

This paper was presented at Digital Resources for the Humanities & Arts:
Dynamic Networks of Knowledge and Practice – Context, Crises, Futures, held
on September 7–9, 2009, at Queen’s University Belfast, Northern Ireland, United
Kingdom.

Creating a Digital Scholarly Edition of the British Library's *Cotton MS.*
Vitellius F. v.
Kevin S. Hawkins

Introduction

Henry Machyn's *Chronicle* of daily life in London from 1550–1563 is a unique resource for historians and linguists studying the 16th century. The *Chronicle* is especially notable because it covers the entire reign of Mary I, a tumultuous period in English history. No modern edition of the text has ever been published, and the last edition, published in 1848, was immediately recognized as deficient for scholarly analysis.

A scholarly edition of this text has been published¹ by the Scholarly Publishing Office (SPO),² a division of the University of Michigan Library where I work. This project is unlike other SPO publications in that it presents primary rather than secondary or tertiary sources, and it utilizes markup and SPO's delivery system in innovative ways. This edition includes a detailed introduction, images of the manuscript, a transcription including supplied text from other sources, a modernization of the text, and images of a 19th-century handwritten transcription of the original manuscript. The project has been marked by complex relationships among the various stakeholders, which evolved over the years and contributed to the project's long gestation period. I hope that SPO's experience with this project will help all embarking on electronic publishing projects avoid similar problems in the future.

First, some background on SPO's publishing services.

The Scholarly Publishing Office

The mission of the Scholarly Publishing Office is to serve the scholarly community—both scholars at Michigan and at other institutions—by providing sustainable electronic publishing services, supporting local control of intellectual assets, and exploring opportunities to extend and disseminate library collections. We prefer to publish through open access but also host subscription-based publications.

Nearly all SPO publications are hosted online using the DLXS suite,³ which includes XPAT (an XML-aware search engine) and middleware to provide a browser-accessible interface to the content. DLXS was designed as an all-purpose digital library platform for digitized collections, including monographs, serials, finding aids, bibliographies, and images. Broadly speaking, DLXS can deliver two kinds of textual content: page images with OCR and fully encoded text. Except for a publication's homepage and other contextual information, all other webpages—serving content, browse lists, and search results—are dynamically generated by the system.

Most of SPO's publications are journals or monographs, delivered as encoded text or page images. The Machyn edition is unique among SPO publications in that it includes both page images and encoded text, the latter with features not found in other SPO publications.

Next, a summary of the features of the Machyn edition published by SPO.

¹ See <<http://quod.lib.umich.edu/m/machyn/>>.

² See <<http://www.lib.umich.edu/spo/>>.

³ See <<http://www.dlxs.org/>>.

Structure of the online scholarly edition

The main entry point of the scholarly edition is based on SPO's template for journal homepages. It contains an image copied from Early English Books Online (EEBO) and also contains brief introductory text. From here the user can read about the edition, browse and search it, or access other resources related to Henry Machyn, including page images of the previously mentioned 19th-century handwritten transcription of the original manuscript.

The scholarly edition itself can be browsed or searched from the homepage, leading to dynamically generated webpages. The edition contains an introduction with illustrations from various sources,⁴ information on how to use the online edition, and the diary entries themselves. Each entry contains an “enhanced transcription”—a diplomatic transcription with supplied text in red—followed by a modernization of the text in contemporary orthography. The dates for each diary entry were regularized and supplied by the editors. Access to the manuscript images is through page break indicators in the enhanced transcription—a blending of the page-image and encoded-text models of DLXS introduced as a customization for this publication. Supplied text used to have no distinctive appearance in DLXS, but we changed the middleware to display it as red by default for all publications hosted in DLXS at Michigan and, consequently, at other institutions using a recent version of DLXS.

Searches can be restricted to the transcription only, the enhanced transcription (which includes supplied text), or the modernization only. Due to a limitation of the XPAT search engine used in DLXS, searching matches only exact phrases in the underlying XML, so intervening XML tags limit the recall in phrase searches.⁵ However, XPAT supports stemming with the asterisk wildcard at the end of a query, a distinct advantage to using XPAT instead of another search engine for searching text in languages with inflectional affixes.

Under “Other resources” is a short explanatory text and link to page images of the 19th-century manuscript. Since these images could not be OCR'd and no one has transcribed this manuscript, the text is not searchable, and the images are simply provided for reference.

