

Jul 7th, 1:30 PM - 3:00 PM

The Big Picture: Conceiving a Digital Edition of Jane Addams' Papers

Cathy Moran Hajo
Ramapo College of New Jersey

Follow this and additional works at: http://repository.brynmawr.edu/greenfield_conference
[Let us know how access to this document benefits you.](#)

Hajo, Cathy Moran, "The Big Picture: Conceiving a Digital Edition of Jane Addams' Papers" (2017). *Women's History in the Digital World*. 8.

http://repository.brynmawr.edu/greenfield_conference/2017/Friday/8

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. http://repository.brynmawr.edu/greenfield_conference/2017/Friday/8

For more information, please contact repository@brynmawr.edu.

The Big Picture: Conceiving a Digital Edition of Jane Addams' Papers

Cathy Moran Hajo, Ph.D.
July 6, 2017



Early in 2014, when I was working at the Margaret Sanger Papers, I was approached by one of our federal funders -- the National Historical Publications and Records Commission, or NHPRC. The Sanger Papers had almost finished its work on our four-volume book edition and the Jane Addams Papers needed a new editor. Was I interested in taking it on?

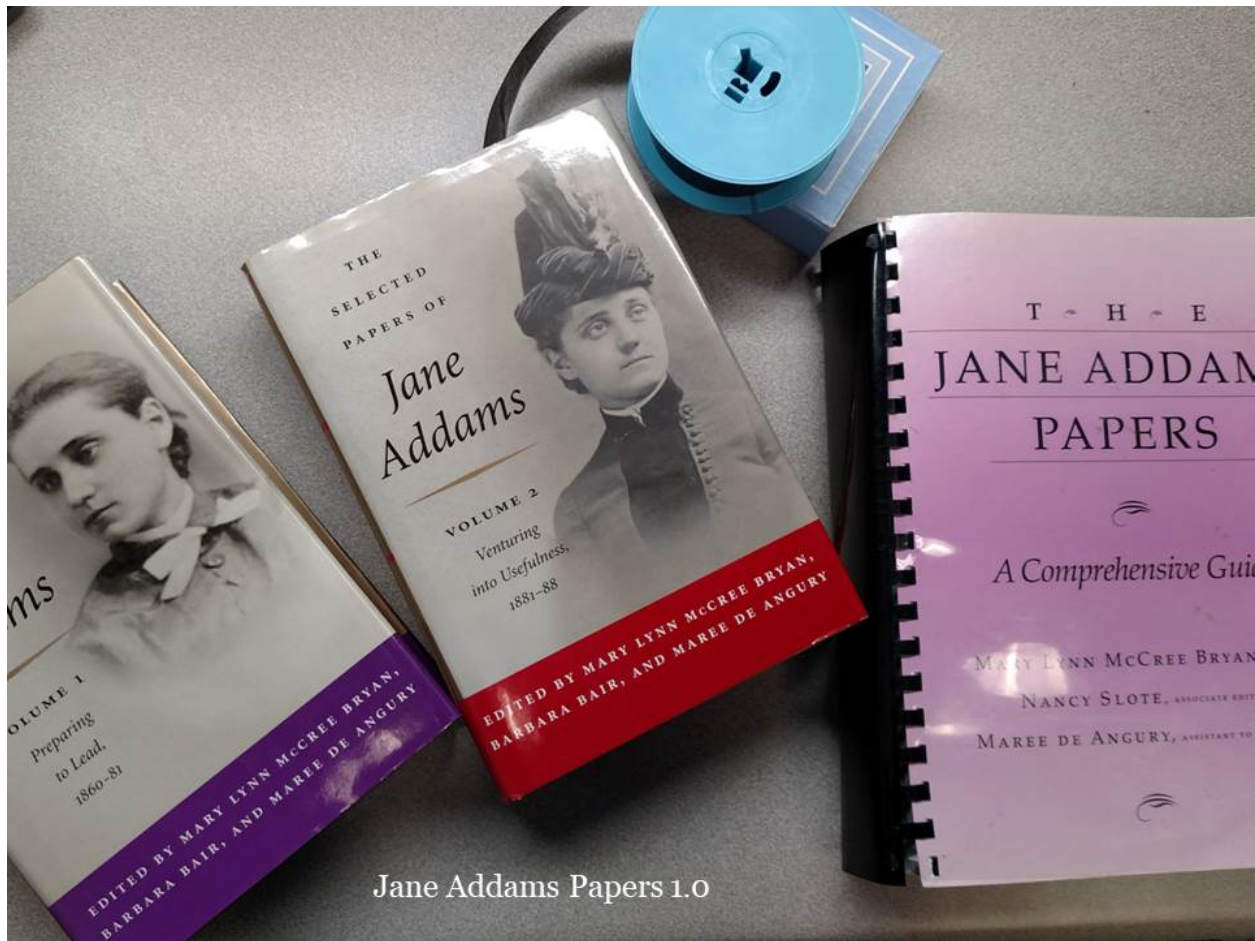
The Jane Addams Papers was, like the Sanger Project, one of a handful of women's editions started in the 1970s and 1980s as microfilm editions.



