A RESEARCH REVOLUTION IN THE MAKING GOOGLE BOOKS AND MORE AS SOURCES FOR WOMEN'S HISTORY

by Phyllis Holman Weisbard

If you follow developments in libraries, publishing, or Googleland, you have probably heard about the Google Books project. Google now has agreements with several major libraries in the U.S. and abroad to digitize vast quantities of their holdings. Harvard, Stanford, Oxford, the University of Michigan, and the New York Public Library are charter "library partners" in the enterprise, and as of this writing they have since been joined by the University of Wisconsin-Madison (including material from the Wisconsin Historical Society), the Midwest universities consortium known as the Committee on Institutional Cooperation ("CIC"), which also includes UW-Madison, the University of California, the University of Virginia, the University of Texas, Princeton University, the Bavarian State Library (Germany), Ghent University Library (Belgium), University of Lausanne (Switzerland) and two libraries in Spain. A current list of partners and links to their involvement in the project is at http://books.google.com/googlebooks/partners.html. The project, formerly known as Google Print, was announced in December 2004 and is well under way; thousands of books are already in the database at http://books.google.com/.² Google Books and other mass digitization projects have the potential to revolutionize research methods and results. With what's already available, Google Books can now greatly enhance student and scholarly quests.

Before illustrating ways that Google Books can be used for research in women's history, there is a bit more background to consider.

The Google Books project has set off a considerable stir among authors and publishers of works still in copyright who consider the digitization of their works without their permission to be a breach of copyright. The Authors Guild and others sued Google in September 2005, and final ruling awaits court action. Google maintains that because it only displays brief "snippets" from a copyrighted work unless there's an agreement with the publisher to display more, this usage is within the "fair use" guidelines of copyright law. Google also gives a second digital copy of each work to the lending library. The libraries generally plan to keep their digital copies of copyrighted works in "dark archives," although libraries' interpretations of what's in the public domain or falls within fair-use guidelines may differ somewhat from Google's.

Google Books has numerous publisher partners; in fact, many publishers signed up before the library program was announced. These publishers control how much of the

content of each title they want displayed, from 20% to 100% of the text. Google Books labels any book with less than 100% displayed as "limited preview." The entry for the book then includes the statement "Pages displayed by permission," and Google displays links to the publisher.

Books that are out of copyright are labeled "full view"; and indeed, each and every page of these can be viewed. In most cases, these books can also be downloaded as PDFs, from which pages can then be printed. Currently, "copy and paste" of text in the PDFs does not appear to be possible (passages can, however, be copied as images); nor can optical character recognition be applied, even when the downloaded PDFs are opened in Adobe Acrobat Professional. This is a feature that library partners might choose to change in the presentation of their digital copies. (The University of Michigan Library catalog is already doing so. For books that have been digitized by Google, there are links to two e-versions: Google Books and "M-Books," for the University of Michigan copy. M-Books can be displayed three ways: as images, as PDFs, or as raw, uncorrected text, labeled "full-text." "Copy and paste" will work from the "full-text" display (although users will want to compare the text to the PDF, because there are errors in character recognition).

Google's is not the only wholesale digitization effort going on. Microsoft, Yahoo, and numerous university presses and other publishers teamed up in Fall 2005 to form the Open Content Alliance (OCA; see http://www.opencontentalliance.org/). The OCA intends to build a multilingual and multimedia collection. The collection policy states that the OCA "will initially concentrate on digitally reformatted monographs and serials which represent diverse times, regions and subjects which are in the public domain or available under a Creative Commons license." OCA books will in the future be accessible directly through the OCA website and indexed through Yahoo, but are best browsed or searched at present through the "texts" section of the Internet Archive site at http://www.archive.org/details/ texts. There is also a demonstration selection of books at http://www.openlibrary.org/.

