

Integrating Collaborative Bibliography and Research

Ryan Shaw

Univ. of North Carolina at Chapel Hill
100 Manning Hall
Chapel Hill, NC 27599
ryanshaw@unc.edu

Patrick Golden

The Emma Goldman Papers
2241 Channing Way
Berkeley, CA 94720
ptgolden@berkeley.edu

Michael Buckland

Univ. of California, Berkeley
102 South Hall
Berkeley, CA 94720
buckland@ischool.berkeley.edu

ABSTRACT

We report on the design and implementation of an innovative shared work-environment “Editors’ Notes,” especially the use of an open-source bibliographical reference management platform (Zotero) in conjunction with a continuously updated corpus of the working notes of three leading documentary editing projects and the curatorial notes of a library special collection. The benefits, constraints, and affordances of using Zotero in a one-way relationship with the text corpus are described and additional possibilities are noted.

Keywords

Collaboration, bibliography, digital humanities, personal information management.

INTRODUCTION

Compiling accurate bibliographical references is an essential part of research and is part of a broader process of personal information management by researchers (Jones & Teevan, 2007). Researchers’ personal bibliographies usually remain personal, with selected items appearing at the ends of publications but, for various reasons, one rarely knows the larger bibliographies from which they were drawn. Researchers are rewarded more for original research than bibliographies.

The effort of publishing bibliographies may be reduced by using reference management software to facilitate the acquisition, organization, and use of bibliographic descriptions. Reference management software and services have been designed for collaborative use, but only recently have widely used platforms for collaborative reference management emerged. (See Marino, 2012 for a comprehensive review of this literature). There are extensive literatures on personal information management and collaborative research, yet relatively few studies of collaborative bibliography as an intersection of the two

areas (Fourie, 2012). Here we present a brief examination of the issues related to our integration of Zotero, a popular platform for personal and collaborative reference management, with Editors’ Notes, a system for managing a collaborative editorial research process and its products.

EDITORS’ NOTES

Editors’ Notes is an open-source hosted service, funded by the Mellon Foundation, for organizing the research work of documentary editors (Shaw & Buckland, 2011). Documentary editors prepare “editions” of documents such as letters, diaries, and essays that have value as evidence for political, intellectual, or social history (Kline & Perdue, 2008). They contextualize these documents with extensive footnotes based on their original archival research. Yet the published footnotes represent just a small fraction of this research. The majority is contained in internal notes made by the editors and their assistants as they develop answers to the questions raised by their documents. It is the rich scholarly content of these internal notes that Editors’ Notes aims to make accessible to and interoperable among various editing projects and the public.

While editors’ research notes may take various forms, typically they are created during the course of researching a particular question. For example, consider the following scenario. A letter from Margaret Sanger to one of the officers of the International Planned Parenthood Federation (IPPF) raises some questions in an editor’s mind regarding the structure of that organization. The editor decides that some research is needed to better understand and explain the organizational structure of the IPPF, and a note is created to track this research. The note is structured as an annotated bibliography, with an entry for each resource found to contain useful information on the organizational structure of the IPPF, including both a bibliographic description of the resource and a summary of the information found. In addition to this annotated bibliography, there may be a summary distilling all the relevant information found in the various sources. This distillation forms the basis for the eventual published footnote.

The editing projects involved in Editors’ Notes have used a variety of reference management systems throughout their decades of existence. These systems include various combinations of physical filing cabinets, library-cataloging software, custom applications built on relational databases,

This is the space reserved for copyright notices.

ASIST 2012, October 28-31, 2012, Baltimore, MD, USA.
Copyright notice continues right here.

and specialized reference management software such as EndNote. To varying degrees, these systems have allowed the projects to organize and manage descriptions of the sources they consult.

Yet the organization and management of the research notes based on these bibliographies has been more haphazard. While reference management software usually provides some means of adding notes to individual entries, this functionality is insufficient to capture the cross-referencing that ties together the threads running through consulted sources. Thus, research notes have tended to live separate lives from bibliographic descriptions. As a result, notes on a particular source or topic may spread across handwritten notes, annotated photocopies, or word processing files, and duplication of content is rampant.

Editors' Notes seeks to improve this situation by integrating reference management into a system for creating, organizing, and maintaining research notes. Rather than creating yet another reference management system, we have integrated Editors' Notes with the Zotero reference management software. This allows us take advantage of the various tools Zotero provides for importing and exporting bibliographic data and focus our limited resources on representing and managing the complex links among research questions, resources, and research notes. By integrating bibliographies describing resources with the explanatory notes made about particular resources' significances, those bibliographies can be enriched with additional layers of information regarding the relevance and quality of the resources described therein. In the following sections, we examine this integration first from a technical perspective and then from a conceptual one.