Jane Addams, born in 1860, was an American social worker, Progressive philosopher and activist for peace, woman suffrage, child labor reform, and social welfare. Addams is best known for co-founding Chicago's Hull-House, a social settlement in 1889. The author of eleven books, hundreds of articles and speeches, Addams won the Nobel Peace Prize in 1931. At her death in 1935, the *New York Times* called her "perhaps, the world's best-known and best-loved woman."¹

The Addams Papers was one of the first women's editions, begun in 1976 by Mary Lynn, almost ten years before the Sanger Project.

¹ "Jane Addams, A Foe of War and Need," *New York Times*, May 22, 1935.



Jane Addams Papers 1.0

After publishing an 83-reel microfilm and a detailed subject index by 1995, Bryan and her editorial team were working on a six-volume print edition of Jane Addams' papers. They published two volumes and were working on a third volume, covering the years 1899-1900. Bryan was ready to retire after completing Volume 3, but had not been able to find a successor.

The NHPRC was primarily interested in ensuring that the book edition would be completed, but from the first, I wanted to create a digital edition. Back in 2009, when I gave the presidential address to the Association for Documentary Editing,² I challenged the editing

² Cathy Moran Hajo, "[Scholarly Editing in a Web 2.0 World](#)," October 16, 2009.

community to reach for broader audiences through digital technology. I believed then, and even more strongly now, that providing free access to as many documents as we could must be our primary goal. That meant revisiting the **microfilm** rather than digitizing the existing books, which only represent about 3% of the papers. And I wanted those documents to reach the **broadest audience possible**, including students both young and old. In many ways, as I thought about whether to take on this project, I thought about that speech and realized that it was time to put my money where my mouth was.

Where would we host it? How would we build it? And how should be envision it?

Institutional Host

I contacted Ramapo College of New Jersey, a small state liberal arts college, about hosting the Addams Papers. It was my alma mater and I had worked closely with them while I worked at the Sanger Papers, supervising Ramapo interns and joining their alumni board. When the Addams opportunity arose, I worked with Dean Stephen Rice of the Salameno School of Humanities and Global Studies to develop a project that worked to meet one of the College's fundamental goals--providing students with hands-on work in digital humanities, providing them with experience working with primary sources, and working in public history. Addams' work in social welfare was also of interest to the College's school of social work, and her international connections offered an opportunity to explore the history of the early 20th century in global terms.



Copyright

As with all projects, one of the first steps was to think through **permissions and copyright**. Nothing could proceed without that. Jane Addams' unpublished writings are in public domain. Check! The major repositories of her papers, the Swarthmore Peace Collection and the University of Illinois at Chicago were enthusiastic. Check! And perhaps most surprisingly, the microfilm publisher, Proquest, agreed as well. Check! We would have to clear the rest of the permissions and copyright, but it was do-able.

Legacy Data

One important question was whether the Addams Papers had created a **database** to control information on the documents, because that would help determine how we would

approach creating a digital edition. **They had not.** In some ways, this made things easier for us, if a bit more time-consuming. We did not have to work with legacy data or try to wrestle it into some more modern system. It meant that we could think creatively about what the Addams digital edition should be and what it should do.

What we had to work with was the microfilm itself and the guide and index.

