

Manuscript submitted for publication in *La médiation numérique : renouvellement et diversification des pratiques: Actes du colloque "Document numérique et société"*, Zagreb, 2013. [Ed. by] Joumana Boustany, Evelyne Broudoux, Ghislaine Chartron. Brussels : De Boeck, 2014. Published version may differ slightly.

Editors' Notes: An Example of Changed Mediation

Michael Buckland, Emeritus Professor, University of California, Berkeley.
buckland@ischool.berkeley.edu

Patrick Golden, Bibliographer, University of California, Berkeley. ptgolden@berkeley.edu

Barry Pateman, Senior Editor, University of California, Berkeley. hiawg@earthlink.net

and

Ryan B. Shaw, Assistant Professor, University of North Carolina, Chapel Hill. ryanshaw@unc.edu

Abstract

Scholarly annotated editions of historically significant texts constitute an important foundation for education and research. Preparing 'documentary editions' requires a sustained investment of highly specialized expertise. Current editorial procedures are still rooted in the pre-digital work practices and space constraints of the printed codex. A collaboration of documentary editing projects has demonstrated how current Web technology can facilitate scholarly editing and increase the return on investment by making the editors' research notes promptly and fully available through Web publication; gaining efficiency through collaborative, shared access to working notes among related projects; and providing interoperability with other scholarly infrastructure.

Keywords Scholarly editing, digital notes, Web technology.

1. Introduction

Our sense of identity as individuals and as a society is influenced by our understanding of our past, which is always incomplete and imperfect. An important resource, in this context, is the publication and explanation of historically significant texts: documentary editions. Large and famous examples of this genre include the *Monumenta Germaniae Historica* and the *Marx–Engels-Gesamtausgabe*, the historical-critical edition of the complete writings of Karl Marx and Friedrich Engels. Documentary editions are usually concerned with the writings, correspondence, and other texts relating to a significant individual. The annual survey in *Documentary Editing* listed 66 published volumes from 65 different editing projects in 2011. Nine volumes were from 'founding father' and/or presidential papers projects, with an 18th century emphasis. The remaining 56 volumes were from projects covering a wide range of individuals, groups (e.g. Cherokees, German Immigrants), and themes (e.g. Ballads, Mexican-American War, Vaudeville) with, mostly, a 19th and/or 20th century emphasis.

The scholars who prepare documentary editions not only select and present texts, they also explain their meaning and significance. Understanding depends on knowing the context: Who was this mentioned person? What is known about that event? What else was happening at this time and place? Why is the choice of words in the text interesting? – and so on. So adequate preparation of a documentary edition usually requires specialized expertise, a lot contextual research, and years of work.

2. The Problem

Scholarly annotated editions of historically significant texts constitute an important foundation for education and research. Documentary editing requires a sustained investment of highly specialized expertise, but long-term funding is difficult. The expert editors and their carefully trained assistants spend much time researching people, places, events, and institutions associated with their papers. Notes resulting from these time-consuming investigations are typically kept in folders in the project offices or hard drives and may result in a few lines of footnote in the eventual published volume. The reality is that most of what is learned (and all of what is learned in some cases) is not included in the published volumes. Instead of being shared with other researchers, records of the research are discarded when grants for publication expire.

This situation is the more poignant for two reasons: First, editorial work tends to be duplicative in the parallel editorial efforts of different projects with overlapping scope. For example, Emma Goldman (1869-1940, the famous anarchist) and Margaret Sanger (1879-1966, birth control and women's rights activist) knew each other and were active in some of the same circles, so the editors of the Goldman papers and the editors of the Sanger papers, located nearly 4,000 km apart, often research the same details, as do scholars working in other capacities: historians, archivists, and curators of special collections. Second, projects expire, but scholarship continues. The ideal would be if the editorial "workshop" could remain ready to support resumed scholarship as and when labor and funding allow.

Current work practice is rooted in influence of the pre-digital work practices and the space constraints of the print-on-paper codex. The great majority of documentary editing projects exist only in relation to the eventual published edition. When the last volume goes to press, the editorial staff becomes unemployed and their materials abandoned. Even if their notes are physically preserved, they are would be hard to locate and difficult to use.

