The Margaret Sanger Papers Project

A Documentary Edition in the Digital Age

Cathy Moran Hajo and Esther Katz
sanger.papers@nyu.edu

The World Wide Web is fast becoming the first place researchers and the curious look when they need to find information quickly. For those with access, the Internet can open the door to a vast storehouse of knowledge. However, for the unwary, the Internet can lead to a maze of poorly-rendered and untrustworthy source material. In the field of history, this most often takes the form of transcriptions of historical documents (or portions of documents) loaded onto the Web without any indication of who has done the transcription, whether it has been proofread or changed, or if it has been excerpted. Sometimes the document is taken out of its original context.

In the case of research on Margaret Sanger, the founder of the American birth control move-

Esther Katz, an Associate Professor of History (ADJ), is the editor and director of the Margaret Sanger Papers Project. Cathy Moran Hajo, a doctoral candidate, is the assistant editor and assistant director.

ment, the search for accurate and reliable source material is further complicated by that fact that anti-abortion groups have taken to using Sanger and her words to tarnish the reputation of the current reproductive rights movement. This is generally done by taking selected Sanger statements (often just a few sentences from a multi-page document) and placing them on the Web as evidence of Sanger's racism, anti-Semitism, pro-abortion stance, or some other polemical point the group mounting the website wishes to make. As a historical editing project striving to provide access to Sanger's words, thoughts and deeds by accurately reproducing Sanger's writings and speeches, the Margaret Sanger Papers Project has undertaken the production of an electronic edition that lives up to the same scholarly standards used for editions of historical documents in print and on microfilm.

The Margaret Sanger Papers Project (**www.nyu.edu/projects/sanger**), has been located at New York University's History

Department since 1987. Supported by grants from the National Historical Publications and Records Commission, the National Endowment for the Humanities and many private foundations, the Project has already published a two-part, 101-reel microfilm edition, containing over 54,000 documents. We are also preparing a printed guide and item-level index that will integrate our two parts with a third one: the 145 reels of additional Sanger material housed on microfilm at the Library of Congress. All three series are currently available to NYU researchers at Bobst Library. (For a list of other libraries which have purchased the microfilm, see the Margaret Sanger Papers website.) The Project is now working on the first volume of a three-volume book edition of selected, transcribed and annotated Sanger documents to be published by Indiana University Press.

Published editions of historical documents have for decades provided access to rare and fragile historical material. Print editions, such as the multi-volume Papers of Woodrow Wilson or Papers of Thomas Jefferson, offer carefully transcribed and annotated documents in editions that are available in libraries and schools across the country. More recently, historical editions have also been available on microfilm, such as The Papers of Susan B. Anthony and Elizabeth Cady Stanton (45 reels). Such edi-

Spring 1998

tions allow many more users to have direct access to originals than was previously possible, though in a contextualized format with each document carefully identified (by author, recipient, date, etc.). The emergence of new digital technologies now offers even more possibilities for publishing historical editions.

Three models for digital editions have been identified: image editions, live-text editions and combined editions. All three editions provide users with retrieval and search tools as well as supplementary contextual material; the difference among them is in how the historical documents are presented. Image editions present original documents as facsimiles, so that the experience is much like viewing the original manuscript. Live-text editions present transcriptions of original documents whose text can be searched by the user. Combined editions offer both images and live-text transcriptions, so that users can search the text but also have access to the format of the original.

Because there are so many editions already in print, a fourth model, the transitional edition, has been identified as a means of providing access to already published volumes. Printed transcriptions can be searched via online indexes.

As useful as the Sanger microfilm and book editions are, we began to consider the option of publishing on the Internet several years ago. Our initial idea was to digitize all three Sanger microfilm series — over 115,000 documents. In so doing, readers would easily be able to view documents across the 246 microfilm reels, bypassing the fixed format

of the three-series film. After scanning and digitizing the microfilm, our plan was to use the Project's already existing database program to link the documents to their identifying information.

Digitizing the entire microfilm proved to be prohibitively expensive, however, as each frame needed special handling to insure a clearly replicated digital image. More importantly, we realized that digitizing the existing microfilm edition would offer users faster and more convenient access to the documents, but not much else. We decided that if we were going to incur the expense of digitizing historical documents, we would want to take advantage of all the capabilities that the new medium had to offer.

We decided not to try to imitate our existing editions. Rather, we would create a wholly new type of historical edition. We would produce a series of small, topic-based electronic editions, each consisting of approximately 200 to 400 documents and each focusing on a major theme in Margaret Sanger's life.