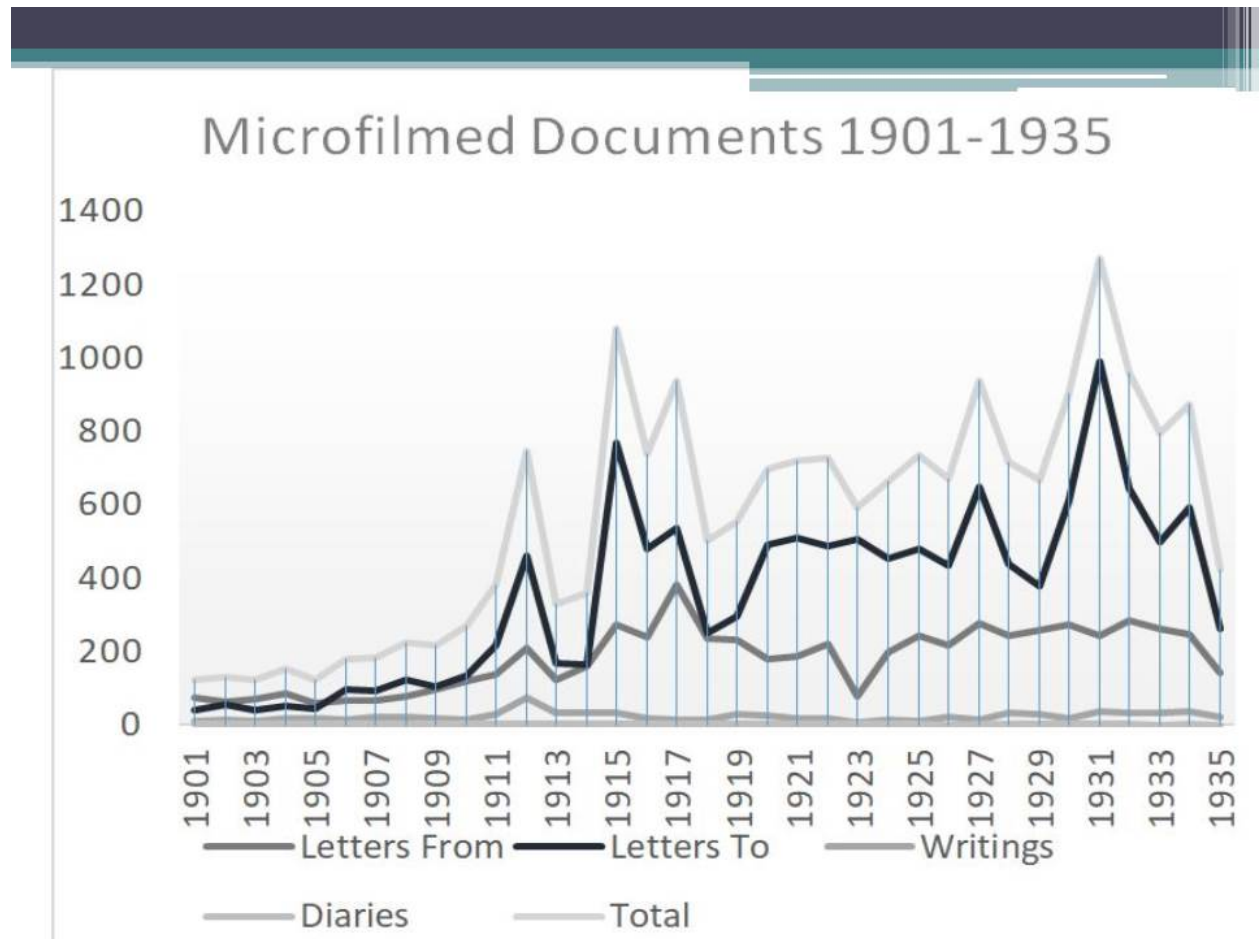
The Jane Addams Microfilm and Guide

The Addams Papers contains correspondence, legal documents, financial records, diaries and calendars, writings, and Addams' reference files. It also includes the records of the Hull-House Association, board minutes, real estates and financial records, records of the many activities and groups that met there, scrapbooks and reports of activities and research conducted there. It also includes a file of newspaper clippings.

The Addams Papers conducted a **search** for documents from 1975-1983, querying over 1,000 archives, libraries, and major newspapers. I knew that we would need to revisit the search, as more archival collections have been deposited and described since 1983, and because the Internet now allows us to locate collections in far-flung archives from our desks. A quick search of archival and newspaper databases pointed to some 75 collections that needed to be searched.

