

Ryan Shaw, Patrick Golden and Michael Buckland

4 Using Linked Library Data in Working Research Notes

Abstract: We describe how we are experimenting with using linked library data to connect the organizational infrastructures of libraries, archives, and special collections with those of researchers who use them. The site of this experimentation is Editors' Notes, an open-source hosted service for organizing the library and archives-based research of documentary editors. We report on an initial trial of incorporating linked data from libraries and other sources into Editors' Notes, and describe how that trial has informed our current effort to usefully exploit linked data to benefit researchers. We conclude with an overview of the potential benefits.

Keywords: Linked data, Digital humanities, Note-taking

Introduction

Documentary editing projects prepare 'editions' of documents such as letters, diaries, and essays that have value as evidence for political, intellectual, or social history (Kline and Perdue 2008, 2). These projects rely heavily on libraries, archives, and special collections as they research the people, places, organizations, events and ideas referenced in the documents they are editing. This research is organized in the form of "working notes" organized around particular questions. These notes are eventually distilled into footnotes in a published 'edition' of historical documents.

The published footnotes represent just a small fraction of the research produced by documentary editing projects. The majority of the research produced by editors and their assistants is represented by the working notes they develop to answer questions raised by their documents. These working notes are typically created during the course of researching a particular question and the specific forms they take may vary with the work practices of individual researchers. For example, consider a specific documentary editing project: the Margaret Sanger Papers Project (MSPP) at New York University, which has since 1985 been editing and publishing the papers of American birth-control pioneer Margaret Sanger.¹ A letter from an Indian birth-control activist to Sanger may raise some questions

¹ <http://www.nyu.edu/projects/sanger/>. Accessed on 23 December 2014.

in an MSPP editor's mind regarding the activist's role in planning a conference on birth control held in Sweden. The editor decides that some research is needed to better understand her role and the other people and organizations involved, and so she creates a "working note" to track this research (Figure 4.1). She structures the note as an annotated bibliography, with an entry for each resource found to contain useful information, and including both a bibliographic description of the resource and a summary of the information found.

The screenshot shows a web interface for 'Editors' Notes'. At the top, there are navigation links for 'Browse' and 'About', a search bar, and a user profile for 'Ryan Shaw'. The main title is 'Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood'. Below the title, there are 'Related topics' and tabs for 'About', 'Note', and 'History'. The 'Note' tab is selected, showing a list of sources with a 'Note' icon next to each. The first source is 'Rama Rau, Dhanvanthi. *An Inheritance: The Memoirs of Dhanvanthi Rama Rau* (New York: 1977)'. The second source is 'Suitters, Beryl. *Be Brave and Angry: Chronicles of the International Planned Parenthood Federation*. London : International Planned Parenthood Federation, 1973.'. The third source is 'Rama Rau, Dhanvanthi, Letter to Margaret Sanger, April 2, 1953 [MSM S41:211]'. The fourth source is 'Rama Rau, Dhanvanthi, Letter to Margaret Sanger, Apr. 28, 1953 [MSM S41:305]'. A text box next to the fourth source contains a summary: 'Still wants to have a strong Indian presence at the Stockholm meeting. "I have managed to get a donation of Rs. 5,000/- for one delegate and as the subjects, 'World Population Trends' and 'Food and other Resources in relation to World Population' figures so largely in the Programme, my own feeling is that we should invite Dr. S. Chandrasekhar of the Baroda University to attend the Conference. If only we could have two more passages paid, we could include in our delegation perhaps Dr. Gore who will be taking up work with us shortly, and a representative of the Research work that is done in India." Will contact Houghton about her ideas on the draft constitution.'. The fifth source is 'International Planned Parenthood Federation, *Fourth International Conference on Planned Parenthood, Report of the Proceedings, 17-22 August, 1953* (Stockholm, Sweden, London, 1953)'. The sixth source is 'Dhanvanthi Rama Rau, "The Family Planning Work of India," speech on Aug. 20, 1953. Edited version.'

Figure 4.1: A documentary editor's 'working note' about an Indian birth-control activist's role in planning the Fourth International Conference on Planned Parenthood.

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. But it is very likely that the working note contains a variety of useful information that will not appear in the published footnote due to either lack of space or unresolved issues with the information. Working notes may include doubts about published accounts; 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.

The MSPP editors seek to communicate the value of the documents they edit by contextualizing them, identifying and explaining within footnotes the people, places, organizations, events and ideas that documents' authors refer to

or imply. Dhanvanthi Rama Rau and the Fourth ICPP are two such topics. In the published edited volumes, these various topics are linked to each other and to sources in libraries and archives via a rich web of explanation and commentary. So while editorial projects are nominally organized around the ‘papers’ of a single historical figure, such as Sanger, projects end up carrying out in-depth research surrounding their studied era, reaching well beyond the narrow scope of their central figure’s biography.

As they look for relevant documents, researchers use the catalogues and finding aids created and maintained by librarians and archivists. But once researchers have found those relevant documents, they begin creating working notes and bibliographic data in separate organizing systems. Sophisticated researchers, such as those involved in documentary editing projects, may even construct structured compilations of data such as chronologies and name authorities. These systems could conceivably benefit from re-using data from catalogues and finding aids, but in practice this rarely happens.