Documentary editors work hard to provide explanations in commentary, footnotes, and appendixes. Financing for these projects is generally difficult, so both editors and funders have an interest in the benefits derived from their work, in the return on their investment. However, benefits are limited because, for several reasons, the full explanation is often not included in the published volume for lack of space. Editors working on related projects do duplicative research without knowing it. Their duplicative research is also not published for lack of space. Significant editorial insights remain inconclusive, incomplete, or too peripheral for inclusion in the published volumes.

We describe an initiative to increase the benefits of document editing through a change in perspective and as well as the use of Web technology. In the project *Editorial Practices and the Web* three major documentary editing projects undertook a fundamental change in how they work. The three collaborating editing projects have adapted their work practices to take advantage of Web technology as a way to share their working notes with each other and with the world.

3. A Solution

There is a simple technical solution. The editors could write their explanations and their working notes in full, then save them as html, hang them on any website, and then continue as before. Web search engines will index these web pages and so the benefits of editorial labor become available immediately in full, for everyone, for free, regardless of what is or is not included in the eventual published volume. Although this move seems simple in a technical sense, it constitutes a change in professional work practice, has extensive consequences, and is, therefore, complex.

Our approach is based on two significant moves: first, a change in perspective. Discussion of the preparation of documentary editions is ordinarily focused on the text being edited and on the eventual published edition. Our concern, however, is not with the text itself directly nor with the resultant editions, whether published in print or online, but, instead, with the *editing process*. Second, a small but very significant move to include Web technology early in that process provides earlier and more complete access to the results of the editors' efforts.

A project entitled 'Editorial Practices and the Web, is examining these issues (Buckland & Golden, 2012; Golden & Buckland, 2012). It is administered by the Electronic Cultural Atlas Initiative (ECAI) in the School of Information, University of California, Berkeley. ECAI works at the intersection of digital libraries and digital humanities and specializes in collaborative projects. See ecai.org

Three long-established documentary editing projects agreed to adapt their procedures and to upload their notes to a shared website:

- the Emma Goldman Papers Project (University of California, Berkeley) is preparing a four volume edition of texts relating to Emma Goldman (1869-1940), the charismatic Russian-born anarchist activist. <http://sunsite.berkeley.edu/goldman>
- the Margaret Sanger Papers Project (New York University) is preparing a four-volume edition of the papers of Margaret Sanger (1879-1966), America's best-known birth control activist. <http://www.nyu.edu/projects/sanger/>; and
- the Elizabeth Cady Stanton and Susan B. Anthony Papers Project (Rutgers University) was competing a six volume edition of the principal sources of women's political aspirations associated with E. C. Stanton (1815–1902) and S. B. Anthony (1820 – 1906). <http://ecssba.rutgers.edu/>

All three are mature projects already in progress for more than twenty years. The participants were chosen because of their overlapping interests in late nineteenth century and early twentieth political radicalism in the United States. The greater the overlap in interests and the greater the geographical separation, the greater the expected benefits from Web-base cooperation.

4. The Present Situation

The situation can be described by considering editing procedures in three stages.

Initial Research Notes. Ideas and notes are remembered and recorded in (often handwritten) notes. With the reducing costs of laptops, notebooks, scanners, and OCR software, the trend is away from paper pads towards scanned documents and keyed notes.

Editors' Working Notes. Notes, collected data, lists, references, clippings, photocopies, etc., are mostly stored in topical folders in filing cabinets, but there are also specialized locally-developed tools such as itineraries, chronologies, and legislative histories. Importantly, the *Editors' Working Notes* include notes on the many unresolved problems that researchers accumulate: reasons to question published accounts; why a claim might be suspect; known false leads; promising clues and lines of inquiry that might be followed up later; notes that someone else knows about some point; references to documents not yet located; citations known to be garbled; unresolved queries; and so on. One might hesitate to publish such working notes openly, but editors and researchers in other projects could find them very useful and others might already know the answer or, at least, be able to suggest where to look. This revives in a small way the nineteenth-century "Notes and Queries" genre of periodical.

There are also locally-made tools. Editors of personal papers usually need to create a detailed itinerary of that person's movements, which were complex in the case of Emma Goldman's lecture tours. Similarly, editors might create uniquely detailed legislative and legal histories of specific topics as the Stanton and Anthony project editors have.