The Addams **microfilm guide** is unique, I think, in that it tried to index subjects. Its editors indexed documents by author and recipient, but also indexed mentions of people. For large topics, like Jane Addams, Hull-House, and the Women's International League for Peace and Freedom, they provided some subject access. It wasn't as detailed as a book index, but it certainly offered some guidance. Strangely, though, the index did not indicate document

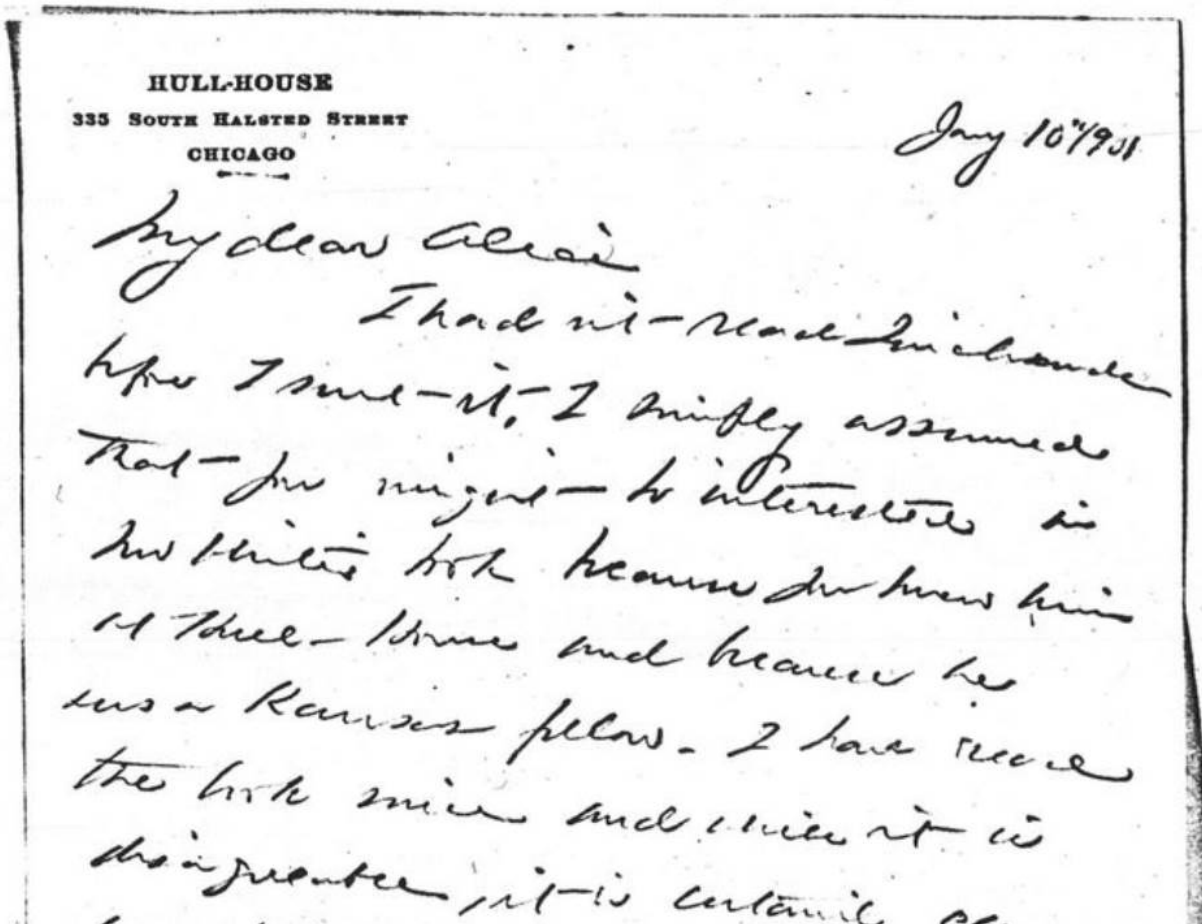
dates, which made it a bit more difficult to use. The guide also included a listing of all correspondence by date, organized by "letters from" and "letters to" Jane Addams. This list does not include any names. This allowed me to, slowly, count the numbers of documents in each year, which helped me get a better sense of the extent of the project.



I decided to focus first on correspondence and writings first. I wanted to work with Addams' diaries as well -- which, sadly, are more like appointment books than reflective notes on her life -- but I moved them to the second tier after seeing how difficult they are to read. They may be more useful to researchers as a spreadsheet and mapped chronology

than as texts. Looking only at correspondence and writings, I counted about 21,000 documents.

But then there was the handwriting!



One look at Addams' hasty scrawl and I knew that simply providing digital images of these documents, even if rich with description, would not make them accessible to a broad audience. Students, the general public, younger people in general, heck, even I had trouble reading some of them! While a fair percentage of the documents on the microfilm were typed, the handwriting would serve as a real barrier to wide use. While it was one thing to

create metadata and images for 21,000 documents, it was another to also provide transcriptions.