This paper describes our ongoing effort to provide useful working tools for library- and archives-based research and our experiments with integrating linked data from libraries and other sources into these tools. First we provide some background describing the kinds of tools three documentary editing projects—the MSPP, the Emma Goldman Papers (EGP),² and the Elizabeth Cady Stanton & Susan B. Anthony Papers Project (ECSSBAP)³—currently use to organize their research. Then we discuss some of the problems with these tools, and introduce Editors’ Notes, an open-source hosted service designed to address these problems. We report on an initial experiment with incorporating into Editors’ Notes linked data from libraries and other sources, and describe how that experiment has informed our current effort to usefully exploit linked data to benefit researchers. We conclude with an overview of what we expect those benefits to be.

Current tools for organizing editorial research

Editorial projects rely on a mix of technologies to organize and manage their working notes and bibliographic data. While some projects still rely on handwritten notes, the general trend among the projects we have worked with has been away from paper pads towards scanned documents and typed notes. As is the case with most historians, documentary editors typically organize their notes

² <http://sunsite.berkeley.edu/goldman/>. Accessed on 23 December 2014.

³ <http://ecssba.rutgers.edu/>. Accessed on 23 December 2014.

in topically organized folders (Rutner and Schofield 2012). These folders reside in the project's filing cabinets or hard drives, and most of the material within them never leaves the project offices. For managing bibliographic data, editing projects may rely on 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, and specialized reference management software such as Zotero or EndNote. To varying degrees, these systems have allowed the projects to organize and manage descriptions of the sources they consult.

In addition to their notes, references, clippings, and photocopies, editorial projects also rely on specialized and locally-developed tools for recording structured compilations of data such as itineraries, chronologies, and legislative histories. Editors of personal papers usually need to create a detailed itinerary of their subjects' movements. The MSPP and ECSSBAP keep these types of records in Microsoft Access and dBase III databases, respectively. Researchers at the EGP keep detailed records of Emma Goldman's lecture tours⁴ and have made chronologies inside Microsoft Word and Word Perfect files for events such as the assassination of President William McKinley, the Preparedness Day Bombing of 1916, and the rise of the Industrial Workers of the World. Similarly, editors might create uniquely detailed legislative and legal histories of specific topics as the ECSSBAP editors have. Additionally, there are hundreds of informal chronologies spread across the working notes of each project.

Some technologically sophisticated projects also maintain specialized and locally developed tools for managing names (Hajo 1991). Names are always a challenge for historical research, and this is especially true for editorial projects. For example, many of the women researched by the MSPP had multiple different married names and alternated between using their maiden name and their married name at a given time. Dorothy Hamilton Brush, a close friend of Sanger's, appears in documents variously as Dorothy Adams Hamilton, Dorothy Brush, Dorothy Dick, and Dorothy Wamsley. Chinese and Indian names appearing in English-language documents may be transliterated in a wide variety of ways. Nicknames and children who share names with their parents or grandparents are also sources of confusion. A name authority database records preferred names, variant names, and identifying codes for every individual researched by the project. Each primary and secondary document controlled by the project is examined for personal names, and each name found is searched for in the name authority database. If a matching individual is found, the document is tagged

⁴ Some of which have been published as linked data at <http://metadata.berkeley.edu/emma/>. Accessed on 23 December 2014.

with an identifying code for that individual, and if the form of the name found in the document is a new variant, it is added to the name authority database. If no match is found, a new authority is created. To verify spelling and usage, periodically the name authority database will be manually checked against external authority files such as the Library of Congress Name Authority file.

Problems with current tools

The tools described above generally enable an editorial project to meet its goal of publishing edited volumes (though these often appear well behind schedule). But the organization and management of the valuable working notes these projects produce leaves much to be desired. 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. So working notes have tended to live separate lives from bibliographic descriptions, and as a result, notes on a particular source or topic may spread across handwritten notes, annotated photocopies, and word processing files. Consequently, duplication and loss of content is rampant. It is a major challenge to enable fine-grained access to and indexing of notes without making researchers feel as if they are working with “a million little pieces”. On the one hand, researchers ought to be able to search for and link to “atomic” notes taken on a single source document as it relates to one narrow topic. Yet they also need to be able to work with flexible aggregations of these small ‘atoms’ in ways that feel natural to them.

One of the ways researchers would like to work with their notes is through visualization. Editors often wish to create visualizations displaying temporal, geospatial, or collective biographical information to concisely express large datasets, reinforce explanations, or offer aesthetically pleasing displays for public exhibition. But the difficulty of aggregating and repurposing the information in their working notes often precludes this. Working notes stored in formats such as WordPerfect documents require non-trivial amounts of special post-processing in order to turn them into structured data that can then be fed to visualization libraries. This requires outsourcing such work to experts who have mastered the use of specialized tools for text processing and data visualization, thereby decoupling the visualization of information from the process of research that created it. This becomes a problem if, for example, a researcher who is a technological layperson wants to correct or amend a timeline created by a since-departed specialist. The problem is not a lack of tools for generating map displays, timelines,

prosopographies, and the like, but, rather, how to incorporate such tools into the work routines of hard-pressed editors and their assistants with an acceptably low threshold of learning and effort.