Editors' Notes, appear, if at all, in very concise annotations, footnotes, endnotes, and appendixes in the eventual published volumes. In contrast, if explanatory notes were written at whatever length an editor considered justified and helpful, with sources clearly stated, and promptly posted on a website, they would be more informative, would be immediately available to everyone, would soon be indexed by search engines, and would easily be found by inquirers. Dated and signed, these notes can provide a steadily growing population of trustworthy research reports that all kinds of scholars and students can benefit from. In particular such notes would benefit scholars outside of research institutions and could facilitate a higher level of scholarship in popular compilations such as the Wikipedia.

What is needed is a sustained move from the upper row of Figure 1 to the lower row.

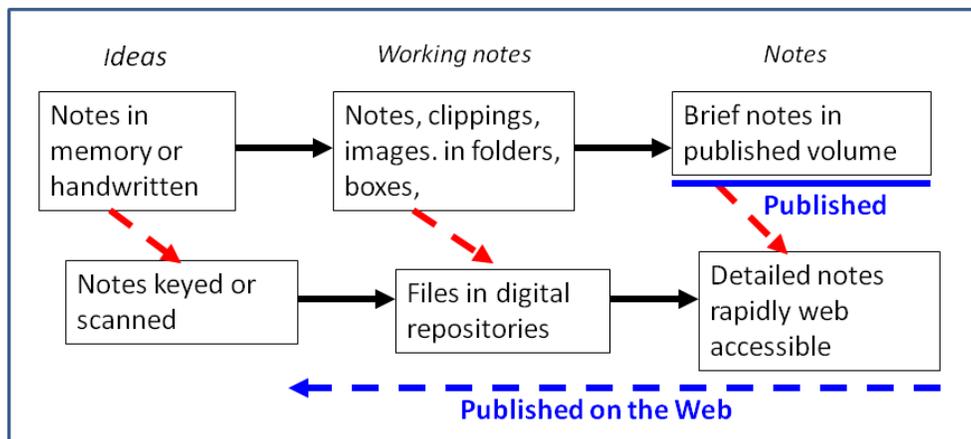


Figure 1. The transition to digital notes

A more systematic, more digital approach to the first and middle stages is mainly a matter of accelerating the existing shift from handwritten to keyed notes and adopting a more structured arrangement of material. Each stage feeds the next.

4. Architecture and Software

Benefits would result from editors' notes being saved as web-documents on any website. However, we created a shared Editors' Notes database and website hosted at Berkeley and designed to accommodate the varied forms of notes in which participating editors and their staffs enter their notes.

<http://editorsnotes.org>. We made many, many minor changes during the past two years to adjust the website's behaviors to suit the needs and preferences of the contributors.

The website was created by combining open source software: Django (the Python web framework), Postgres (using native support for XML fields), Xapian (for full-text search), South (for database migrations), Disqus (for discussion threads), Zoom.it (for high resolution scans), Zotero (for input and editing of bibliographic data), and Open Refine (formerly Google Refine), for duplicate detection. Django is an open source web application framework originally developed for the rapid production of news reports. Its primary goal is to ease the creation of complex, database-driven websites. Django follows the model-view-controller architectural pattern, emphasizes reusability and "plugability" of components,

rapid development, and the principle of DRY (Don't Repeat Yourself). Python is used throughout, even for settings, files, and data models. Django is opensource software now administered by the non-profit Django Software Foundation <<http://www.djangoproject.com/>>. The Xapian search engine is fast, flexible, well-documented, and fully open-source. It is also well-integrated with the Django web framework that we plan to use as a platform <<http://xapian.org>>.

The emphasis on open source software departs from the use of proprietary software, which can cause severe difficulties for projects requiring continuity over decades as products become unprofitable or obsolete and licenses and campus technical support ends.

The underlying data structure has three kinds of records.