The Open Content Alliance offers what appears to be higher-quality scanning than Google's and numerous display options, including simple text (which is the best type to view when you want to copy and paste) as well as PDF and DjVu (a high-resolution, high-compression program). OCA also "welcomes all efforts to create and offer tools (including finding aids, catalogs, and indexes) that will enhance the usability of the materials in the archive." OCA plans to allow its books to be indexed by other search engines, including Google, whereas thus far Google Books are searchable only within Google. Having all accessible digitized books indexed in one search engine would be very advantageous to researchers, especially if they are searchable in a subset limited to books.

Thousands of books have been digitized by other projects at universities and research institutions, such as the Making of American Project at the University of Michigan (10,000 nineteenth-century books) and Wright American Fiction 1851–1875 (2,887 volumes) at Indiana University, but it is often difficult to find the digitized version of a title within a

general search engine, since the search engine also picks up every mention of that title on Web pages. There are projects that attempt to index available e-books. The Digital Book Index (DBI) at http://www.digitalbookindex.org/ has kept up reasonably well with existing projects — as of this writing it has indexed 130,000 books, of which 90,000 are freely accessible,⁵ and the Online Books Page (http://digital.library.upenn.edu/books/)⁶ links to 25,000 free e-books — but these figures will be dwarfed by the Google and OCA projects, and it is hard to see how either DBI or the Online Books Page will be able to keep up.

Since Google Books is further along than the Open Content Alliance project and has captured the imagination by the sheer breadth of its undertaking, my examples will mainly come from Google. Here are some ways to use Google Books, with examples coming from topics in women's history.

Accessing known books that were published before 1924. It's worth checking Google Books if you wish to consult *any* book you are already aware of that was published before 1924 (copyright cut-off date). For example, you are an undergraduate student at the University of Wisconsin. You're working, in your dorm room on the weekend, on a paper that's due Monday on the history of the term "feminism." You find a University of Wisconsin–Madison Libraries' catalog record for What Women Want: An Interpretation of the Feminist Movement, by Beatrice Forbes-Robertson Hale (New York: Frederick A. Stokes Company, 1914), and you want to check the definition of "feminism" put forward by Hale. The catalog shows two holdings, one in the Cairns Collection, a non-circulating collection of works by American women writers through 1930 held in the Special Collections Department of Memorial Library, and a microfilm version in the History of Women Collection (reel 819, no. 6595). Is it worth your time to tromp to the library? Once you get there, you'll find out that Special Collections is only open during the daytime on weekdays, so you won't be able to look at the print copy until Monday (your due date). And although you could consult the microfilm, you've never yet used any microfilm resource, and the prospect is intimidating. So, instead, you visit Google Book Search Advanced, where you type "What Women Want" in the title field, and up pops the entire text of this book, whose first line is already most of what you need: "Feminism is that part of the progress of democratic freedom which applies to women" (p.3).

Better access to the contents of books. Because Google indexes all the words in each book, you can find material in new ways. For example, there is a useful biographical dictionary of prominent women of the nineteenth century, edited by Frances Willard and Mary Livermore, that first appeared in 1893 and has been reprinted many times and under slightly variant titles. Google Books has thus far digitized the 1897 edition, called *American Women: Fifteen Hundred Biographies With Over 1,400 Portraits* (New York: Mast, Crowell and Kirkpatrick). In none of the editions, including the most recent (*Great American Women of the Nineteenth Century: A Biographical Encyclopedia* [Amherst, NY: Humanities Books, 2005]), is there an index by *place*. If one wanted to find all the women who have lived in

or been associated with Wisconsin it would be a tedious task. A search for "Wisconsin OR Wis. OR Wisc." within the digitized Google book turns up thirty-five hits, including Osia Joslyn Hiles, an advocate for Indian rights who was the first secretary of the Wisconsin Indian Association; Sarah Dyer Hobart, a poet; and Ellen A. Dayton Blair, a temperance speaker. Even more significantly, Google Books performs a major city search automatically throughout the full-view book, then "Google-maps" the cities mentioned and lists the book pages on which that city's name appears. Since Milwaukee is one of the major cities, the city display adds several other pages that are relevant for someone interested in Wisconsin women, where the text includes "Milwaukee" but not "Wisconsin." One of the interesting women whose biography can be found this way is Ellen Palmer Allerton, who wrote for Milwaukee and Chicago newspapers and was at one time the book review editor for the Milwaukee Sentinel.