Each edition contains images of the actual documents, surrounded by contextual information. The look of the electronic edition mirrors that of the microfilm: above each document is a descriptive "target" which identifies the author, recipient, document date, the location from which it was written, and the repository that holds the original. In addition, we added a list of individuals and organizations mentioned in each document.

However, while the identifying targets on our microfilm edition are static (readers must

consult the printed index to locate specific documents on the film), the electronic targets are active, allowing the reader to search them for related documents as well as supplementary information we provided. Users can now easily search the documents in a variety of ways: by the name of individual authors or recipients, by document type, by date, etc. For example, when users click on the name of the author of the document, they can access a short biography of the individual that includes the nature of their relationship with Margaret Sanger. The user can also obtain a list of all documents in the edition written by, written to or mentioning that individual. When a user clicks on the name of a law, the text of that law is displayed; when they click on the date, they can access a day-to-day chronology of what Sanger was doing and where she was at the time the document was written.

Documents in the electronic edition are not arranged in a fixed format. Instead, users may access them in any number of ways. Providing documents in this unstructured format allows readers to approach the material from a variety of perspectives. One can follow Sanger's development by looking at the material in chronological order, trace her relationship with a single individual by accessing documents by authors and recipients, or highlight the different positions she held in public and private by contrasting her published writings with her personal letters. The important difference here is that the reader, not the editor, is determining how the edition will be read. The flexibility and fluidity allow readers to make the

intellectual connections for themselves.

In 1990, we took our plan and joined the Model Editions Partnership (**mep.cla.sc.edu**), a consortium of historical editors and technical experts, funded by the National Historical Publications and Records Commission (**www.nara.gov/nara/nhprc**) to devise models for converting historical editions to electronic form. The Sanger model differed from most of the other samples in the Partnership because it was image-based rather than text-based, and because we included a good amount of supplementary material that had not appeared in our published editions, such as biographical essays on individuals mentioned in the documents.

The 200-page sample we prepared for the Model Editions Partnership documents Margaret Sanger's 1914 publication of *The Woman Rebel*, a feminist journal advocating birth control and radical reform. The model prepared for the Partnership will be expanded by the Sanger Project to constitute the first in its electronic edition series. *Margaret Sanger and the Founding of the Birth Control Movement* will not only include documents dealing with *The Woman Rebel*, but will continue the story to cover Sanger's founding of the Brownsville clinic in 1916 and the formation of a national movement by 1921.

From the start, the Partnership has been committed to creating digital editions in a standardized format that would be independent of commercial software products. This would ensure that the data would be transferable, should delivery systems or software change. For electronic text,

this meant using the Standard Generalized Markup Language, or SGML. Because we were dealing with historical documents, the Partnership decided to create a subset of SGML designed specifically for scholarly humanities texts.

The tagging structure of the Model Editions Partnership has been designed to use both descriptive and formatting tags that specifically address the needs and problems of historical editors. With formatting tags, editors identify the framework of the document (document headers, signatures, and postscripts) and signal the ways in which text is physically rendered (strikeouts, interlineations, and margin notes). Descriptive tags are used to classify words found in the documents. For example, different tags would be used for different types of names, such as names of people versus names of places. This will allow users to search for the person "Washington" without having to wade through every instance in the edition in which a city, street or business of that name is mentioned.

Because the Sanger sample uses images rather that transcriptions of the documents, our problems were somewhat different from those of the other partner projects. In the Sanger edition, the "text" to be tagged is not the actual words written by Sanger and her contemporaries, but the editorial material we have created to identify and contextualize the edition by the Project editors. Thus, we did not need many of the elaborate textual tags used by other editions, such as those used to render interlineations or changes in a document

through various drafts. As a result, the Sanger sample was initially seen as the simplest of the model editions. Compared to the other editions, it contained very little actual text to be tagged – just the targets, biographical essays and topic essays – since our edition relied upon links between documents to provide context. However, in the actual work of tagging and testing the samples, our sample has turned out to be the most complicated of the seven. Most of the problems came about because the software currently available to display SGML text is not as powerful at linking discrete documents as HTML browsers are. Thus, over 20,000 tags were needed to describe and link the documents together.

The experience of creating an electronic edition has changed the way we think about publishing historical material. By exploring the relationships between the documents in our edition, highlighting the individuals mentioned in each document, and creating links between them and the documents, we have come to a deeper understanding of Margaret Sanger's world in 1914, the network of individuals with whom she associated, the development of her political thought, and the intersection of her public and private life. We believe that researchers using this electronic edition will be able to use these primary sources in a much more fluid manner, which will encourage more original thinking and interpretation.

We continue to refine the Sanger Project's model edition. Once it is completed, a sample will be available on the Project's website at **www.nyu.edu/ projects/sanger**. 💾