survey of the strategies employed by these and other editions, we will discuss how TEI's new chapter on manuscript description alleviates some of the problems previous projects solved with local solutions. The chapter on Manuscript Description builds on the work of two separate initiatives which have been recently combined: *MASTER* project (1999-2001), an EU-funded project headed by Peter Robinson, and the work of the *TEI Medieval Manuscripts Description Work Group* (1998-2000), headed by Consuelo Dutschke and Ambrogio Piazzoni. The new elements available in this tagset provide for detailed description of primary texts including transmission, physical description, the relationship between parts of the manuscript (for example, when a poem is enclosed with a letter), dimensions, location, manuscript identification, provenance, and history of ownership.

Another area to be discussed is the difficulties in developing an ontology or controlled vocabulary for a correspondence. The ontology, the backbone for the search page, is more difficult to develop for a collection of letters than other document types. Subject headings, such as the *Library of Congress Subject Headings (LCSH)*, which are used to describe entire collections or self-contained bodies of information, are not suitable for this project which describes each letter individually. The problem with using schemes such as *LCSH* is twofold: one, the letters cover many subjects and follow no formal organization pattern, making it difficult to make a faceted indexing schema like *LCSH* worthwhile; secondly, the subject headings were meant to be used in the cataloging of cohesive works or collections, and were not designed to be brief entries in the index for a specific work or collection.

The indexing done for this edition more closely resembles back-of-the-book style indexing in terms of its description of the details of the text. Standard controlled vocabularies that might be used in this type of indexing, like the *Getty Art and Architecture Thesaurus*, on the other hand, are too specific and

terms do not sufficiently summarize or categorize the topics discussed. Capturing, representing, and, indeed, interpreting a multitude of topics present in any given letter — from general subjects to more intimate personal details — is of paramount importance. If ontology is defined as a "formal, explicit specification of a shared conceptualization" (Fensel 11), the burden of interpreting by a third party what a "shared conceptualization" of a text written for an intended audience of one is immense. Indeed, as the correspondence itself often indicates, meaning is often misconstrued by the intended recipient. Given these difficulties, other types of structured data, such as annotation and abstracts, may be used to mitigate issues of keywords conveying different meanings when taken out of textual context.

Another challenge when editing documentary texts for electronic publication is choosing a philosophy by which to encode. This is particularly true in the case of editing modern correspondence. Editors have had to traditionally decide whether the purpose of the encoding is to capture the physical appearance of the page (regardless of the text's logical sequence), or whether it is to record the textual/ontological flow (regardless of the text's physical appearance). In traditional print publications, editions (except for facsimiles) reflect a logical sequencing of the text. For example, text which appears in the margins is placed where the editor feels it belongs logically, even when the writing crosses page boundaries (such as finishing a letter in the margins of the first page when the author ran out of room on the last).

This edition is exploring methods of encoding both the physical appearance of the page, as well as the letter's logic. This is particularly challenging when encoding, for example, marginalia. To represent the marginalia within the logical sequence of the text, the editor must decide where it is to be anchored within the textual flow. To represent it in a physical representation, the editor must provide coordinates that will anchor the text vertically and horizontally in relation to the main body of the work. While some of this positioning is absolute, for example, anchoring text at the top of the page, other positioning is relative, for example, anchoring marginalia relative to the paragraph it appears next to. While the encoding must take into account, in some measure, the technologies available to us today, XSLT, CSS, and JavaScript, for example, at the same time it must also be encoded with a view to future presentations, independent of current technologies.

These are a sampling of issues that will be discussed.

## Bibliography

Chestnutt, R. David. "The e Model Editions Partnership: 'Smart Text' and Beyond." *DLib Magazine* (July/August 1997). <[http](#)

>[://www.dlib.org/dlib/july97/07chesnutt.html](http://www.dlib.org/dlib/july97/07chesnutt.html)>

DALF: *Digital Archive of Letters in Flanders Project*. Centrum voor Teksteditie en Bronnenstudie (KANTL). Accessed 2005-03-21. <<http://www.kantl.be/ctb/project/dalf/>>

DeRose, Steven J., David G. Durand, Elli Mylonas, and Allen H. Renear. "What is Text, Really?" *Journal of Computing in Higher Education* 2.1 (Winter) (1990): 3-26.

Farrow, John. "All in the Mind: Concept Analysis in Indexing." *The Indexer* 19.4 (1995): 243-247.

Fensel, Dieter et al. *Spinning the Semantic Web*. Cambridge, Massachusetts: MIT Press, 2005.

Matthews, Douglas. "Indexing Published Letters." *The Indexer* 22.3 (2001): 135-141.

Renear, Allen H., Elli Mylonas, and David G. Durand. "Refining Our Notion of What Text Really Is: The Problem of Overlapping Hierarchies." *Research in Humanities Computing 4: Selected Papers from the ALLC/ACH Conference, Christ Church Oxford, April 1992*. Ed. Susan Hockey and Nancy Idle. Oxford: Oxford University Press, 1996. 263-280.

Schreibman, Susan. *The Thomas MacGreevy Archive*. Accessed 2005-03-21. <<http://macgreevy.org>>

*TEI Guidelines P4*. Accessed 2005-03-15. <<http://www.tei-c.org/Guidelines2/index.html>>

*TEI Guidelines P5, Manuscript Description Chapter*. Accessed 2005-03-15. <<http://www.tei-c.org/Activities/MS/FASC-ms.pdf>>