Finally, just as the research of individual editorial projects is unnecessarily fragmented across different working notes due to organizational issues, the funding and organization of editorial projects leads to unnecessary fragmentation of research across different projects and over time. The majority of the research produced by editorial projects is not included in the published volumes, is not shared with other researchers, and is discarded when grants for publication expire. 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 overworked colleagues with repeated requests for help. This results in fragmentation across parallel but related research efforts. Another kind of fragmentation occurs over time. Scholarly editing requires a sustained investment of highly specialized expertise, but long-term funding is difficult to find and usually narrowly limited to support for the eventual published edition. When the manuscript of the final volume is ready for publication, the editors and staff retire or move on and their working notes become effectively inaccessible if not discarded. But even as projects expire, scholarship continues. The ideal would be if the editorial “workshop” could remain ready to support resumed scholarship as and when labor and funding allow.

Editors’ Notes

It is to address the problems discussed above that we have developed Editors’ Notes,⁵ an open-source hosted service for organizing the research of documentary editing projects (Figure 4.2, p. 54). Editors’ Notes enables the publishing, sharing, and editing of working notes throughout the research process, from the initial posing of specific research questions all the way through to the authoring of polished footnotes and articles. The mission of the service is much the same as the one William Thoms (1849, 2) identified when he founded *Notes and Queries*: to provide a ‘medium by which much valuable information may become a sort of common property among those who can appreciate and use it.’

Editors’ Notes is organized around *notes*, *documents*, and *topics* (Figure 4.3, p. 54). A note is any kind of working note written by an editor. Editors can use a WYSIWYG interface to easily edit notes, and all past versions of edited notes are

⁵ <http://editorsnotes.org/>. Accessed on 23 December 2014.

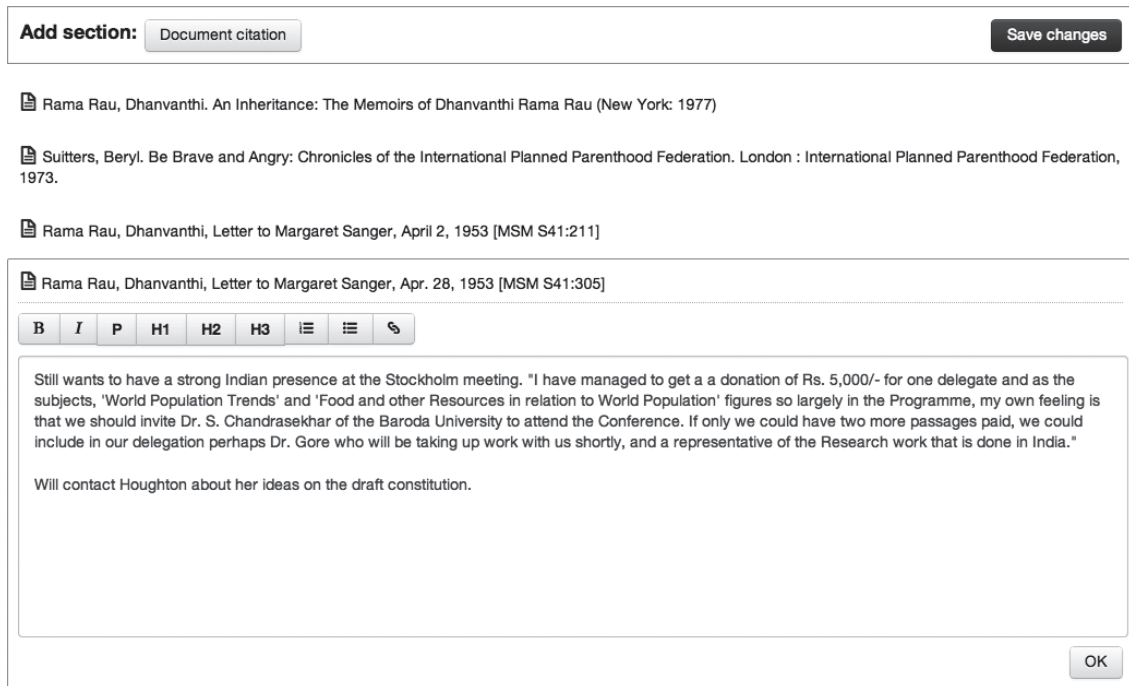


Figure 4.2: Using Editors’ Notes to edit one section of a working note on ‘Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood’.