- i. *Notes*. Notes consist of text written by an editor. They are stored as html so that they may have hyperlinks and all the other features that html enables. Each Note is categorized based on its completeness: notes are "Open" when they require more work; "hibernating" when a resolution remains desired but appears impractical or of low priority; and "closed" when deemed completed. Any Note can be revised at any time and all prior versions are retained and could be restored. The intended separate category of *Queries* is adequately handled by the "Open" and "Hibernating" categories.
- ii. *Documents*. Documents are records of source material that may be cited by an editor. We have used Zotero to manage Document metadata (e.g. item type, author, title, archive), enabling the input and output of Documents as structured bibliographic records. Documents may have attached Scans, Transcripts (with optional annotations), and hyperlinks to external websites.
- iii. *Topics*. All Notes and Documents are indexed using terms drawn from a controlled vocabulary of Topics which the interface uses to aggregate the Notes and Documents relevant to a specific person, organization, place, event, publication, or theme. Topics may be person names, organization names, place names, event names, publication names, or names of topics or themes. We can think of these as subject authority records, with support for variant spellings, aliases, etc., but they can go beyond that, with support for various kinds of relations among Topics, e.g. personal relations between persons, involvement of persons and organizations in events, and so on.

Part of the architecture of the site is shown in Figure 2.

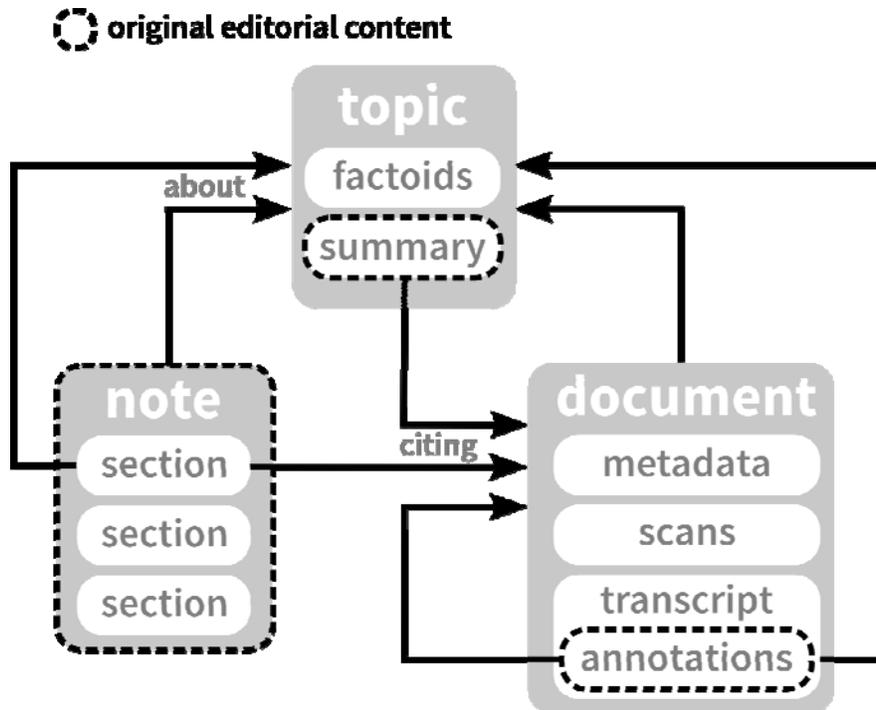


Figure 2. Part of the Editors' Notes data model. Notes, sections of Notes, and Topic summaries may cite Documents. Document annotations are linked to the Topics to which they relate. The Topic's "summary" is for free-form textual description of the Topic. "Factoids" are the pieces of structured data created locally and/or imported from a trusted resource. These could be the source for specialized search and visualization interfaces.

Permissions to view or edit different items are handled by Projects, which are made up of one or more users. Each Note, Document, and Topic is associated with all users who have edited it, as well as the Projects those users belong to. Projects may choose to restrict public access to certain items that they "own" as they see fit. Full records of changes to each item are stored, making it possible to view or revert to earlier versions. In August 2012 password control to view the site was quietly removed making the site openly available to both humans and web crawlers. By September, after Web search engines, including Google, Bing, and Baidu (China), had indexed the contents, the resources on the Editors' Notes site were being viewed by scholars from around the world.

A sample editors' note is shown as Figure 3.

