PROPOSAL NARRATIVE  [CONDENSED]

The Problem Addressed

The expert editors of scholarly editions and their carefully trained assistants spend a good deal of time researching people connected with their papers in one way or another. Notes on these time-consuming investigations are typically kept in folders in the project offices and may result in a few lines of footnote in the eventual printed volume. The reality is that most of what is learned about these individuals (and all of what is learned in some cases) is neither included the eventual printed volumes nor shared with other researchers. This editorial work could be better utilized.

Editorial work tends to be duplicated in the parallel editorial efforts of different projects with overlapping scope. For example, Emma Goldman and Margaret Sanger 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 over 2,000 miles away from each other, often research the same individuals, unwittingly. We estimate some 20% overlap in persons researched. One could imagine some saved effort if the two editorial teams shared the same suite of offices but that won’t happen for institutional reasons and, even if it did, they both overlap with editorial work at the Samuel Gompers papers in Maryland, the Eleanor Roosevelt papers project in Washington, DC, the Bertrand Russell papers in Hamilton, Ontario, the Wellcome Library in England, and on and on. It is remarkable how these contemporary leaders knew each and moved in shared networks.

So far we have mentioned research about people, but, of course, the problem is much larger because the editorial efforts on places, institutions, events, and other topics are also comparably duplicative.

Current editing work practice is rooted in the lingering impact of pre-digital work practices and the space constraints of the printed codex. The situation could be transformed by modest changes in procedures, which fall into three stages:

Initial Research Notes: Ideas and notes are remembered and recorded in (often handwritten) notes;

Editors’ Working Files: 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. Editors can and do ask editors elsewhere for help on specific topics, but answering may be time-consuming and they do not want to burden over-worked colleagues with repeated requests for help; and

Editors’ Notes, which appear, if at all, in very concise footnotes, endnotes, and appendixes in the eventual printed volumes.

Let us consider these three stages working backwards:

Editors’ Notes: If explanatory footnotes were written at whatever length an editor considered justified and helpful, with sources clearly stated, and promptly posted on the project’s website, they would be more informative, would be immediately available to everyone, would soon be indexed by Google and other search engines, and would readily 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 could
facilitate a higher level of scholarship in popular compilations such as the Wikipedia. To see what these could look like follow the link to two draft Editors’ Notes in the middle of http://sunsite.berkeley.edu/Goldman/index-notes.html These Editors’ Notes are attached in Appendix D.

Editors’ Working Notes take two forms: folders of topical notes, and specialized local working tools. The contents of the folders in the filing cabinets contain more detailed scholarship than the Editors’ Notes. If done tidily, these notes could provide a clear record of the progress of research (and the directions that research has taken) and move it out of the filing cabinet and out of the editor’s memory into plain sight. In digital form, Editors’ Working Notes could be shared with other editors in a variety of different ways (e.g. password controlled federated search; or depositing back-up copies in a shared, searchable repository).

Importantly, the Editors’ Working Notes (more than the Editors’ Notes) could include notes on the variety of unresolved problems that researchers accumulate: reasons to question accepted assumptions; 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; reasons to question published accounts; and so on. One might hesitate to publish such working notes openly, but some editors and researchers in other projects could find them very useful and others might already know the answer or, at least, how to resolve some of them. This revives in a small way the nineteenth-century “Notes and Queries” genre.

The second category comprises locally-made tools. For example, 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. We took a WordPerfect text listing created by the Emma Goldman Papers Project of Goldman’s travels and massaged it to create the searchable database and map interface at http://gray.ischool.berkeley.edu/emma/. (See the screen shot in Appendix E.) Similarly, editors might create uniquely detailed legislative and legal histories of specific topics as the Stanton and Anthony project editors have.

Initial Research Notes: With the reducing costs of laptops, notebooks, scanners, and OCR software, the trend is away from paper pads towards keyed notes and scanned documents.

What is needed is a move from the first row below to the lower row.