INTEGRATION FROM A TECHNICAL PERSPECTIVE

Zotero is an open-source reference management platform developed by the Center for History and New Media at George Mason University. The platform consists of a client that runs as a browser extension or a standalone program, and a server for collaboratively sharing and maintaining bibliographies. Together, these allow researchers to build bibliographies made up of items either added manually or pulled from various sources, including library catalogs, journal databases, newspaper websites, or other reference management software. These bibliographies can be maintained purely locally, or shared via the Zotero server.

By relying on the Zotero platform Editors' Notes can avoid "reinventing the wheel" of reference management, while benefiting from interoperability with the other tools that use the platform. Two features the Zotero platform provides are particular importance: its server API and its standard data model for bibliographic data.

Bibliographic Data Server API

Bibliographies that are shared via the Zotero server can subsequently be read from and written to using Zotero's server API. From its inception, the Zotero project has

sought to encourage its integration with other software and services by providing APIs (Cohen, 2008). The server API makes it possible for users to build a bibliography using a Zotero client and then share that bibliography with Editors' Notes. Since Zotero clients can import bibliographic descriptions from other reference management software, as well as using the Zotero "translators" to add bibliographic descriptions directly from various online sources, this considerably reduces the labor required to build bibliographies within Editors' Notes.

Standard Bibliographic Data Model

While Editors' Notes uses the Zotero API to read data from Zotero bibliographies, for reasons of performance and reliability it does not actually rely on Zotero to store its bibliographic data. Instead, all bibliographic descriptions read from Zotero are redundantly stored within the Editors' Notes server. As a result, bibliographies can also be created and modified within Editors' Notes without using Zotero.

Even when the Zotero client and server API are not used, however, Editors' Notes maintains compatibility with the Zotero platform by adopting its data model. All bibliographic data is stored in the JSON (JavaScript Object Notation) format used by the Zotero API. As a result, we can take advantage of tools that use this format. For example, the Zotero platform includes processors for rendering its bibliographic data in a variety of citation styles. Since Editors' Notes uses the same data format, we can use these processors as well without modification.

Currently the integration of Editors' Notes with Zotero is only one-way: bibliographies can be pulled from the Zotero server, but modifications or additions to those bibliographies made within Editors' Notes cannot be pushed back. However, since we have maintained compatibility at the schema level, this could be enabled in the future without too much work.

INTEGRATION FROM A CONCEPTUAL PERSPECTIVE

A standardized data model and server API make integration with a collaborative reference management platform possible from a technical perspective. But data models, APIs, and the specifics of technical integration change. A conceptual perspective can move beyond these specific details to consider less transitory issues of integration. Here we draw upon the conceptual model of bibliography developed by Bates (1976) and extended by Hendry, Jenkins, and McCarthy (2006) to characterize the "division of labor" between Zotero and Editors' Notes.

Presentation Constraints

Integrating with a platform for reference management means giving up control over the specification of many constraints related to the presentation of bibliographies. With this loss of control, however, comes the opportunity for more consistent searching and browsing of bibliographies across projects that have submitted to these externally imposed constraints.

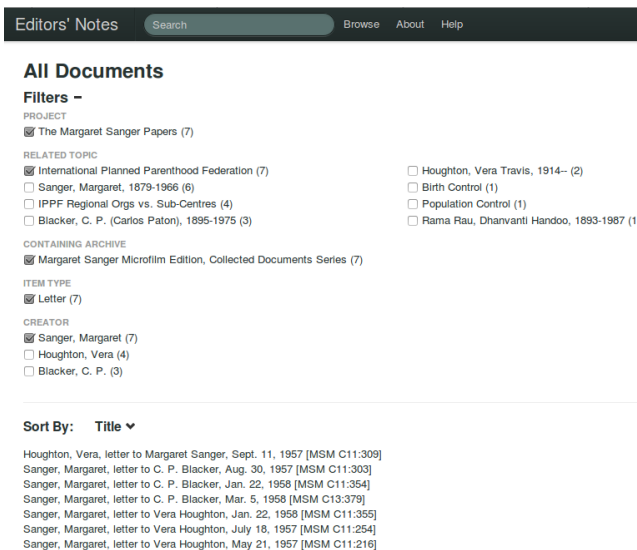
Bibliographic Units

Zotero defines a core set of “item types” such as Book, Journal Article, Letter, Manuscript, and so on (Zotero 2011). It is not possible to modify these core item types or add new ones without breaking compatibility with the Zotero platform. Thus, these types effectively serve to define the allowable bibliographic units of any bibliography stored in Zotero. This may be a problem for projects with an existing taxonomy of bibliographic units that cannot be cleanly mapped to the Zotero taxonomy. For example, the documentary editing projects that use Editors’ Notes often treat archival collections (rather than just the individual documents within them) as bibliographic units, yet Zotero currently has no Archival Collection type (thought there are plans to add one). Similarly, the Stanton and Anthony Papers project has traditionally recognized scrapbooks as a distinct and important type of document, yet Scrapbook is not a Zotero item type. Notably, both of these examples involve ambiguous bibliographic units that can be described either as individual resources or as collections of resources. These kinds of units are generally problematic for Zotero, given its “flat” taxonomy that does not recognize hierarchical relationships among item types.