Editors' Notes
Browse
About

Search

Sanger and the Third ICPP

Unlike the coverage in secondary sources, planning on the Third ICPP started far earlier than See Watumull and the Third ICPP and British and Third ICPP for more. Focus in these notes

Add section: Document citation

 Sanger, Margaret, letter to Katharine Dexter McCormick, Jan. 8, 1951 [MSM S33:712]

 Sanger, Margaret letter to Pearl S. Buck, Jan. 9, 1951 [MSM S33:717]

B
I
P
H1
H2
H3
☰
☷
↶

Notes that if they had been able to keep the East-West going, it would have made a great and blundering, not only in Washington but wherever but leaders are, and of course most fr civic affairs given to military leaders. We seem to learn nothing from history and one would make our military leaders hesitant to build up a powerful military organization. Germany dic on their knees."

"There is a possibility of our having a population Conference in the Orient somewhere in 19 from India recently tells me many of the 65 clinics that I organized in 1936 are still going. I the health groups to that information as well as supplies will not have to be limited to peop ahead and I am so happy to feel well again."

 Sanger, Margaret letter to Eleanor C. Jenkins, Jan. 9, 1951 [MSM S33:719]

 Brush, Dorothy letter to Margaret Sanger, Jan. 21, 1951 [MSM S33:768]

Figure 3. Sample editor's working note.

5. Future Developments

Progress has been significant because routine day-to-day procedures have moved from the desktop to a Web environment. Additional developments are desirable.

5.1. Incorporating Digital Humanities Technology

Editors and their staff continue to work primarily with simple flat text files and scanned images. There is a chasm between the daily routines of ordinary scholars and the impressive technical achievements of experts in the large-scale, complex projects reported at Digital Humanities conferences with dazzling visualizations created from complex databases by experts using sophisticated software. How might the latter be harnessed for use by the former, who have so little capacity for absorbing additional workload or complexity? There are tools for using name authorities or generating map displays, timelines, prosopographies, and the like. The challenge is software integration and interface design with very low thresholds of user effort by non-specialist researchers. Editors may maintain name authority files (e.g. Hajo, 1991). One priority is making links with Geonames, VIAF, Wikidata, and other resources and enabling editors to download records and also to contribute additional place names, people, and organizations.

Similarly, efficient tools are needed for importing records to enrich local data. These tools would have an added benefit of removing some tedious, duplicative work from everyday research. Editors would be able to import contextual details of, for example, persons (e.g. birth and death dates, place of birth, other names) or of places (alternative names, containing jurisdiction, latitude and longitude) without researching or transcribing these details at every mention. Using a link can bring the benefit of automatic updating as additions and corrections are made to the resources to which they are linked (Shaw & Buckland, 2011). Simple interfaces are also needed to allow users to invoke visualizations (maps, timelines, and network graphs) based on targeted local data.

5.2. Preservation and Access

Our project has drawn attention to the loss of resources when editorial projects end. The funding for documentary editions is narrowly limited to support for the eventual published edition. When the manuscript of the final volume is ready for publication, the editors and staff depart and their working notes become effectively inaccessible if not discarded. Grants do not (yet) fund the preservation of the editors' working papers. Elite projects generally have strong continuing institutional support but they are not typical and nobody seems to know about the rest. Only the published printed volumes remain available in most cases.

Projects end, but scholarship continues! Could the legacy of working notes of completed projects cost-effectively support future scholarship instead of being discarded? To the extent that editors' working notes have been handled digitally along the lines of our project, they could remain available and accessible with minimal overhead. Thinking tactically, we could examine what low-cost procedures could keep these editorial resources accessible as a more-or-less arranged and preservable archival resource. Thinking strategically suggests that the relationship between the editorial working notes and the published editions should be reconsidered. Currently, the published editions are the one and only product. The editorial expertise and project working resources are treated as expendable means to that sole objective. But the change in mediating technology makes it imaginable to reverse that relationship. In this view the editorial "workshop" (expertise and working notes) could be enduring assets and the published editions would become intermittent valued by-products. Scholarly communication could be greatly extended if it were feasible not only for scholars anywhere to have *sustained access* to the working notes, but also for scholars anywhere to *add supplementary notes, corrections and additions to them* (with clearly separate attribution) in the future *as and when interest, ability, and resources allow*.