Transcribing documents is an essential editorial task. It makes difficult texts legible, and in a digital medium, also makes them searchable. Adding 21,000 transcriptions to our digital edition would add a new dimension to the project and many more hours of work. But it would make the edition much more useful to our audience. We had to think through:

- How we wanted to represent difficult text-- words that we can't read, or words that were crossed out or overwritten.
- How we wanted to deal with misspelled or abbreviated words.
- How we wanted users to search transcriptions.

And it was at this point that we needed to give some hard thought to the technological platform we would use.

Platforms

For a good twenty five years we have been told that scholarly editions had to use XML because it was platform independent and allowed rich content encoding of documents. I had worked on a small digital edition of Margaret Sanger's speeches and articles, which was built using the **Text Encoding Initiative and XML**.³ In that edition we used encoding to track people, organizations, and titles of works mentioned. We used metadata stored in the TEI header to identify the author, date, publication source, and developed a detailed list

³ "Jane Addams, A Foe of War and Need," *New York Times*, May 22, 1935.

of subjects. While TEI and XML are powerful tools, it seemed to me that they were overkill for the basic kinds of encoding we wanted.

Working with TEI with an ever-changing staff of student interns and workers was a challenge. While students took to encoding, their work was difficult to proofread at the level we were accustomed to. We had to ensure that the tags were applied correctly, that the TEI files were valid, and that the metadata was properly constructed. And we had to proofread the transcriptions on the public site in order to correct any errors that crept in after encoding. We did not include images in our edition, as most of the documents were published and not terribly interesting to look at. But the main problems we had with our TEI site were publishing it to the web, as well as upgrading it as guidelines changed. The site looked good in 2003, but by 2015 it looked tired and we did not have the programming skills or funding to improve its appearance. When the TEI released P5, its newest guidelines, we were still stuck in a variant of P4, which meant that none of the new web publishing tools worked with our edition.

Looking at other TEI-based projects, available as digital editions, I found that few of them offered the kind of searchability and flexibility I hoped to offer to readers of the Addams Papers. In fact, since starting the Sanger digital edition in 2003, it did not seem TEI-based editions had advanced very far in how they looked or operated. I was also reluctant to take on a TEI-based project at a college that did not have support for digital humanities work. TEI is a remarkable tool for working with complicated text, but I didn't think that it was the best fit for us.

I wanted to include the images and build rich descriptive metadata, and wanted something that would be stable but easier to work with. I had used Omeka to build a digital archive for teaching, and decided to adopt it for the digital edition.



Omeka, if you have not heard of it, is an open-source digital content management system created by the Roy Rosenzweig Center for History and New Media⁴ and designed for small archives, museums, and other cultural heritage sites. It uses the Dublin Core metadata to store data in a MySQL database. It can export the metadata records in XML and other formats. Omeka was designed to be easy to work with. It uses entry forms to build rich

⁴ For more on Omeka, see <https://omeka.org/>.

metadata and allows projects to customize the entry screens to help insure that data goes in cleanly. Omeka shines in other ways--chiefly its ease in publishing your digital content on the web. After the difficulties of transforming XML for publication on the web, Omeka's simple check box to publish an item seems a wonderful feature. Another was the ability to extend Omeka by using and developing plugins. These are small modules that can be added to customize your content, adding, for example, the ability to map your data by adding the Geolocation plugin, or add text pages to your site by using the Simple Pages plugin.

I spent time looking at other Omeka-built sites and other digital editions to think about what we wanted.⁵ Seeing the kinds of resources that other projects provided helped us think about what we wanted our edition to do. But working with the documents themselves was equally as important.

One thing we lost by selecting Omeka was the ability to encode our transcriptions. Rather than use brackets around an iffy word, we could have encoded it as <unclear> and indicated why in the tag. Or we could have tracked the editorial changes Addams made to a document, indicating where text was deleted and where it was added. I was willing to give up that small advantage. Most of our documents did not have a lot of complexity, and it seemed that we could still show those complexities intellectually using typography.

⁵ For links to Omeka-based sites see <https://omeka.org/showcase/> and <http://omeka.org/codex/Sites Using Omeka>.

Access

One of the things that I and other editors of women's papers have devoted our lives to is making these primary sources available. We had always been limited by the technologies that were available and the perceptions of what was "important" enough. When projects based on the Founding Era were publishing comprehensive book editions of everything their great white man ever touched, women's projects were relegated to the "microfilm ghetto." We were told that we couldn't get funding for book editions that included more than a tight selection of between 1 and 5 percent of what we had put on microfilm. And just as we started completing our 4 or 6 volume sets, we learned that the book was "dead." Then once digital publication of editions began, it was the Founding Fathers' volumes that went up first, leaving women's projects also in a technological ghetto.

