Taking scholarly note-taking to the Web

Michael Buckland
University of California, Berkeley

Ryan Shaw
University of North Carolina at Chapel Hill
the past

1. Documentary editing

2. Current state of Editors’ Notes

3. Roadmap: exploiting linked data

4. Roadmap: hibernating scholarship

the future
Documentary editing

- Editors prepare collections of documents: letters, articles, diaries, essays, etc.

- Printed volumes provide **context** for better understanding subjects’ experiences and general milieu through footnotes, images, chronologies, articles.
Documentary editing: workflow

1. Gather documents
2. Contextualize select items
3. Publish final product
4. Repeat as funding allows
Case study: Emma Goldman Papers
Dear Sir

I am in receipt of your note of the 18th inst. and in reply would like to say this, I do not know whether Czolgosz was an Anarchist, nor have I the right to say that he was not. I have not known him sufficiently to be acquainted with his political Views. In fact no one has, including Mr Isaak and Mr Shilling and if they have expressed opinions either for or against his being an Anarchist they have done an injustice to the young man, who died such a brave death, as you and you, dear Sir will be guilty of the same offence if you attempt to prove that Czolgosz was not or was an Anarchist. No one can with certainty say, that the man was an Anarchist, since he was but little known, and since he has never proff made a public statement to that affect. Still less can one prove that he was not one; there is nothing in his act by which one can deny him the right of being an Anarchist. It is true, the Philosophy of Anarchy does not teach Invasion, but it does teach self defence, and Czolgosz’s act was an act of self defence and nothing else. You may question this, since Czolgosz was not personally attact by McKinley, quite true, but Czolgosz belonged to the Oppressed, to the Exploited and Disinherited Millions, who lead a life of darkness and despair owing to those, of whom McKinley was one, therefore he was personally attacked by the ex President, or rather he was one of the Victims of the McKinley regime and those McKinley catered to. The act of Czolgosz may have been impractical or inupportune, I will not argue this point now, but I insist it had nothing unanarchistic about it, since, as I said before Anarchism claims the right of Defence against Invasion and Aggression of every shape and form and no one, who has his eyes open will and can deny that those in Power are the Invadors, and McKinley certainly was one of them. I send you with this, 2 Articles

1. Leon Czolgosz, who assassinated President McKinley at the world’s fair in Buffalo in September 1901, was judged sane before his trial by three psychiatrists for the prosecution and two psychiatrists for the defense. Later, Tufts University Medical School psychiatrist Walter Channing assessed Czolgosz as mentally unstable, based strictly on the assassin’s personal history, in a October 1902 article published in the American Journal of Insanity.

2. Cleveland anarchist Emil Schilling had been Czolgosz’s first contact in the anarchist movement in May 1901, lending him reading materials and answering his questions. According to Schilling they were estranged even before Czolgosz’s arrival in Chicago. Abe Isaak, editor of Frei Society, introduced Czolgosz to EG in Chicago; Isaak found Czolgosz’s behavior suspicious.
Patrick—

Lenin:
Had any of his family members beside his brother, been imprisoned?

What was the book he had written on ‘political economy’ that was used in Russian Universities?

New York (Evening?) Post, September 1918 editorial on IWW verdict for the huge IWW trial in Chicago.
Lenin, Vladimir Ilich [V.I. Ulyanov] (1870–1924) Russian revolutionary leader and founder of the Bolshevik party, a faction of the Russian Social Democratic Labor Party formed in 1903. Lenin grew up east of Moscow in the city of Simbirsk in a relatively wealthy family. When he was a teenager, his brother was executed for his role in plotting to kill Tsar Alexander III. While studying to become a lawyer, Lenin began his involvement in radical and Marxist circles in the late 1880s and 1890s. By the last years of the nineteenth century, he was a prolific writer and published many newspapers and articles encouraging revolution in Russia. To avoid being punished for his activities by the Tsarist police, Lenin spent much of the decades preceding the 1917 Russian revolution in exile in London, Paris, Geneva, Zurich, and other European cities, working with other émigrés to encourage revolution in Russia. After the overthrow of the Tsar in February 1917, Lenin planned his trip back to Russia, arriving in Petrograd in April. His return was controversial because his train from Switzerland was allowed to pass through Germany, one of Russia’s enemies in World War I. This fact led many to believe that Lenin was
Problems

- Published volumes & necessary work are expensive
- Lack of space for all footnotes
- Much of research done is either glossed over in footnotes or not included at all
  - Fact checking
  - Falsification or dead ends
  - Tangential biographical details
- Preservation & legacy
1. Documentary editing