<table>
<thead>
<tr>
<th>Notes in editor’s memory, hand-written notes.</th>
<th>Notes, photocopies, clippings, etc., in folders in filing cabinets.</th>
<th>Truncated footnote (maybe) in eventual printed volume.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial notes keyed or scanned.</td>
<td>Files in digital repository open to authorized users.</td>
<td>Editor’s note rapidly published on Web.</td>
</tr>
</tbody>
</table>

The Emma Goldman Papers Project is experimenting with the design of web-published Editors’ Notes. The importance of this small change in work practices is modest in the context of a single editorial project. The real impact of a new “Editor’s note on the Web” genre would come
from its adoption as a standard best practice for editorial projects generally, internationally. Indeed, with time, the compounding cumulation of Editors’ Notes could have a substantial impact far beyond the editorial tasks that engendered them and could bring an increased visibility and recognition that the projects need.

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.

Three points are worth noting. First, high tech skills are hardly needed. The Emma Goldman Editor’s Notes are in html. Much of the benefit would accrue from merely writing a note using any word processor, then saving and posting the text as a plain html file. Second, this is primarily an acceleration in the evolution of existing work practices, which, if found worthwhile, would be sustained. Third, the cumulative effect over time of widespread adoption could be very substantial.

Placing relatively simple text Editor’s Notes on open websites and more complex Editors’ Working Notes in sites with restricted access would have a transformative effect on access to the fruits of editorial research, but would be a significant change in procedure for most editorial projects. The Central Team will work with each project individually (and with its IT support) to facilitate the process.

At the same time such technically-simple documents provide a starting point for the application of the more innovative techniques now being deployed in the digital humanities. For the purposes of this project, it is proposed that the Central Team do some proof-of-concept demonstration on the materials made available.

The Bigger Picture

Access to the results of scholarly editing projects has been constrained in three ways, two of which are noted above: First, their findings are incompletely accessible to the public and web-accessible Editors’ Notes are good remedy; second, work-in-progress is inaccessible to other editors and scholars, for which current technology can support limited access to Editors’ Working Notes.

The third current weakness is lateral: Other memory services -- archives, libraries, museums, mass digitization repositories, compilers and publishers of reference resources – although sometimes institutionally related, tend to be somewhat “silico-ized.” They have their own work practices and their own emergent standards. The Electronic Archival Description (EAD), which originated at Berkeley, now provides an internationally accepted standard for archival finding aids and is being extended with the Electronic Archival Context (EAC). Libraries have their own evolving standards (AACR2, MARC, Z39.50) and practices. In particular, librarians have a long-established tradition of creating guides for users, including detailed general guides to subject areas and shorter guides to specific topics, more or less tailored to local users and local resources, commonly called “pathfinders.” The expertise of a special collection curator provides a basis and an opportunity to go well beyond conventional cataloging and conventional pathfinders to provide guidance more comparable to that of an Editor’s Note. Please see the example on 25th page of the Appendix. However, user-related research and development has tended to emphasize the use of telecommunications tools to provide 24/7 access to a librarian rather than facilitating users’ self-help.

The Labadie Collection, part of the University of Michigan Libraries, is a notable special collection of materials relating to anarchism and radical movements, especially in the USA. There are four reasons to include the development of Curator’s Notes at the Labadie Collection as part of this project.
The first two reasons are pragmatic: First, the contents of the Labadie Collection are an exceptional resource for the editors of the other projects and any mechanism that facilitates access to the Labadie’s resources will materially help anyone engaged in research on anarchist and radical themes. Second, web-accessible Curators’ Notes made at Labadie will be enriched by the incorporation of the substantial file of hand-written and typed notes of the first Curator, Agnes Inglis.

The third reason is developmental: Archivists developed finding aids and then the Electronic Archival Description to shape best practices and standardized mark-up. Librarians developed how-to-find-out guides both book length and concise specialized guides commonly known as Pathfinders. (See, for example, the State Library of Iowa Pathfinder Project http://www.thepathfinderproject.org/ and the Internet Public Library collection of pathfinders at http://www.ipl.org/div/pf/ ). The typical standard template illustrated by the Pathfinder Project site works well for introductory purposes, but we believe that for special collections a new, more discursive style akin to the Editors’ Notes shown in the Appendix is indicated and we seek to explore this possibility.