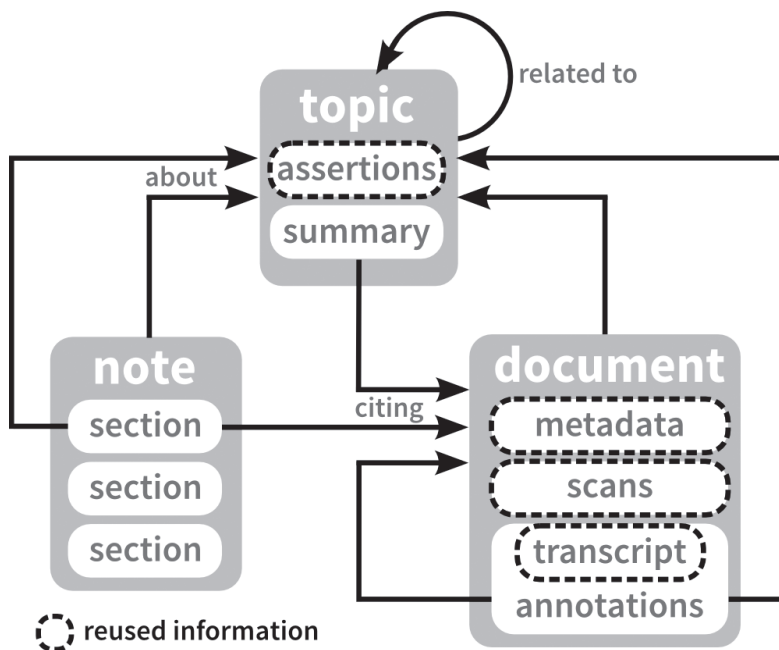


Figure 4.3: Part of the Editors’ Notes data model. Factual assertions, metadata, scans and transcripts may be authored locally or linked to elsewhere.

saved. Notes are stored as HTML, so they may have hyperlinks and all the other features that HTML enables. Notes are structured into topic-specific and

source-specific sections, allowing researchers to search for and link to note sections taken on a single source document as it relates to one narrow topic. For example, a working note on “Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood” (Figure 4.2) might reference dozens of documents and relate to other topics such as the Family Planning Association of India. Researchers can work with these note sections in the context of the broader working note, or they can pull together all the note sections about Rama Rau, whether or not these were taken in the course of researching “Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood”. Each note is categorized based on its completeness: notes are open when they require more work; closed when deemed completed; and hibernating when a resolution remains desired but appears impractical, of low priority, or unobtainable given available sources.

Notes cite or explain documents. A document is anything that an editor is editing (e.g. the letters, diary entries, speeches, etc. that are the focus of the project) or is citing (any supporting evidence found in the course of the editor’s research). Documents may have attached scans, transcripts (with optional annotations), and hyperlinks to external websites. Transcripts can be footnoted, and footnotes can cite other documents. Editors’ Notes integrates with Zotero to manage document metadata (e.g. item type, author, title, archive), enabling the input and output of documents as structured bibliographic records (Shaw, Golden and Buckland 2012).

Notes and documents are indexed using terms drawn from a controlled vocabulary of topics. Topics may be person names, organization names, place names, event names, publication names, or names of concepts or themes. Each topic may have a summary, a free-form textual description or explanation of the topic. We can think of topics 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.

Integrating linked data into Editors’ Notes

The topics in Editors’ Notes provide a natural point for integration with and consumption of linked data from libraries and other cultural heritage institutions. Topics can be augmented with structured assertions created locally and/or imported from trusted external datasets. These assertions can then be used as the basis for specialized search and visualization interfaces. Assertions may

relate separate topics, and just as topics are used to index notes and documents, relationships between topics might also be used for indexing and discovery. For example, Emma Goldman and Alexander Berkman were not only lovers but also co-editors of the anarchist journal *Mother Earth*. Their co-editing relationship might be used to index a document that is relevant to the latter relationship but not the former.

To investigate these possibilities, we added to Editors' Notes capabilities for harvesting and editing linked data (Shaw and Buckland 2011). The harvester ran periodically, looking for new linked data related to topics. For each topic name, the harvester queried a co-reference service (Glaser, Jaffri and Millard 2009) to obtain sets of candidate URIs for that name. Each set contained URIs that had been asserted to refer to the same entity. For example, when queried with the name "Emma Goldman", the first set of URIs returned by the co-reference service includes the identifiers for Emma Goldman from VIAF (Loesch 2011) and Freebase (Bollacker et al. 2008). For each set in the response, each URI in the set was dereferenced and the resulting data was examined. If, for any of the dereferenced URIs, this data included a valid label, and the value of this label matched the topic name, then the whole set of candidate URIs was accepted. Otherwise, the set was rejected and the next set was examined. In this way, the harvester obtained a set of zero or more URIs for each topic. The harvester then stored all assertions obtained by dereferencing the URIs.

Editorial control over assertions about topics

Each assertion was stored in two separate graphs: one graph contained all the candidate assertions about a given topic, while the other contained all the assertions from the same source (e.g. DBpedia, VIAF, Deutsche Nationalbibliothek, etc.). The topic-specific graphs made it simple to display all the assertions found for a given topic, while the source-specific graphs made it easy to request fresh data from a given source.