2. Current state of Editors’ Notes

3. Roadmap: Exploiting linked data

4. Roadmap: Hibernating scholarship
Editors’ Notes

- http://editorsnotes.org/
  http://ecai.org/mellon2010/

- Finding a safe place for the “debris” of research

- Improving return on investment for documentary editing projects

- Central focus on changing work practices of editors and researchers rather than digitizing what already exists
Design principles

- Minimal amount of “friction” for researchers
- Flexibility for different work habits
- Consistency in data models
- Existing technology wherever possible
- Adherence to web standards
Data model
Documents

Zotero for document metadata

High quality, zoomable scans

Transcripts in HTML with interface to annotate passages of text
Primary method of indexing items

Classified by type

Interface for clustering and merging
Notes

- Most difficult part of the project
- Notes are messy, and purposefully so
- How to model something so chaotic & idiosyncratic?
- Goals: Easy to use; flexible but consistent
Notes

Description

Status: Open, closed, hibernating

Assigned users

Sections

Citation with optional notes

Stored as HTML

Revision history
Examples

- Aftermath of Sanger’s trip to India
- Test Cases during Reconstruction
- "Address at the Kate Richards O’Hare Testimonial Dinner"
What changed for researchers?

Free text → Structured blocks

Implicit people, places, events → Explicit linkable entities

Filing cabinets → Open access
Benefits

- Connections linking topics are freed from the minds of editors & researchers and indexed for anyone to see
- Standardized records of work can easily be revisited from within a project or from outside
- New way of seeing the outer edges of humanities research
- Evidence of intense, often messy, scholarship behind concise, clean footnotes
Open source

- https://github.com/editorsnotes
- Django web framework
- PostgreSQL database
- Haystack for full-text searching
- Zotero for document description
- Google Refine for duplicate detection
- Mozilla Persona for ID management
1. Documentary editing

2. Current state of Editors’ Notes

3. Roadmap: exploiting linked data

4. Roadmap: hibernating scholarship
Desired enhancements

- Better sorting, filtering and aggregating of notes
- Improved naming control and discoverability
- Temporal, geospatial, and relational visualization
Adding structured data

3 approaches:

- Schema-based authoring
- Automatic extraction
- Human-in-the-loop reconciliation
Goldman, Emma, 1869-1940

**From VIAF about Goldman, Emma:**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Name</strong></td>
<td>Goldman, Emma</td>
</tr>
<tr>
<td></td>
<td>Goldman, Emma, 1869-1940</td>
</tr>
<tr>
<td></td>
<td>Goldmann, Emma, 1869-1940</td>
</tr>
<tr>
<td></td>
<td>Gol’dman, Emma, 1869-1940</td>
</tr>
<tr>
<td></td>
<td>Amme, Ema, 1869-1940</td>
</tr>
<tr>
<td><strong>Date of Death</strong></td>
<td>1940</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>Person</td>
</tr>
<tr>
<td></td>
<td>Person</td>
</tr>
<tr>
<td><strong>Date of Birth</strong></td>
<td>1869</td>
</tr>
</tbody>
</table>

**From Freebase about Emma Goldman:**

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Date of Death</strong></td>
<td>1940-05-14</td>
</tr>
<tr>
<td><strong>Series Written (or Contributed to)</strong></td>
<td><em>Emma Goldman: A Documentary History of the American Years</em></td>
</tr>
<tr>
<td><strong>Country of Nationality</strong></td>
<td>United States of America</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td>Ashkenazi Jews</td>
</tr>
<tr>
<td><strong>Cause of Death</strong></td>
<td>Stroke</td>
</tr>
<tr>
<td><strong>Jewish people</strong></td>
<td></td>
</tr>
</tbody>
</table>
Lessons learned

- Possible to automatically harvest relevant linked data from libraries and other institutions

- Editorial control over the harvested data needs to be better integrated into the note-taking process

- Did not adequately demonstrate the benefits of structured data

- Do not simply aggregate and edit linked data—need to usefully exploit it to researchers’ benefit.
Editors create topics to label and index their notes; later reconciled to external identifiers in a separate batch process.

Old

Editors fluidly create, link to, and reconcile topics within the note-taking process.