The fourth and most powerful reason transcends the scope of the present project. If one really wanted the hidden treasures of special collections made fully available to scholars everywhere, consider how it might be done. A nineteenth century solution would be a multi-volume reprint of everything in the collection, like Migne’s Patrologia (387 vols) or the Monumenta Germaniae Historica (>300 vols), but the price is huge and an unedited reprint without scholarly explanations would be opportunity missed. A twentieth century solution would be a microfilm edition, but it would be difficult to have editorial annotation ready at the time of microfilming. But now, in the twenty-first century, not only is mass scanning and digitization the “reprinting” solution of choice, but ongoing curatorial annotation can be added as energies and opportunities permit and built up in a dialogue between a knowledgeable Curator and expert readers. Seen in this light, the inclusion of the Labadie Collection in this project paves the way for a renaissance in the curation of scholarly special collections.

These developments in archives, libraries, and other areas have strengthened each area. The time is now ripe for more attention to developing interoperability between these areas in ways that go beyond the happenstance that Google might index any of them. In the longer term, we want Editors’ Notes to “talk to” archival finding aids, library pathfinders, and the like, in the sense that they would be enriched with mark-up that could link them with each other either directly or, increasingly, indirectly through links to the controlled vocabularies in the namespaces that are emerging in the semantic web environment. In addition, provided that these different resources are not incompatible, digital humanities tools can increasingly harvest across different genres for detailed analysis and discovery.

In brief, three kinds of access to Editorial work need improvement:

1. Direct access by the public to editorial research findings;
2. Shared access between editorial projects to enhance collaboration; and
3. More interoperability between editorial projects and other kinds of scholarly enterprise.

A Strategy

This situation invites the following strategy:

1. Start very carefully, with a small group of closely-related editing projects with highly overlapping interests and editors who already regard each other as friends and colleagues and who have a record of helping each other;
2. As experience, confidence, and consensus develop, the circle of collaborating editorial projects would be gradually expanded to include increasingly diverse editing projects: other periods, other regions, other emphases (e.g. literary papers), and projects in other countries. As examples, the Marx-Engels papers project (MEGA), mainly in Berlin, would overlap nicely with the collaboration proposed here and, in the U.K., the same tools could be highly beneficial if adopted by the compilers of the monumental *Victoria County Histories* and the *Survey of English Place Names*.


   This strategy would have three additional benefits:

1. If editors knew, in any detail, what research had been done at other projects, then they would know reliably when they could ask for help or themselves make use of other projects’ resources.

2. Institutional and funding patterns favor personality-based editorial projects, typically based of the papers of a single important individual. Personal papers and biography provide an attractive and practical introduction to history. Editing requires examination of contexts and so the more that the contextual notes can shared, the better the personality-based projects can also support the history of groups and themes in social history.

3. Individual editing projects eventually end. What remains then? There may be little more than the printed volumes. Even if the working files are preserved they will be relatively inaccessible after the editors have moved on.

   Based on this wider view, we propose a modest start working with a Central Team, three closely related editing projects, and one lateral initiative.

**PROJECT DESCRIPTION**

The long-term objective is a pivotal shift in editorial work practice internationally, but little impact can be expected without visibly successful adoption at the initial sites. Each site is different and we cannot fully predict what the issues, obstacles, and concerns will be. It seems prudent to start slowly and very carefully, on a small scale, with very highly motivated partners.

The Central Team will provide overall project management and start by assuming full responsibility for everything except the initial creation of plain text by the four collaborating projects. Then, as mutually acceptable and interoperable styles and formats emerge, templates and other tools will gradually be migrated to the collaborating projects’ sites or wherever else these institutions wish them to be. The life-cycle of the notes is seen as creation, initial publication, and back-up (depository storage).