Going first meant that the Founders' editions had to figure things out on their own, and they chose to work with publishers—and thus developed a digital edition model that tried to replicate their books in a digital format, employing a paywall to try to recoup costs. Demand for free access resulted in more funding that made transcriptions, but not annotations available for free through the Founders Online version of the Rotunda Founding Era. But as the bar for digital publication has been lowered, projects like the Addams Papers are actually in a good place. We have more options for publishing digital editions, we can include images as well as transcriptions, and we can design our editions for the public and provide free access.

This, for me is the most exciting part of working with the Addams Papers digital edition. Women's papers, hidden in archives or on microfilm for so long, can finally be truly accessible to the public. If we do our jobs right, they will be **more accessible** than some of

the “great men.” And this accessibility comes not just from the fact that the site is open access, it is also in the way that we are designing it—to make the documents both accessible and understandable by a broad audience.

Administering the Project

One of benefits of designing a digital edition system from scratch is that you can build a tool that can also help run the project. When you are adding features to an existing or long-term project, you often end up working in several systems. My goal in using Omeka was to track as many things as possible within the system.

- Copyright and permission management -- We needed to be able to identify documents that need permissions, and to have an easy way to update them as permissions were secured. As we complete a set of documents, we can generate lists of people that need to be cleared and start the process.

Workflow

- **Metadata** (incomplete, complete, proofread, ready to publish, published)
- **Transcription** (pending, complete, proofread, ready to publish, published)
- **Translation** (not needed, pending, completed, proofread, ready to publish, published)
- **Research** (pending, assigned, completed, proofread, ready to publish, published)

Permissions

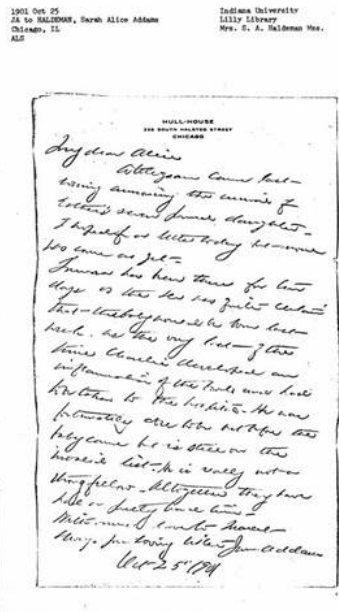
- **Facsimile Permission** – from owner of original (Pending, Cleared, Restricted)
- **Copyright Permission** (pending, cleared, restricted)

-
- Work flow management -- We need to know what steps are completed for each document in the process of writing descriptive metadata, transcribing the text, translating the text, and researching identifications. We want to be able to access documents that need to be proofread, research that needs to be done, and documents that are ready to be published. We also want to track the history of changes to the digital record.
- Counts -- We can get up to the minute information on the number of documents, people, organizations and events. We can see how many are in each stage of work.

We found that one of Omeka's advantages -- its simple one click **publication process** -- quickly became **too** simple for our needs. Our documents were complex, made up of three parts--images, metadata, and transcriptions (and in some cases translations).



Can we publish it?



Copyright? – Yes, Jane Addams is in public domain.

Repository permission? – Yes, the Lilly Library has granted permission.

Metadata? – Yes, it has been proofread.

Transcription? – Yes, it has been proofread.

Translation? – None needed.

Yes, we can publish it!

In addition, we had to track repository permissions and copyright. And we needed to insure that an editor had reviewed the student work, proofreading metadata and transcriptions. When was a document ready to publish?

- We could have waited until everything was cleared and proofread, like we would in a book edition.
- We can publish documents or parts of documents in stages as work is completed.
 - To publish an image, we need both repository permission and copyright.

- To publish a transcription we just need copyright.
- To publish metadata only, we don't need any permission.
- We hold back documents that need more work.
- We can use the varying status fields to organize our work.