Here's another example: You are a scholar of gender and disasters. You want to understand the gendered nature of women's experiences during the San Francisco earthquake and fire of 1906 and their aftermath. Must you scour all 721 entries in WorldCat for works about this disaster for mention of women? Can Google Books short-circuit the process? Absolutely! For this quest, a more complex search statement helps: "San Francisco' (woman OR women OR lady OR ladies) (fire OR earthquake)," and a click on "full view books." (For Boolean aficionados: Google searches support OR'ing, if OR is put in capital letters; Google automatically ANDs; hence there are no ANDs between the grouped OR'd items.) There are 618 hits, of which 436 are available in "full view." This may seem only somewhat better than the 721 figure from WorldCat, number for number, but the Google result is much more precise, since it retrieves the books and zeroes in on the pages within that meet the criteria. One of the first entries is for The History of the San Francisco Disaster and Mount Vesuvius Horror, by Charles Eugene Banks and Opie Percival Read (Chicago, 1906). Google takes us to page 116, where the authors quote the dispatch sent by reporter Helen Dare to her paper, the Chicago American, describing her personal experience during the quake. Banks and Read call it "the most graphic recital of an eye-witness to the destruction of San Francisco." Dare seems particularly attentive to women she observes in various states of undress, carrying babies, dragging a sewing machine, etc. ("A drawer of it falls out and they halt to gather up the precious scattered spools. Poor little seamstress, this is her all now" [p.121]).

For this topic, a search of what's available currently via the Internet Archive is also productive. Alice Gerstle Levison's oral history, *Family Reminiscences* (conducted in 1967 by Ruth Teiser for the Bancroft Library, University of California, Berkeley), shows up because Ms. Gerstle Levison experienced the San Francisco earthquake and discussed it in her interview, and the cataloging record includes subject headings for this. (It should be noted that a search of "regular Google" also turns up this oral history transcript — it just doesn't appear in Google Books because it hasn't been digitized specifically for that project.)

Using materials that are not readily available on your campus. Let's say you are a historian of women's health. You teach at a small college that lacks a significant history of

medicine collection and doesn't have the perceived demand for or the funding to acquire databases such as *Eighteenth Century Collections Online*. You would like to examine some eighteenth-century tomes, such as *A Physical View of Man and Woman in a State of Marriage: With Anatomical Engravings*, by M. de Lignac (2 v., 1798), which, for example, the Ebling Library for the Health Sciences at the University of Wisconsin–Madison keeps in a "historical vault" and does not circulate. You are on a first-name basis with your interlibrary loan librarian, but even she can't spring books out of non-circulating collections elsewhere. You try Google Books and find that the New York Public Library's copy has been digitized. You can read the volumes through, search them for instances of particular terms, or go directly to many of the engravings, which Google presents in a "selected pages" area from the opening screen. (The OCLC/WorldCat record and Google Books note that the work was "[p]rinted on the Continent; most of the copies imported were seized by the Customs owing to the alleged indecency of the plates.")

Finding useful passages where you wouldn't otherwise think to look. This is the most exciting aspect of mass digitization with full-text indexing. You retrieve relevant material in contexts you would otherwise not run across without considerable effort. Only serendipity compares, but one can't rely on chance encounters to support research.

Example: You are researching Emmeline Pankhurst and are interested in contemporary impressions of her efforts during World War I. Are there works that do not center on her sufficiently to come up under a subject heading for her, yet which include relevant information? What can Google Books offer for this research? Here are two items that turn up in a "Pankhurst" search in Google Books that would be hard to discover by systematically searching library catalogs or following bibliographic citations:

- (1) You know that Mrs. Pankhurst visited Russia, intending to rally women to organize and to keep Russia in the war. But which of the 2,159 entries in WorldCat with the subject heading "World War, 1914–1918 Soviet Union" should you look at for mention of Pankhurst? The answer from Google Books: *Runaway Russia*, by Florence MacLeod Harper (New York: Century Co., 1918). Harper writes: "I do not know what good was accomplished in the end, but I was very sorry that Mrs. Pankhurst had come to Russia, because she was foredoomed to failure... Her motive was good; she was sincere; but unfortunately the women of Russia were too busy revolutionizing to bother being organized" (p.162). Over the next several pages, Harper goes on to describe a Pankhurst meeting she attended that in her view went nowhere, and she speculates as to why Mrs. Pankhurst made no headway with Russian women. It is unlikely that a researcher would be led to the account in *Runaway Russia* by following footnotes in other scholarship, either. Harper and *Runaway Russia* are not referenced, for example, in June Purvis's recent thorough biography of Pankhurst (*Emmeline Pankhurst: A Biography* [New York: Routledge, 2002]) or in Paula Bartley's more concise study (*Emmeline Pankhurst* [New York: Routledge, 2002]).
- (2) Short Rations: An American Woman in Germany, 1915...1916, by Madeleine Zabriskie Doty (New York: The Century Company, 1917). Its title doesn't hint that Doty

includes a chapter on "London and the Suffragettes," about a visit she made from Germany to London in the summer of 1915. During her stay she observed with distaste a patriotic women's procession organized by Mrs. Pankhurst, who in her view had abandoned the cause of women's rights by urging women to take over the jobs of men who were away fighting, no matter the wages offered, in order to support the war effort. By contrast, Doty participated in a subsequent women's march led by Emmeline's daughter Sylvia, calling for fair wages for women workers. She noticed that Sylvia walked by her mother and sister without speaking to them. Although a diligent researcher might eventually find Doty's book by examining all titles that share the subject heading "World War, 1914–1918 — Personal narratives," there are more than 2,300 in WorldCat, so it would be quite a daunting task.

"Search inside" and more. Thus far I've focused on full-text results, but much can be gleaned as well from works available only in excerpts. In these cases, researchers discover books that are worth exploring further, either through purchase or in libraries. For this topic, I'll leave Google for awhile — even though "snippets" and "limited views" in Google Books apply — and turn to another product: Amazon.com's "Search Inside" feature. For each "Search Inside" book, with the agreement of the publisher, Amazon displays excerpts from the book and applies various word analyses. For this example, you are interested in doing a comparative study of women's autobiography as a source of history. You would like to include the memoirs of Glückel of Hameln, a seventeenth- and eighteenth-century German Jewish businesswoman, and you want to identify some recent scholarship on her. You find Natalie Zemon Davis's study, Women on the Margins: Three Seventeenth-Century Lives (Harvard University Press, 1995), through your library catalog, since Glückel is one of Davis's three women and the cataloguer has provided a subject heading for her. You wonder if there is anything since then. Among the results from Amazon searches for Glückel (using variant spellings Glueckel or Gluckel or Glikl) are three academic books that can be "searched inside." They are Being For Myself Alone: Origins of Jewish Autobiography, by Marcus Moseley (Stanford University Press, 2006), Autobiographical Jews: Essays in Jewish Self-Fashioning, by Michael Stanislawski (University of Washington Press, 2004), and Marion Kaplan's Jewish Daily Life in Germany, 1618–1945 (Oxford University Press, 2005). Searching inside the first two, you readily see from the tables of contents that Glückel is a chapter topic, so you'll want to obtain these titles. And you can be tantalized further by reading some of these books' pages that deal with her, but sooner or later you will bump up against the end of the allowed excerpts. For Jewish Daily Life in Germany, you don't find Glückel among the chapters or "key phrases" of people, places, and topics that Amazon notes in the book, but when you use the word search within the book, you are given thirty-six quotations in which Glückel (or, in this case, with the spelling Glikl) is mentioned, and you realize that this book will also be useful to you.