The project comprises the following tasks:

#1. Project website at Berkeley.

2. Make the production of *Editor’s Notes* routine standard practice at the Goldman, Sanger and Stanton-Anthony projects. Circumstances and preferences differ at each collaborating site, so this task will require careful separate consultation and is expected to be an iterative, evolutionary process.
#3. Introduce Editors’ Working Notes as standard procedure at multiple editing projects. Again, since circumstances and preferences differ, careful separate consultation will be needed.

#4. Evolve, collaboratively, mutually acceptable format and markup guidelines.

#5. A pilot study of Curators’ Notes similar to the Editors’ Notes and specialized library pathfinders.

#6. Demonstrate to project staff and interested local scholars at each site what advanced humanities computing techniques could do with these notes.

#7. Estimation of overlap in editorial effort.

#8. Migration of IT support for the generation of Notes and Working Notes and of the publication and repository sites to the locations preferred in each case, probably but not necessarily locally.


#10. Documentation, dissemination, and promotion of wider adoption of what proves beneficial.


**Tentative Architecture and Software**

The plan is that the Central Team will establish an Editors’ & Curators’ Notes database and website hosted at Berkeley and designed to generate the varied forms of notes. An underlying data structure of four kinds of records is foreseen:

1. **Notes.** Notes consist of text written by an Editor or Curator. The text is stored as html, so it may have hyperlinks and all the other features that html enables.

2. **Queries.** Queries also consist of text written by an Editor or Curator, a special kind of Note identified as asking something rather than explaining or describing something. The distinction exists solely so that Queries can be presented differently in the interface. A Query may be converted to a Note if it is answered.

3. **Citation.** Citations are bibliographic descriptions of (and references to) source documents. These may be references to offline documents or links to online documents, and may include a scan of the document itself.

4. **Terms.** All of the previous three kinds of records must be indexed using terms drawn from a controlled vocabulary. Terms 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 go beyond that, with support for various kinds of relations among terms, e.g. personal relations between persons, involvement of persons and organizations in events, and so on. Just as Terms are used to index Notes, Queries, and Citations, relationships between Terms might also be used for indexing and discovery. For instance if John Doe and Fred Smith were both schoolmates and business partners, their business partnership relation might be used to index a Citation that is relevant to the latter relationship but not the former.

Notes, Queries, and Citations are associated with their creators (of which there may be more than one in collaborative documents). Records of every change to these records are kept so that it is always possible to return to earlier versions. Notes, Queries, and Citations may be made public or kept private. Private records are only visible to editors and curators, not the general public.
The interface will use the indexing terms to aggregate all of the Notes, Queries, and Citations relevant to a specific person, organization, place, event, publication, or theme. The resulting aggregation is expected to look something like the E.B. Foote and L.A. Times notes in Appendixes D. Snapshots of the aggregated term pages, as well as materials like the scanned documents, can be periodically stored at the Internet Archive or elsewhere.

This is not a software development project, but an experiment in changing work practices in keeping with the general move from a print on paper to a digital environment. Numerous more or less suitable software tools are available and more will doubtless emerge during the next two years. The goal of widespread adoption dictates the use of software that is already widely used, well-supported, economical, and favored by local IT support services. We have identified the Django web framework (http://www.djangoproject.com/) and the various components available as the software of choice and the Xapian search engine for full-text search. These choices could change if more satisfactory options emerge.

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 pluggability 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>.

Where collaborating institutions prefer other tools, we will migrate them from Django to their local software of choice before the end of the project.

We seek to avoid local software development unless we really need it. Any new software would be made freely available as open source at no cost.

PARTICIPATING INSTITUTIONS

The Electronic Cultural Atlas Initiative, University of California, Berkeley

The Electronic Cultural Atlas Initiative (ECAI) is exceptionally positioned to undertake this project. ECAI was established in 1997 as a unit reporting to the Dean of International and Area Studies [Eff. April 1, 2010 reporting to the Dean of the School of Information] with a mission to transform teaching and research in the humanities and social Sciences through improved attention to place and time in an increasingly digital environment.