The ambition would be to move beyond a short term tactical solution (graceful retirement into a passive archival collection) toward a working collection in which at least the finding aids and research guides could be updated and enriched as scholarship continues, a new genre somewhere in between a conventional (static) archive, a library special collection, and an ongoing research program. This is a logical consequence of the rationale for our project and follows from the move to digital technology and a networked environment. There appears to be little precedent for this, except in local community archives and open note-book science.

5.3. Horizontal Interoperability

In addition, the Labadie Collection of documents on radical history and social protest movements at the University of Michigan Library has also contributed numerous notes created by curators concerning their often obscure materials. These notes resemble documentary editors' notes in form, function, and utility and can create a bridge between the separate worlds of librarians and documentary editors. The rationale for including the Labadie collection was not simply the usefulness of making detailed, expert curators' notes openly available but that this course of action opens up a possible renaissance in the active curation of library special collections. A similar experiment with the finding aids and research guides created by archivists is planned.

6. Summary and Outcomes

Scholarly annotated editions of historically significant texts constitute an important foundation for learning and research in the Humanities. Scholarly editing, however, requires a sustained investment of highly specialized expertise and long-term funding is difficult. Given the right software, minor changes in work practices can make the painstaking editorial research much more organized, convenient and rapid, widely accessible, and permanent, thereby increasing utilization, efficiency, and the return on investment. The move to web-accessible notes was a pivotal change. Future tasks are to bring the functionality of advanced Digital Humanities projects to the editor's workbench and to ensure that resources so carefully assembled by editors remain capable of supporting future projects.

In a field dominated by the limitations of the print-on-paper codex, a relatively small technical change in mediation has had extensive consequences: the contextual research of the editors now becomes immediately available to everyone; duplicative research is reduced; every online visitor becomes a potential helper; the preservation of notes becomes feasible; editors' notes can be linked horizontally with librarians' notes, archivists finding aids, biographical dictionaries, etc.; linked data mark-up could provide to name authority services and support map displays and other forms of visual analysis.

This experience is a case study in changed mediation, increased return on investment, and the future of cultural scholarship, including a reversal of the relationship between scholarship and publication.

Acknowledgments

We are grateful for the financial from the A. W. Mellon Foundation and the Coleman Fung Foundation and the indispensable collaboration with Esther Katz and Cathy Moran Hajo (New York University), Ann Gordon (Rutgers University), Julie Herrada (Labadie Library, University of Michigan) and their assistants.

References

The project webpage is at <http://ecai.org/mellon2010/>. The database and website are available at <http://editorsnotes.org>. Note the contact form at <http://editorsnotes.org/about>.

BUCKLAND (M.) and GOLDEN (P.) *Editors' notes: a case-study of collaboration and convergence in the humanities*. Taiwan E-learning and Digital Archives Program ("TELDAP") conference, Taipei, Feb. 24, 2012. Viewed 15 Jan 2013. <http://people.ischool.berkeley.edu/~buckland/teldapplenary12.ppt>

GOLDEN (P.) and BUCKLAND (M.) *Using the Web to change editorial research practice*. Pacific Neighborhood Consortium & Electronic Cultural Atlas Initiative Joint Meeting, Berkeley, Dec. 7, 2012. Visited 15 Jan 2013. <http://metadata.berkeley.edu/ednotespnc12.pdf>

HAJO (C. M.), *Computerizing control over authority names at the Margaret Sanger Papers*, in *Documentary editing*, XIII, no. 2, 1991, pp. 35-39.

SHAW (R.), GOLDEN (P.) and BUCKLAND (M.), *Integrating collaborative bibliography and research*. Text for poster, American Society for Information Science and Technology Annual Meeting, Baltimore, 29 Oct. 2012. Viewed 15 Jan. 2013. <http://metadata.berkeley.edu/ShawGoldenBuckland2012.pdf>

SHAW (R.) and BUCKLAND (M.). *Editorial control over linked data*. American Society for Information Science and Technology Annual Meeting, New Orleans, Oct. 10, 2011. Viewed 15 Jan. 2013. <http://metadata.sims.berkeley.edu/posterasist2011.pdf>