Once a set of candidate assertions about a topic had been obtained, editors could use the linked data editor to accept or reject them (see Figure 4.4, p. 57). Accepting and rejecting assertions could happen at different levels of specificity. An editor could reject a single assertion that she judged to be inaccurate. Or she could choose to reject all assertions that shared a given predicate that had been judged irrelevant to the editing project. For example, many DBpedia (Auer et al. 2007) resources contain assertions about what templates are used on their corresponding Wikipedia pages, and this information is not likely to interest editors.

Finally, an editor could accept all the assertions about a given topic, or all the assertions from a given source. When an editor accepted assertions from a given source, this was treated as evidence that the identifier from that source referred to the same entity, and an owl:sameAs assertion was created linking the topic to that identifier. Thus the process of accepting assertions had the effect of linking Editors' Notes topics to standard identifiers in external systems. Accepted assertions were inserted into a graph associated with the editor who accepted them. This way the provenance of published assertions was made clear, and editors could choose whether they needed to further assess assertions accepted by less expert contributors (i.e. student assistants).

The Editors' Notes Project

Logged in as Ryan Shaw

Add Topic Add Note Add Document Member Projects Help Log out

Goldman, Emma, 1869-1940

Related topics: Content, Harold A. (1887-1944), Havel, Hippolyte (1869-1950)

Article Discussion Related Notes & Queries (1) Related Documents (172) Facts

From <http://viaf.org/> about Emma Goldman:

name	Goldman, Emma ✓ ✕	Goldman, Emma, 1869-1940 ✓ ✕	Goldmann, Emma, 1869-1940 ✓ ✕	Gol'dman, Èmma 1869-1940 ✓ ✕	Goldmann, E. 1869-1940 ✓ ✕	Görudoman, Ema, 1869-1940 ✓ ✕	אמנוע, ואמדולאג ✓ ✕	ימדולוג, הומא, 1869-1940 ✓ ✕
date of death	1940 ✓ ✕							
type	Person ✓ ✕	Person ✓ ✕						
date of birth	1869 ✓ ✕							
image	Emma_Goldman_seated ✓ ✕						http://rdf.freebase.com/ns/m.02bdhh ✓ ✕	

Figure 4.4: The interface for accepting or rejecting assertions in the first iteration of the Editors' Notes linked data editor.

Lessons learned and current direction

Our pilot project demonstrated that it was possible to automatically harvest relevant linked data from cultural-heritage institutions and other datasets. A simple interface for accepting or rejecting harvested assertions made it easy for researchers to exert editorial control over the harvested data. Ultimately, however, our pilot did not demonstrate that the benefits of additional structured data were worth editors spending their (or their interns') time accepting or rejecting it. Thus

our current development effort is focused on not simply aggregating and editing linked data but usefully exploiting it to researchers' benefit. Before turning to a discussion of those expected benefits, we will briefly describe the new tools we are currently developing and how they exploit linked data.

The first change involves how topics are created and linked to. In the pilot implementation, editors created topics to label and index their notes and documents, and these topics were later reconciled to external identifiers in a separate batch process. In the new system, editors fluidly create, link to, and reconcile topics within the note-taking process. When in the course of note-taking an editor types the name of a place, person, organization, or other such entity, the interface offers elective autocompletion from a ranked list of matching candidates found among existing topics or in external datasets. Using this interface an editor can create and link to internal topics or external entities without interrupting her research workflow.

For example, suppose an editor is creating a note about Rama Rau's role in planning the Fourth ICPP in Stockholm, Sweden. After she types the letters "R-a-m-a R-a-u" she presses the tab key, triggering a search of the Editors' Notes entity index for existing topics or previously indexed external entities with labels containing those letters. The label "Rama Rau, Dhanvanti Handoo, 1893–1987" appears in a small menu below the cursor, and the editor presses enter to confirm that this is the Rama Rau referred to in the note. Because this label is from an external entity not previously used within the site, it is automatically established as a new Editors' Notes topic. The text of the note is left unchanged, but a link is created between the string "Rama Rau" in the note text and the new topic.

In case no suitable internal topics exist and none of the previously indexed external entities are relevant, the system will enable incremental creation and "offline" asynchronous reconciliation of new topics. For example, suppose the editor types the letters "S-t-o-c-k-h-o-l-m" and presses the tab key, again triggering an entity search. This time no suitable results are found, so instead of candidate labels, a small "Create Topic?" menu appears with the options Person, Place, Organization, Event, and (selected by default) Topic. The editor selects "Place" and is prompted to edit the default topic label "Stockholm". She adds "Sweden" to complete the topic label, and a link is created between the mention of "Stockholm" in the text of the note and the newly-created topic labelled "Stockholm, Sweden". The editor then continues with her research and note-taking.

In the background, Editors' Notes has begun querying a configurable set of linked data endpoints, trying to reconcile the label Stockholm, Sweden with identifiers managed by those endpoints. It discovers the following candidate identifiers:⁶

⁶ All URLs given in the text were accessed on 24 December 2014.

- <http://viaf.org/viaf/153530943> (VIAF geographic name Stockholm (Sweden));
- <http://sws.geonames.org/2673730> (Geonames place Stockholm), and
- <http://dbpedia.org/resource/Stockholm> (DBpedia populated place entity Stockholm).