ECAI works at the intersection of digital libraries and digital humanities and specializes in collaborative projects. The Institute of Museum and Library Services, the Luce Foundation, the National Endowment for the Humanities, and the National Science Foundation have funded substantial ECAI projects in recent years. The present proposal arises directly from the participation of the Emma Goldman Papers Project in two projects: “Bringing Lives to Light: Biography in Context” (an IMLS National Leadership project) and “Context and Relationships” funded by the joint NEH and IMLS Advancing Knowledge program.

ECAI is jointly led by two Co-Directors: Professor Lewis R. Lancaster, Emeritus Professor of East Asian Languages and Cultures, who pioneered the digitization of ancient Asian religious texts; and Michael K. Buckland, Emeritus Professor (and former Dean) in the School of Information, who has exceptional experience of innovation in library services. Major current collaborations are with Academia Sinica (Taiwan); the École française d’Extrême-Orient (Paris); Fudan University (Shanghai), and The Queen’s University Belfast (Northern Ireland).
ECAI collaborates with numerous individual scholars around the world with a rich program of conferences, workshops, e-publications, advancement of best practices, and proof-of-concept projects. On campus ECAI works closely with the U.C. Berkeley School of Information. (For details see ecai.org)

The University of California is a leading public university system with a Land-Grant mission dedicated to instruction, research, and service. The Berkeley campus is the original and best known of the ten campuses.

**Collaborating Institutions**

The four collaborating programs are:

- *The Emma Goldman Papers Project*, University of California, Berkeley. [http://sunsite.berkeley.edu/goldman](http://sunsite.berkeley.edu/goldman)
- The Elizabeth Cady Stanton and Susan B. Anthony Papers Project, Rutgers, The State University. [http://ecssba.rutgers.edu/](http://ecssba.rutgers.edu/) and
- The Labadie Collection, University of Michigan Libraries. [http://www.lib.umich.edu/labadie-collection](http://www.lib.umich.edu/labadie-collection)

These four programs are a good initial selection because their interests overlap closely, the staffs are well-known to each other, and all four are eager to participate. The Labadie Collection of materials on radical history is an apt choice for this project: It is a resource for the other three projects; the staff performs a quasi-editing role through the preparation of finding aids that correspond loosely to the notes of the editing projects; and this project can capitalize on an extensive set of hand-written or typed notes created through 1952 by Agnes Inglis, the first curator.

**PROJECT ORGANIZATION**

The project is based at the University of California, Berkeley, and structured as a collaboration of five units: a Central Team, based in the Electronic Cultural Atlas Initiative, and four collaborating programs: One at U.C. Berkeley (Emma Goldman Papers Project) and three at other institutions: University of Michigan (Labadie Collection in the University Library); New York University (Margaret Sanger Papers, in the Department of History); and Rutgers University (Stanton & Anthony Papers Project).

The Central Team at UC Berkeley (M. Buckland, B Pateman, and Research Assistants) is responsible for overall project management; support and guidance for each collaborating editing project; and demonstration of future possibilities.

The three Collaborating Editing Projects commit to the routine generation of *Editor’s Notes*; to *Editors’ Working Notes* with limited access by other editors and scholars; and to cooperation in documentation, demonstration, and evaluation.

The Labadie Collection commits to the creation of *Curators’ Notes* designed to be analogous to (and interoperable with) *Editors’ Notes* and specialized library pathfinders.

**Long-term Sustainability**

The objective of this project is to develop small but significant changes in work practices. Success is defined by their incorporation into existing programs. If that is achieved, long-term
sustainability is achieved until such time as they are, in turn, displaced by subsequent improvements.

**SUMMARY AND CONCLUSION**

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. Minor changes on work practices could make the painstaking editorial research much more widely accessible, more rapidly and more permanently, thereby increasing utilization, efficiency, and the return on investment. We see the move to web-accessible *Editors’ Notes* as a pivotal change, with much more to follow.