New
Motivating structured data use

enabled storing and editing of structured data, but provided no incentive for editors to do this

Old

storing and editing structured data immediately enables sorting and filtering and creating simple visualizations

New
## Women Already Voters: Test Cases during Reconstruction


<table>
<thead>
<tr>
<th>Article</th>
<th>Related Notes &amp; Queries (8)</th>
<th>Related Documents (66)</th>
</tr>
</thead>
</table>

**Filters**

<table>
<thead>
<tr>
<th>+ Item Type</th>
<th>Publication Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Publication Title</td>
<td></td>
</tr>
<tr>
<td>+ Author</td>
<td></td>
</tr>
<tr>
<td>+ Archive</td>
<td></td>
</tr>
<tr>
<td>+ Representations</td>
<td>Revolution (11)</td>
</tr>
<tr>
<td></td>
<td>Woman's Journal (6)</td>
</tr>
<tr>
<td></td>
<td>Chicago Legal News (4)</td>
</tr>
<tr>
<td></td>
<td>Chicago Daily Tribune (3)</td>
</tr>
<tr>
<td></td>
<td>Norwalk Gazette (3)</td>
</tr>
<tr>
<td></td>
<td>Philadelphia Inquirer (3)</td>
</tr>
<tr>
<td></td>
<td>Detroit Tribune (2)</td>
</tr>
<tr>
<td></td>
<td>Albany Law Journal (2)</td>
</tr>
<tr>
<td></td>
<td>New York Times (2)</td>
</tr>
<tr>
<td></td>
<td>Rochester Union &amp; Gazette (1)</td>
</tr>
<tr>
<td></td>
<td>[Philadelphia Press] (1)</td>
</tr>
<tr>
<td></td>
<td>Woman's Journal (1)</td>
</tr>
<tr>
<td></td>
<td>Central Law Journal (1)</td>
</tr>
<tr>
<td></td>
<td>Detroit Advertiser and Tribune (1)</td>
</tr>
<tr>
<td></td>
<td>Santa Cruz Sentinel (1)</td>
</tr>
<tr>
<td></td>
<td>New Northwest (1)</td>
</tr>
<tr>
<td></td>
<td>St. Louis Dispatch (1)</td>
</tr>
</tbody>
</table>
Filter and sort notes not only using the dates of the cited documents (as they currently can), but also using:

- locations and birth and death dates of the people referenced in the notes
- locations and dates of existence of the organizations referenced
- locations and dates of the events referenced
Light and Shadows: Emma Goldman 1910-1916

Search for a place Emma Goldman visited:
Seattle, Washington

May 24, 1910

```
May 24, 1910: 4 related documents
```

May 25, 1910
Emma Goldman lectured on "White Slave Traffic" at Columbia Lodge Hall, Seattle.

```
May 25, 1910: 4 related documents
```

June 9, 1912
Emma Goldman began a week long series of lectures on anarchism and socialism in Seattle. Washington. Seattle Times created furor over anarchism: "Will not some patriots save Seattle from the treasonable utterances of Emma Goldman?" 500 veterans of the Spanish War threatened to meet Emma Goldman at the train station and drive her away. Mayor refused requests to prevent Emma Goldman from speaking, but assigned police to her meetings to prevent an outbreak of trouble.

```
June 9, 1912: 4 related documents
```

June 11, 1912
Emma Goldman scheduled to lecture on "Art and Revolution" at Eiler's Recital Hall, Seattle.

```
June 11, 1912: 1 related document
```

June 12, 1912
Emma Goldman scheduled to lecture on "Chantecler" at Eiler's Recital Hall, Seattle.

```
June 12, 1912: 1 related document
```

June 13, 1912
Emma Goldman scheduled to lecture on "Communism, the Most Practical Basis for Society" at Eiler's Recital Hall, Seattle.

```
June 13, 1912: 1 related document
```

June 14, 1912
Emma Goldman scheduled to lecture on "Sex, the Great Element of Creative Work" at Eiler's Recital Hall, Seattle.

```
June 14, 1912: 1 related document
```

June 15, 1912
Emma Goldman and Ben Reitman scheduled to lecture on *An...".
A note on Dhanvanthi Rama Rau & the Fourth International Conference on Planned Parenthood can become viewable as:

- a **map** of specific locations in Stockholm and Bombay
- a **timeline** of dates associated with the conference
- a **network** of relationships among people and organizations.
Expected benefits

- Working notes become repurposable
- Working notes become more discoverable
- Shift of focus from one-shot product to continuous data curation process
1. Documentary editing
2. Current state of Editors’ Notes
3. Roadmap: exploiting linked data
4. Roadmap: hibernating scholarship
Thank You

We are grateful for funding from:

The Andrew W. Mellon Foundation

Coleman Fung

Michael Buckland  buckland@ischool.berkeley.edu
Ryan Shaw  ryanshaw@unc.edu
Project information  http://ecai.org/mellon2010/
Project site  http://editorsnotes.org/
Source code  https://github.com/editorsnotes