Once this search for candidate identifiers has finished, an unobtrusive notification appears, letting the editor know that she can finish the reconciliation process when she wishes. Not wanting to interrupt her work, the editor ignores the notification.

The next day, the editor logs into Editors’ Notes again. On her dashboard she sees a reminder that the new topic “Stockholm” has not yet been reconciled. She clicks on the reminder and is taken to a page displaying data from VIAF, Geonames, and DBpedia. Scanning the data, she sees that the entities identified within each of these services are equivalent to one another.⁷ She decides that they are also equivalent to the Editors’ Notes Topic “Stockholm, Sweden” and clicks a button to finish the reconciliation process, thereby associating the three external identifiers with the new topic.

Our pilot implementation enabled storing and editing assertions about topics from external datasets, but provided no incentive for editors to do this. In the new system, storing and editing external assertions about topics happens in the context of sorting and filtering lists of notes and documents and creating simple visualizations of the information held in working notes. Thus aggregating and editing linked data is not an unmotivated activity, but is seamlessly integrated into the process of managing and using working notes. Thus through a combination of selectively importing data from external links and entering it locally, topics are gradually and incrementally enriched from a mere list of generic “things” to a structured group of semantically distinct and descriptive entities suitable for advanced querying and manipulation. Importantly, researchers have full editorial control over this data, ensuring its high quality and compatibility with their painstaking scholarship. In addition to editing linked data associated with topics, editors also can create notes consisting of structured data rather than free text. This is expected to be useful for creating highly structured working notes such as itineraries, chronologies, and legislative histories.

Many interactions with Editors’ Notes involve working with lists of note sections and documents:

1. Upon browsing to the topic “India, birth control movement in”, one sees lists of note sections and documents related to that topic (Figure 4.5).

⁷ This may be asserted by an external linking framework such as Silk (Volz et al. 2009).

2. The note on “Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood” is a list of note sections on various letters and reports that reference that event (Figure 4.1).
3. A search for “Rama Rau” returns a list of notes and documents with that name in their text (Figure 4.6, p. 61).

Editors' Notes Browse About Search Ryan Shaw

India, birth control movement in

Related topics: Ahluwalia, Gopaljee | All-India Women's Conference | Datta, Alexandrena (Rena) Carswell, 1886-1978 | India, gender roles in | India, government (1858-1947) | Indian Birth-control Society | International Planned Parenthood Federation, Indian Ocean Region | Mysore State (India) | Peters, Hannah | Phadke, Narayan Sitaram, 1894-1978 | Pillay, Aliyappin Padmanabha (AP) (1889-1956) | Rama Rau, Dhanvanti Handoo, 1893-1987 | Shastri, P. D. | Vembu, Elfriede | Watumull, Gobindram Jharmandas, 1891-1959

Article Related Notes & Queries (34) Related Documents (91)

Sanger and the Third ICPP
Last modified: July 9, 2013
Related topics: International Conference on Planned Parenthood, Third (Bombay, 1952) | Family Planning Association of India | International Committee for Planned Parenthood | Sanger, Margaret, 1879-1966 | Rama Rau, Dhanvanti Handoo, 1893-1987 | Watumull, Ellen Jensen, 1897-1990

Indian Impressions of Sanger, 1930s
Last modified: April 15, 2013
Related topics: India, doctors in | India, contraceptives in

Dhanvanthi Rama Rau and Margaret Sanger relations
Last modified: April 15, 2013
Related topics: Rama Rau, Dhanvanti Handoo, 1893-1987 | Family Planning Association of India | International Planned Parenthood Federation

Indian birth control movement
Last modified: August 11, 2012
Related topics: India, population | India, death rates in

Figure 4.5: Notes and documents related to the topic “India, birth control movement in”.

Working flexibly with lists like these requires the ability to filter and sort them easily. Currently researchers can filter and sort lists of documents (and the notes that cite them) using the bibliographic metadata associated with the documents. But when not only documents, but also places, people, organizations, and events have structured data associated with them, we can provide far more powerful facilities for filtering and sorting notes and documents. Given a list of notes on “India, birth control movement in”, users can filter and sort them not only using the dates of the cited documents (as they currently can), but also using the locations and birth and death dates of the people referenced in the notes, the locations and dates of existence of the organizations referenced, or the locations and dates of the events referenced.

Furthermore, by taking advantage of structured spatial and temporal metadata, we are no longer restricted to presenting lists of notes and documents textually. The note on “Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood” can become viewable not only as a text document, but also as a map of specific locations in Stockholm and Bombay, a timeline of dates

The screenshot shows a web interface for 'Editors' Notes'. At the top, there are navigation links for 'Browse' and 'About', a search bar, and a user profile for 'Ryan Shaw'. Below the search bar, it says 'Search results for Rama Rau:'. A 'Filters' section is visible with a '+ Result type' button. The search results are a list of 8 items:

1. Note: Dhanvanthi Rama Rau - basic bio
...Rama Rau - basic bio Lady Dhanvanthi Handoo Rama Rau (1883-1987), president of the Indian FPA and a prominent woman's activist, was organizing the Third International Conference on Planned Parenthood ...
2. Note: Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood
...Rama Rau & the Fourth International Conference on Planned Parenthood Rama Rau is a major player in the IPPF. Need context for quotes. MS wrote: "We were all set up on the platform, Dr Blacker the..."
3. Document: Rama Rau, Dhanvanthi. *An Inheritance: The Memoirs of Dhanvanthi Rama Rau* (New York: 1977)
...Rama Rau. Dhanvanthi. *An Inheritance: The Memoirs of Dhanvanthi Rama Rau* (New York: 1977)
4. Note: Dhanvanthi Rama Rau and the Third ICPP
...Rama Rau and the Third ICPP Rama Rau was the hostess for the Bombay Conference, held in 1952, which introduced her to the international planned parenthood community. Need context for: Blacker wrote: "...
5. Document: Rama Rau, Dhanvanthi, letter to Margaret Sanger, Sept. 5, 1951 [MSM C9:172]
...Rama Rau. Dhanvanthi, letter to Margaret Sanger, Sept. 5, 1951 [MSM C9:172]
6. Note: Dhanvanthi Rama Rau trip to the Soviet Union in late 1950s
...Rama Rau trip to the Soviet Union in late 1950s We think she went in 1957 or 1958.
7. Document: Rama Rau, Dhanvanthi, letter to Margaret Sanger, Dec. 11, 1951 [MSM C9:279]
...Rama Rau. Dhanvanthi, letter to Margaret Sanger, Dec. 11, 1951 [MSM C9:279]
8. Topic: Rama Rau, Dhanvanti Handoo, 1893-1987
...Rama Rau. Dhanvanti Handoo, 1893-1987 Influential Indian birth control activist, president of International Planned Parenthood Federation.

Figure 4.6: A search for documents with the string “Rama Rau” in their text.

associated with the conference, or a network of relationships among people and organizations. Maps, timelines, and network graphs correspond most naturally to places, events, and personal relationships, but any topic which has geographic coordinates can be mapped, any topic with time points or ranges can be put on a timeline, and any relationships among topics can be visualized as a network. These three genres of visualization together are, therefore, broadly applicable to any kind of structured data about topics that might be gathered. Documents have equivalent data (when and where published; authorship) allowing the same types of visualizations for them too. So, for example, a map display can show any location(s) mentioned in a note, with options to display the locations of other related topics and documents in any number of ways as determined by the interests of the editors.

Now, instead of doing unmotivated data wrangling, editors can do it gradually and incrementally as part of the process of sorting, filtering, and visualizing their working notes. For example, suppose an editor decides she wants to visualize the network of people involved with the International Planned Parenthood Federation (IPPF). She generates an initial visualization by importing assertions from Wikipedia to show a network of relations among people involved with the IPPF including Margaret Sanger, Dhanvanthi Rama Rau, Margaret Pyke, and Lloyd Morain. But the visualization is missing labels for the links between these people and the IPPF, since that information is not provided in Wikipedia. She clicks on the node for “Pyke, Margaret Chubb, 1893–1966” in the visualization and

is shown all the structured data associated with that topic, including Pyke's birth-date, birthplace, date of death, and place of death. She presses the edit button and is presented with an editable tabular view of the facts. She edits the assertion linking Pyke to the IPPF to reflect that she was head of the Family Planning Association of England, a national affiliate of the IPPF. Although this information was added to the "Pyke, Margaret Chubb, 1893–1966" topic, it will now also appear among the facts related to the "International Planned Parenthood Federation" topic.

Benefits of consuming linked data

The tools described can potentially remove some tedious, duplicative work from everyday research. Editors can import contextual details of, for example, persons (e.g. birth and death dates, place of birth, and other names) or of places (alternative names, containing jurisdiction, latitude and longitude) without researching or transcribing these details at every mention. Data accumulation is incremental and comes as a by-product of the routine editing effort. Linking to external datasets can bring the benefit of automatic updating as additions and corrections are made to the resources to which they are linked. These are welcome conveniences. But the major potential benefits of consuming linked data are threefold: making working notes repurposable, replacing name authority files with naming services, and shifting the focus of editorial projects from product to process.

Making working notes repurposable

Researchers already spend considerable time and effort producing working notes. But because these notes consist primarily of unstructured text, they are not easily repurposable. More structured and therefore repurposable working notes enable researchers to use their research in new ways. For example, a researcher can aggregate the chronologies of several different events together into a larger timeline to look for connections or patterns that might not have been evident when viewing each separately. Reducing the expertise required to create such visualizations enables and encourages experimentation with the different kinds of information held in working notes. Giving researchers without specialized knowledge the ability to author, edit, and control structured documents and derived visualizations has the potential to change the nature of how researchers use their working notes. Instead of visualizations like maps and timelines being

handcrafted for one-off exhibitions, they become works in progress, able to be controlled with the same editorial touch as the textual record.