If you decide to pursue the uses of Glückel's memoir by earlier historians before the advent of attention to women's history, Google Books will have much to offer in "full view," including the full text of Solomon Schechter's 1908 essay, "The memoirs of a Jewess of the

Google Books and More

seventeenth century [Frau Glückel von Hameln]" in his *Studies in Judaism* (Philadelphia: Jewish Publication Society). There is a subject heading for Glückel for this book in library catalogs, so having it available in Google Books is more convenient than essential. But the power of Google Books to unearth material that is not addressed in subject headings can be seen in another hit for this search: Joseph Jacobs' *Jewish Contribution to Civilizations: An Estimate* (Conat Press, 1920), p.263. The context is that Jacobs is taking issue with Werner Sombart's theory that Jews are responsible for modern capitalism, calling it exaggerated. In his *The Jews and Modern Capitalism*, Sombart had discussed Glückel as part of his argument. Jacobs retorts that Christian businesspeople were similar, saying, "We might easily parallel Glückel of Hameln by Thackeray's picture of old Mrs. Newcome, the head of the bankinghouse and the support of evangelical missions and charity." Curious about the Sombart passage? No problem — it is easy to check, as both the German original (1911) and the English translation (E.P. Dutton, 1913) have already been digitized.

We are just at the dawn of mass digitization of the human record. Within the memory of current scholars, bibliographic searching jumped from tedious look-ups of single concepts in print indexes and catalogs to electronic databases that allowed for coordinated retrieval of instances that matched numerous concepts at once, and accessing full-text articles became the norm. Mass digitization of books is another giant leap whose promise is only now beginning to be glimpsed.

Notes

- 1. For details about the CIC agreement announced on June 6, 2007, see http://www.cic.uiuc.edu/programs/CenterForLibraryInitiatives/Archive/PressRelease/LibraryDigitization/index.shtml.
- 2. Results from Google Books also appear within the standard, universal Google search, but mixed in within various other search results from websites, news sources, etc. A search for content within books digitized by Google is better performed within the part of Google specific to the Google Books project.
- 3. Webopedia.com definition: "An archive that cannot be accessed by any users. Access to the data is either limited to a set few individuals or completely restricted to all. The purpose of a dark archive is to function as a repository for information that can be used as a failsafe during disaster recovery." See http://www.webopedia.com/TERM/D/dark_archive.html

- 4. "Creative Commons" is explained on the website of that organization (http://creativecommons.org/) as providing "free tools that let authors, scientists, artists, and educators easily mark their creative work with the freedoms they want it to carry. You can use CC to change your copyright terms from 'All Rights Reserved' to 'Some Rights Reserved."
- 5. Digital Book Index (http://www.digitalbookindex.org/_publishers/publisherinfo01. htm, visited March 23, 2007) "lists all digital or electronic books distributed by both print or electronic publishers, including Adult, Professional, Business, Technical, Reference, and Children's books. It does *not* generally list databases, theses, dissertations, magazines, or other unpublished works. The exceptions to the above are on-line Dictionaries, Encyclopedias, Glossaries, Thesauruses, and other similar reference materials." DBI is browsable by subjects, including "Feminism," "Home economics," "Witchcraft," "Women writers and writing," Woman's suffrage," "Women: girls," "Women: social and charitable organization," "Women: wages: employment, incomes, cost of living," "Women and education," "Women and health and medicine," "Women and industry, business, professions," "Women and law," "Women and work," "Women's roles (family and society)," and "Women's studies: history and issues."
- 6. The Celebration of Women Writers, http://digital.library.upenn.edu/women/, is a project developed in collaboration with the Online Books Page. Women writers' books digitized under the auspices of that project or linked to from the project are also indexed in the Online Books Page. In addition, Celebration of Women Writers links to biographical and bibliographical information about women writers.
- 7. While it is the case that catalog records also in effect provide indexing at the chapter-heading level, through keyword searching of contents notes, not all books are so endowed. The record for the Stanislawski title has a contents note, but Moseley's doesn't. WorldCat provides a link to the table of contents for the Moseley book, but the content of these links is not retrieved through a WorldCat keyword search.

[Phyllis Holman Weisbard, Distinguished Academic Librarian at the University of Wisconsin—Madison, is also Women's Studies Librarian for the University of Wisconsin System.]