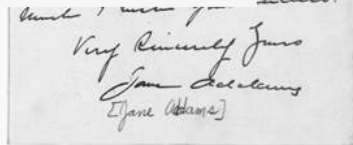
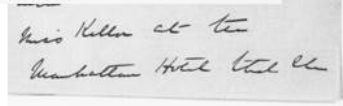
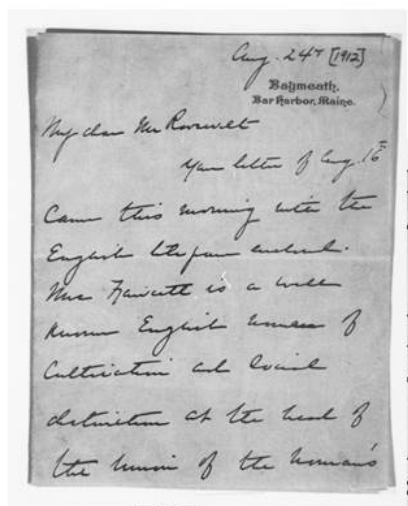
The more fine-grained publication process available in digital publication also opened new questions for us. Editors don't like to admit it when we can't read something, and when it comes to book publication, we might review a document over and over until publication, trying to resolve an illegible word. With digital publication we have an opportunity to publish parts of the edition as soon as they are ready, and we also have the opportunity to correct problems seamlessly. This opens up the question of when a transcription is ready to be published. Should we publish when there are a few illegible words, on the assumption that providing 95% of a transcription is better than none, or should we wait until we have gone over it over and over? I think that we should provide the 95% now and continue improving the transcription as we go. We welcome help from our readers—we want to allow readers to comment, suggest readings of words or interpret what a document is saying in the comment section.

Annotation and Contextualization

One of the things that editions do that separate them from digital archives is that they provide context for the texts, usually in some form of annotation. While we are working on traditional print volumes that will highlight a very select group of documents, we had to think carefully about how we would provide context for readers of the digital edition. With over 21,000 documents, it was not feasible to provide the same kind of intensive

annotation that we will create for the book edition. We don't have the staff or time for that, and it would replicate the work done for print.

In many ways, the metadata that we gather around each document serves as annotation. When we identify dates, authors, and assign subjects, map points, and tags we are helping contextualize the documents and are helping users find documents that are related to their interests.



- Author
- Recipient
- Dates
- Titles
- Type of document
- Repository/Collection
- Locations
- Subjects and Tags
- Names mentioned
- Organizations mentioned
- Events mentioned
- Publications mentioned
- Related documents (drafts, enclosures, etc.)

Part of the process of deciding upon what metadata we want to gather is trying to think about what our different audiences might be looking for. How might a student researcher want to access documents as opposed to a scholar or graduate student? How might our contributing archives want to see the documents they own?

We also needed the capability to add comments to a specific document, explaining issues that only relate to that text. We might note that the enclosures mentioned in a letter were not found, that page 3 is missing, or clarify the dating of a document. All of the other annotations we create are stand-alone, a main entry on the topic that is linked to every mention.

When we thought about what kind of information we could add consistently, we turned to the kinds of things that most historians look for -- **people, places, events, organizations, and publications**. As we were already highlighting those relationships in the documents, it seemed well worth it to provide a short identification of each one. We envisioned building a sort-of web-based encyclopedia of resources about Jane Addams' world which would serve as a resource on its own, as well as a way to navigate the documents.



Louise DeKoven Bowen, ca. 1905.

Name
Birth and death date
Birth place
Biography
Occupation
Tags
Rightsholder
Locations

We gather metadata about these people, creating a rich network of the Progressive Era, including politicians and laborers, women suffrage leaders, peace activists, and Chicago's movers and shakers. And as we enter more and more documents, our network expands and grows as Addams broadened her reach. Many of these people are not readily findable on the web, and our entries provide citations that allow readers to find out more. Addams began her work locally in her neighborhood, then expanded to Chicago, to state-level activism, and then to the national stage. We are just beginning to explore Addams' international connections. **Is it a lot of work? Yes, no doubt**, but it is work that helps researchers whether they are interested in Jane Addams or not, and it provides context for

our texts. Readers decide whether or not they want to follow the links to learn more about those described in the documents.

One thing that many digital editions lack is subject access. Many people think that if you can search the document text that you don't need subjects. While text searching provides impressive results, it is not the same as an interpretive subject index. An example might be the category "Jane Addams, and family." Addams rarely uses the word "family" in the letters she writes, she might use "sister," "brother," or "nephew," but more often not. Editors group documents in which Addams discusses her family or corresponds with them in a way that explores her relations with them.

Subject indexing is an intellectual process by which we analyze document content and create a system of knowledge to organize it. We expect to see it in a book, but not necessarily in a digital edition. With the increased numbers of documents we can fit into a digital edition, subject searching becomes essential-- it helps readers cut through the forest to the trees that they are interested in. We built a list of subject terms to be used in the edition by consulting other editions, histories of the Progressive Era, and Addams biographies. We do occasionally add new subjects, but usually do a search to locate any previously entered documents that might have fit it.