While a standardized taxonomy of bibliographic units may be viewed as an intolerable constraint by some, it is precisely this standardization that eases collaboration among unrelated projects. In practice, we have found that the need to standardize has led to healthy examination and discussion of the treatment of bibliographic units by the various participating projects. Still, it would be ideal if Zotero offered a mechanism for sub-classing item types in its taxonomy without breaking compatibility.

Information fields

For each of its item types, Zotero defines an associated set of information fields. For example, the Case item type has Case Name, Reporter, Court, and Docket Number fields, among others. In contrast to the taxonomy of bibliographic units, we have not encountered any limitations related to the information fields defined by Zotero. This is probably due to the fact that the Zotero fields have been based on the same authoritative bibliographic formats already in use by the various editing projects. Yet while the choice of information fields has not posed any problems, control over the values provided for those fields has been an issue. As Zotero has been designed for use by a wide range of users, it sensibly does not enforce many constraints on the values of its information fields. But editorial projects often need to enforce standards for recording values such as dates, especially when those dates are imprecise or uncertain. As Zotero cannot enforce these standards, this requires implementing a separate validation step, thus complicating data synchronization, or else relying on users to manually enforce these standards. Editors’ Notes currently does the latter.



Editors' Notes Search Browse About Help

All Documents

Filters -

PROJECT

- The Margaret Sanger Papers (7)

RELATED TOPIC

- International Planned Parenthood Federation (7)
- Sanger, Margaret, 1879-1966 (6)
- IPPF Regional Orgs vs. Sub-Centres (4)
- Blacker, C. P. (Carlos Paton), 1895-1975 (3)
- Houghton, Vera Travis, 1914-- (2)
- Birth Control (1)
- Population Control (1)
- Rama Rau, Dhanvanti Handoo, 1893-1987 (1)

CONTAINING ARCHIVE

- Margaret Sanger Microfilm Edition, Collected Documents Series (7)

ITEM TYPE

- Letter (7)

CREATOR

- Sanger, Margaret (7)
- Houghton, Vera (4)
- Blacker, C. P. (3)

Sort By: Title ▼

Houghton, Vera, letter to Margaret Sanger, Sept. 11, 1957 [MSM C11:309]
Sanger, Margaret, letter to C. P. Blacker, Aug. 30, 1957 [MSM C11:303]
Sanger, Margaret, letter to C. P. Blacker, Jan. 22, 1958 [MSM C11:354]
Sanger, Margaret, letter to C. P. Blacker, Mar. 5, 1958 [MSM C13:379]
Sanger, Margaret, letter to Vera Houghton, Jan. 22, 1958 [MSM C11:355]
Sanger, Margaret, letter to Vera Houghton, July 18, 1957 [MSM C11:254]
Sanger, Margaret, letter to Vera Houghton, May 21, 1957 [MSM C11:216]

Figure 1. Faceted browsing of bibliographies.

Organization

By adopting the Zotero data model, Editors’ Notes has accepted the constraints that model places on the choice of bibliographic units and information fields. Yet by keeping its own copies of the bibliographic data, Editors’ Notes is free to implement its own organization of that data.

For example, Editors’ Notes provides an interface for faceted browsing and filtering based on the Zotero information fields (Figure 1). Resources, and the notes that cite them, are more easily discoverable when facets for “Item type,” “Publication date,” “Creator,” or other fields are provided, yet faceted search is not a feature provided by either the Zotero clients or its server API.

But the lack of faceted search is simply a missing feature of the Zotero platform, and one that could easily be added. More important is the organization achieved by situating bibliographic descriptions in a specific model of the research process, thus relating those descriptions to specific research topics, questions, and notes.

Selection Constraints

While Zotero imposes a number of presentation constraints on the bibliographies it manages, it rightly imposes few selection constraints at all. By specifying a fixed taxonomy of bibliographic units, Zotero does place some very broad constraints on the *domain* of bibliographies: the universe of possible resources that might be described. In practice, this is unlikely constrain the domain of a bibliography much at all, given that this universe still includes anything describable as a document.