None of this is particularly revolutionary to any of us familiar with online scholarly editions. But what makes this project unusual is the complex relationship among the project's stakeholders. Let me give a short history of the project, showing how the plan changed many times before ending up as what I've just described.

Involvement of the University of Michigan Press and ACLS

Richard Bailey, one of the editors, began discussions in 2000 with the University of Michigan Press about publishing the material prepared up to that point. The Press felt that a modernization of the text would be more widely read than a transcription of the original, so he and Colette Moore, another editor, prepared this version. In January 2003

⁴ Permission was obtained from the holding institution for all of these images as well, though it's questionable whether this is necessary since copyright claims to reproductions of public domain works are tenuous according to US law. Often the holding institution instead charges for the right to reproduce the image, or charges a fee for them to reproduce it for you.

⁵ This is a known shortfall in XPAT, but building a work-around is difficult and has not yet been accomplished.

Richard Bailey signed an agreement with the Press to publish “the work,” which was taken to mean just the modernization. By signing this agreement, the editors transferred their copyright to the Press, and the Press agreed to edit and publish the work.

The Press later entered into an agreement with the ACLS History E-Book Project (now called “ACLS Humanities E-Book”), which SPO coincidentally hosts in DLXS as a subscription-based resource. According to this agreement, ACLS would distribute this book online to subscribers of the project while the Press would retain copyright in the work. While the page count was no longer a concern for the Press after entering into this agreement, it still did not want to invest in editing the transcription and supplied text and therefore was not interested in publishing this portion unedited.

The role of SPO

The Press’s decision not to publish the transcription with supplied text led Richard Bailey to approach the Scholarly Publishing Office about publishing this other material separate from the Press/ACLS version. In addition, he wanted to deliver the images of the original manuscript online, plus images of the 19th-century handwritten transcription, which he had recently purchased. SPO accepted this project because supporting publishing projects from university faculty members is a part of its mission and because the DLXS suite is suited to delivering both page images and more heavily encoded text.

ACLS, being aware of SPO’s role in publishing a parallel edition, was excited about the opportunity to experiment with linking between the two editions (between each entry’s transcription and modernization) and to provide links to the page images in both editions. While the two editions would be stored separately, users would be able to view them together for scholarly analysis assuming that they had subscription access to the ACLS History E-Book Project. Richard Bailey hoped to overcome this barrier to access by having SPO publish the *entire* text—transcription and modernization—on its site, making it freely available yet interspersing the modernization with the transcription in this version as a concession to ACLS and the Press, which wanted to keep value in their subscription content by publishing the only “uninterrupted” modernization. ACLS rejected this plan, and the agreement for parallel, non-overlapping editions was reaffirmed.

SPO completed its version (according to the original plan) before the Press and ACLS did theirs, and, at the urging of Richard Bailey, was prepared to release it to the public with a note that the accompanying Press/ACLS version was forthcoming. At this point, ACLS decided that its involvement in the project was not benefiting anyone, so ACLS broke its contract with the Press, agreeing to compensate the Press for its expenses in editing the text. The Press in turn agreed to relinquish its copyright in the introduction and modernization to the editors and give the text to SPO, with all of the Press’s revisions, for publishing as a single online resource. SPO released all components of the scholarly edition at one time in November 2006.

Thoughts for future projects

The experiences of SPO and the project editors yield some valuable lessons in intellectual property for any digital scholarly edition. We are reminded that signing an academic book contract with a publisher usually means the authors or editors must turn

over their copyright, putting significant restrictions on their freedom to reuse their scholarship elsewhere. As for the failure of the plan of cooperation among SPO, ACLS, the Press, and the editors, it is difficult to determine a single cause of the breakdown in communication. SPO and the editors attempted to make their plans clear to ACLS and the Press on many occasions, but with all parties lacking a common language for discussing the structure of the edition and lacking a common business model, more thorough discussions were needed at an earlier stage to prevent so much lost time. All stakeholders in such a complicated project need to understand which components will be available to whom and agree on the degree of autonomy allowed by each party.

Nevertheless, the customizations made to DLXS allow SPO and other organizations using the DLXS suite to publish other scholarly editions with similar components. This project also shows how an electronic resource could be made available through a mixed model of free and restricted access or entirely free, either way using software designed for hosting generic digital collections.

The rapidly changing market for scholarly literature requires flexibility in publishing models, and cost pressures for university presses and libraries require scalable publishing solutions such as those offered by SPO. We would like to think that, despite the convoluted journey to publication, the final product is published online in a sustainable digital library environment and in a way that serves the needs of the editors and readers.