We only track subjects for the document texts, not for the biographies, organizations, events and publications. But we think that some readers might want subject access to those entries. We decided to use tags as a broader level of subject access, but one that applied both to documents and identifications. So if you search for "Medicine" you will get documents that discuss medicine, health care, sicknesses, and medication as well as entries

about doctors, medical conferences, and events. We think that this quick sorting of digital items will appeal to people who are exploring themes covered in the edition.

Reaching New Audiences



The purpose of preparing all this metadata, digitizing images, transcribing, and research is to reach a broad audience. Many editors see their primary audience as a small community of scholars, but I think that one of the greatest advantages of digital publication is that we can reach many, many more people. So the Jane Addams Papers is geared towards serving that broad audience, consisting of school kids working on History Day projects, teachers who want to create primary-source teaching materials, high school and college students

working on papers, genealogists and individuals trying to trace their family history, and those working for social change who are interested in their history. Scholars and digital humanists too, of course! But if we are honest, we know that scholars will find us and know how to read these documents-- it is our more general audience who needs an easy to navigate and flexible website.

We want to explore using digital tools to analyze, study, and present Addams' life and work, and to enable others to create their own digital work based on our project. We want to be able to export the data we are gathering on Addams' networks--tracking participation in events, organizations, and documents-- so that interested scholars can build social network diagrams and conduct research on the changes in Addams' network over time. How interesting would it be to see what happened to Addams' contacts during World War I, when she was considered a pariah for opposing the war? Do those who cut her off then return by 1931 when she was awarded the Nobel Prize? We can look at the historical figures that Addams discusses in her writings, or analyze the gender of her associates to see whether she worked largely in a community of women. We can map her comings and goings and determine how much (or little) time she was spending at Hull-House in her later years.

Scholars may want to run Addams' writings through text analysis software to look for changes in rhetoric, or compare her writings to male Progressives to see differences in approach on similar topics. To enable these kinds of collaboration we need to provide an easy way to export our data for scholarly use in other platforms.

We also want to reach out to the public and invite them to join our site, and ask questions or make comments on the digital edition. We want to build a crowdsourcing portal where people can transcribe documents or create metadata. We are thinking about experimenting with allowing our readers to rate documents, to enable a new sorting option – highlighting the "best" documents on Jane Addams' views on peace -- which might help younger and more casual users to find the most relevant documents quickly. Relying on crowdsourcing to replace student or editorial work isn't our goal, and I don't think that there is any way to guarantee participation, but there are sites that have attracted great interest. If we can build a committed group of volunteers, we may be able to start digitizing other sections of the microfilm, like Addams's early letters, the Hull-House records, or her diaries.

We are exploring the idea of creating spaces for scholars, students, and teachers for their own research on Addams. They could save searches of documents, comment on them, and repurpose them for digital exhibits. A teacher might present a series of documents with questions for students to answer; a high school student might create an exhibit on an event or topic, bringing in their own illustrations and writing their analysis. A scholar might create an exhibit comparing Addams to another figure, or embedding digital maps, timelines, and visualizations that they created using the digital edition. We would want these kinds of research to be a part of the digital edition in a separate navigational space, but think that seeing how others use and interpret the documents will be a great help to our readers. Hosting their creations will make our edition a richer place.

Just last week we received funding from the New Jersey Council for the Humanities to explore additional outreach options. We will work with teacher's education students at

Ramapo College to create primary source-based lesson plans, and write guides to help middle- and high school students develop National History Day projects. This national program encourages students to use primary sources for their interpretations, and the competition includes presentations at the regional, state and national level in categories ranging from individual and group papers, exhibits, websites, and performances. Each year they focus on broad themes like this year's "Taking a Stand in History."

A New Kind of Edition

With the Jane Addams Digital Edition, we are rethinking our roles as scholarly editors. Publishing in digital media had changed what editions can be and who uses them. Free digital publication of historical documents has already expanded our audience in ways that we could not always anticipate. We are now one Google search away from the 3.6 billion people with access to the World Wide Web and that is a heady responsibility.⁶ Rather than continue down a narrow path designed for scholarly book publication, we need to open up the edition and share what we do with the world.

Thank you!

⁶ Internet Live Stats (<http://www.internetlivestats.com/internet-users/>) accessed on June 28, 2017.