Thus, the specification of a bibliography’s domain will primarily be the task of the integrating system. In the case of Editors’ Notes, the domain is defined implicitly by the kind of documentary editing projects for which it was designed. These projects rely on primary sources: archival

materials such as letters, diaries, and contemporary news articles. Textbooks, encyclopedias, and other secondary literature are mainly excluded. They embrace an *exhaustive* bibliographic selection principle: every resource in the domain that is found to meet the scope is included. As described in the following section, Editors' Notes enables a workflow in which editors are able to guide the scope and domain of their bibliographies through discussion threads, status indicators, and an approval mechanism.

Workflow Constraints

Fostering a collaborative research environment requires instating a system that indicates to potential contributors where and how work should be done. In the Editors' Notes framework, much of this structure is dictated by a project's central editor or editors, borrowing from the structure of a typical documentary editing project. Editors are able to manage project rosters, comment on the status of existing research, suggest new areas where new avenues of research are necessary, and moderate contributions to their projects' notes. Actions on documents and notes are built around a project-based permissions system. Project members are able to add, edit, or delete any of their project's material, while non-members are only able to amend that project's material, optionally at the behest of an editor's approval.

As notes are created in response to research queries, as in the IPFF example mentioned above, project editors are able to describe the status of the note's research. While this is currently accomplished in a "discussion" section, we have identified a model for three status indicators that take into account both the exhaustiveness of a note's bibliographic scope and its success in sufficiently addressing the query from which it was spawned. "Open" notes are ones that have not addressed the research query but have obvious further paths of research to be followed. "Closed" notes have sufficiently addressed the research query, and their augmentation with further documentary evidence is not a high priority. "Hibernating" notes have not addressed the research query and have no more obvious leads for further source material. They exist in a sort of stasis awaiting contribution of new sources and evidence.

This underlying structure of notes' workflows gives users of the Editors' Notes an idea of where contribution is most necessary and points interested parties toward those problems which are yet unresolved. For example, an independent researcher viewing a "hibernating" note can recognize that this status is an invitation to contribute further bibliographic evidence that he or she may possess.

CONCLUSION

Accurate scholarship depends on maintaining reliable control of successive versions of multiple documents, especially working notes that are continuously under revision. In addition, accurate bibliographical citation of resources used is fundamental to scholarship in all fields and a variety of standards and bibliographical software

systems have been developed toward that end. Substantial complexity arises when there is collaborative use of either the notes or the bibliographic references as well as in the relationships between notes and references. The "Editorial Practices on Web" project has developed and deployed a shared collaborative work environment for three leading documentary editing projects (Emma Goldman; Margaret Sanger, and Stanton-Anthony) and a related library special collection (Labadie). These are at four different universities, have significantly varying work practices, and involve some very specialized needs. The challenge is two-fold: to harmonize notes and references and to do that across four different projects. A major component is the partial integration of the Zotero reference management as harmonizing mechanism for the bibliographical references. The rationale, implementation, assessment, and future possibilities have been summarized.

ACKNOWLEDGMENTS

We are grateful to the Andrew W. Mellon Foundation for funding "Editorial Practices and the Web" (<http://ecai.org/mellon2010>), and for the cooperation and feedback of our colleagues at the Emma Goldman Papers, the Margaret Sanger Papers, the Elizabeth Cady Stanton and Susan B. Anthony Papers, and the Labadie Collection.

REFERENCES

- Bates, M. J. (1976). Rigorous systematic bibliography. *RQ*, 16(1), 7–26.
- Cohen, D. J. (2008). Creating scholarly tools and resources for the digital ecosystem: building connections in the Zotero project. *First Monday*, 13(8).
- Fourie, I. (2012). Collaboration and personal information management (PIM). *Library Hi Tech*, 30(1), 186–193.
- Jones, W., & Teevan, J. (2007). *Personal Information Management*. Seattle: University of Washington Press.
- Kline, M.-J., & Perdue, S. H. (2008). *A Guide to Documentary Editing*. Charlottesville: University of Virginia Press.
- Marino, W. (2012). Fore-cite: tactics for evaluating citation management tools. *Reference Services Review*, 40(2), 295–310.
- Hendry, D. G., Jenkins, J. R., & McCarthy, J. F. (2006). Collaborative bibliography. *Information Processing & Management*, 42(3), 805–825.
- Shaw, R., & Buckland, M. (2011). Editorial control over linked data. *Proceedings of the American Society for Information Science and Technology*. New Orleans: American Society for Information Science and Technology.
- Zotero project. (2011). Zotero client data model. Retrieved June 10, 2012 from http://www.zotero.org/support/dev/data_model.