Replacing name authority files with naming services

Linking names to external authorities enables project-specific name authority databases to be replaced or augmented by shared naming and reconciliation services, which has a number of advantages. The process of checking spelling and usage can be further automated, and every project benefits whenever the naming service is updated. A naming service can power an auto-completion function that allows editors to quickly and easily link names to identifiers in any context, such as authoring a working note. And rather than using the separate name authority database to produce reports of variant names to be manually consulted when searching, a naming service can be used to automatically expand searches to include variant names. Finally, all of these amenities can be extended beyond personal names to other kinds of names (events, organizations, places, subjects). Another benefit is that by mapping their topics and entities to identifiers for those topics and entities elsewhere, editorial projects can make their research products more widely accessible. By linking their working notes to external identifiers for people, places, organizations, events and ideas, editorial projects make their work far more interoperable with other scholarly projects, library catalogues, archival finding aids, and open knowledge projects such as Wikipedia.

Shifting from product to process

Used shrewdly, the web and linked data have the potential to dissolve the silos that isolate research in separate editing projects and to enable the serendipitous discovery of useful material by other researchers and the wider public. They also have the potential to shift the focus of editorial project funding and organization away from the published edition as the one and only product. Currently editorial expertise and project working resources are treated as expendable means to that sole objective. But changed 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 and valued by-products. Scholarly communication could be greatly extended if it

were feasible not only for researchers anywhere to have sustained access to the working notes, but also for researchers 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.

Conclusion

Documentary editing projects exemplify a form of library and archives-based research that partially replicates the organizational infrastructure of libraries, archives, and special collections. Were that infrastructure to be made usable in a standardized form such as linked data, these projects could both reuse data from and contribute data back to catalogues, finding aids, and other sources. These tools would no longer be used solely to find documents but would contribute to the ongoing organization of research that uses and expands upon those documents. Ideally, the result would be a continuously updated compilation of facts related to specific people, places, organizations, events and ideas.

For this vision to become reality, it will be necessary to find ways of consuming and using linked data that accord with the working practices of researchers. User interfaces for consuming independently originated linked data must save users more time and effort than they require. Achieving this ideal will require close attention to specific use cases and contexts. Linked data tools and infrastructure must adapt to these use cases and contexts, rather than vice versa. It remains to be seen whether this challenge can be met by the linked data research community, which has tended to put the cart before the horse by focusing on the technical details of data formats and communication protocols instead of trying to understand potential contexts of use.

Acknowledgements

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 Joseph A. Labadie Collection.

References

- Auer, Sören, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak and Zachary Ives. 2007. "Dbpedia: a nucleus for a web of open data." In *The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11–15, 2007, Proceedings*, edited by K Aberer et al., 722–735. Berlin: Springer.
- Bollacker, Kurt D., Colin Evans, Praveen Paritosh, Tim Sturge and Jamie Taylor. 2008. "Freebase: a collaboratively created graph database for structuring human knowledge" *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data – SIGMOD '08*. doi:10.1145/1376616.1376746. Accessed on 23 December 2014.
- Glaser, Hugh, Afraz Jaffri and Ian C. Millard. 2009. "Managing co-reference on the Semantic Web" *Proceedings of the WWW2009 Workshop on linked data on the Web*. http://ceur-ws.org/Vol-538/ldow2009_paper11.pdf. Accessed on 23 December 2014.
- Hajo, Cathy Moran. 1991. "Computerizing control over authority names at the Margaret Sanger Papers" *Documentary Editing* 13(2): 35–39.
- Kline, Mary-Jo and Susan Holbrook Perdue. 2008. *A Guide to Documentary Editing*. 3rd edition. Charlottesville: University of Virginia Press.
- Loesch, Martha Fallahay. 2011. "VIAF (The Virtual International Authority File) – <http://viaf.org>" *Technical Services Quarterly* 28(2): 255–256. doi:10.1080/07317131.2011.546304. Accessed on 23 December 2014.
- Rutner, Jennifer and Roger C. Schonfeld. 2012. *Supporting the Changing Research Practices of Historians*. New York: Ithaka S+R. <http://www.sr.ithaka.org/sites/default/files/reports/supporting-the-changing-research-practices-of-historians.pdf>. Accessed on 23 December 2014.
- Shaw, Ryan and Michael Buckland. 2011. "Editorial control over linked data" *Proceedings of the American Society for Information Science and Technology* 48(1): 1–4. doi:10.1002/meet.2011.14504801296. Accessed on 23 December 2014.
- Shaw, Ryan, Patrick Golden and Michael Buckland. 2012. "Integrating collaborative bibliography and research" *Proceedings of the American Society for Information Science and Technology* 49(1): 1–4. doi:10.1002/meet.14504901245. Accessed on 23 December 2014.
- Thoms, W. J. 1849. "Notes and queries" *Notes and Queries*, 1st series 1(1): 1–3. doi:10.1093/nq/s1-l.1.1 Accessed on 23 December 2014.
- Volz, Julius, Christian Bizer, Martin Gaedke and Georgi Kobilarov. 2009. "Silk—a link discovery framework for the web of data" *Proceedings of the 2nd Workshop about linked data on the Web (LDOW2009), Madrid*. . http://ceur-ws.org/Vol-538/ldow2009_paper13.pdf. Accessed on 24 December 2014.