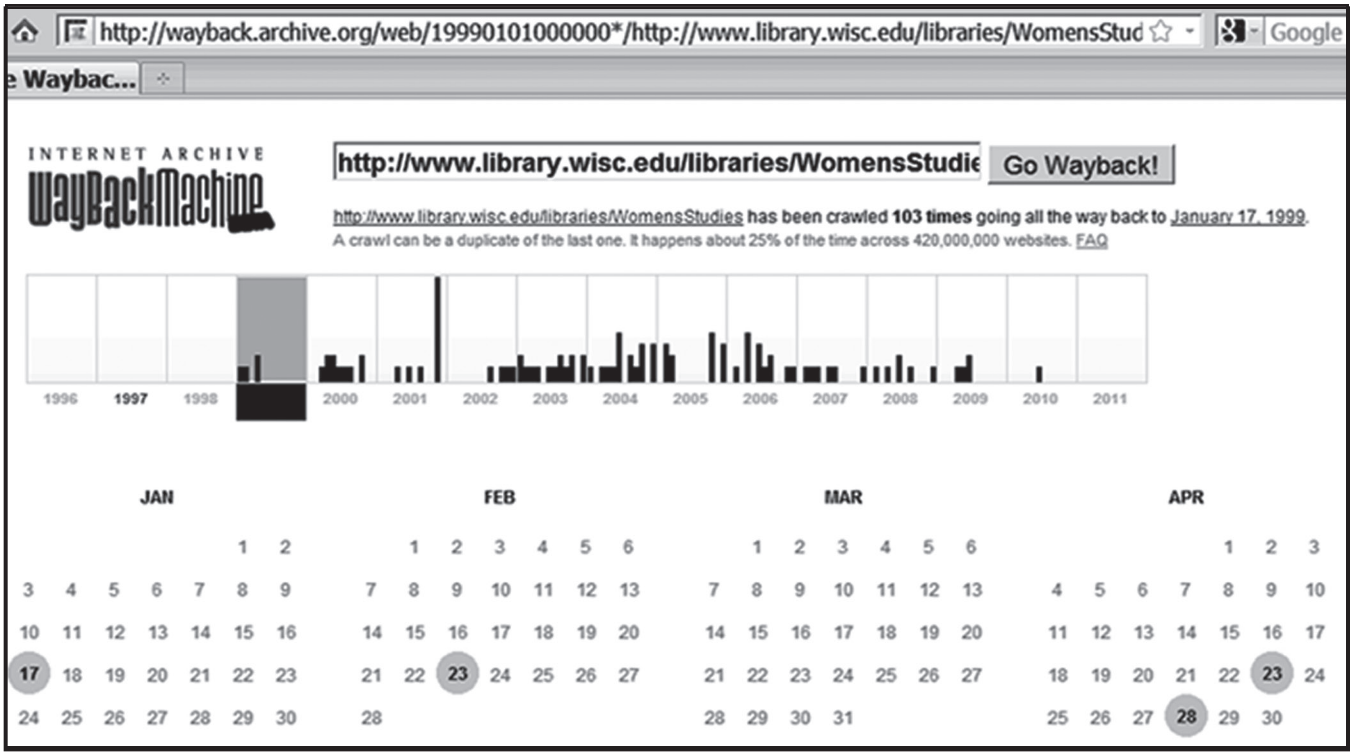
Oldies but Goodies: Archiving Web-Based Information

by Phyllis Holman Weisbard

As more and more information gravitates to electronic-only format, historians must have queasy feelings in their stomachs that their successors will never be able to do research as they have. They are right. On the positive side, research methods have been eased considerably by the availability of digital versions of so much print material from the past. But, perversely, the material that never had a print equivalent — the "born digitals" — are most in danger of disappearing when the eager e-zinesters or earnest organizations that gave birth to them lose their interest, time, or financial backing. Their Web domains lapse (and, in the case of women-focused sites, seem most often to be purchased by pornography purveyors); their efforts are lost to history. Librarians and archivists have realized this, too, but given the billions of Web pages, blogs, tweets, and other items that have ever existed, preserving what's been distributed online is a Sisyphean undertaking. Even with several Big Players now in the picture, it is likely that only a small fraction of what's published online has been or is now being captured periodically.

The Internet Archive

Perhaps the biggest Big Player is the **Internet Archive** with its **Wayback Machine** (**http://www.archive.org/web/web.php**), which "crawls" through millions of Internet sites on a regular basis, saving the version it finds as of the crawl date. The Wayback Machine (using Alexa Internet software) has been doing this since 1996. The Wayback Machine accomplishes two purposes: it lets users see how a particular site has changed over its existence, and it preserves the site, even if it is no longer on the active Web. The entry point for the Wayback Machine is the URL of the website, blog, zine, or whatever you have reason to know existed at some point in the past, whether or not it still exists. You type in the URL and the Wayback Machine presents you with a calendar of the crawl dates for that site. You then click on a date



http://wayback.archive.org/web/19990101000000*/http://www.library.wisc.edu/libraries/WomensStud

http://www.library.wisc.edu/libraries/WomensStudie

http://www.library.wisc.edu/libraries/WomensStudies has been crawled **103 times** going all the way back to January 17, 1999.

A crawl can be a duplicate of the last one. It happens about 25% of the time across 420,000,000 websites. FAQ

JAN

					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

FEB

	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28						

MAR

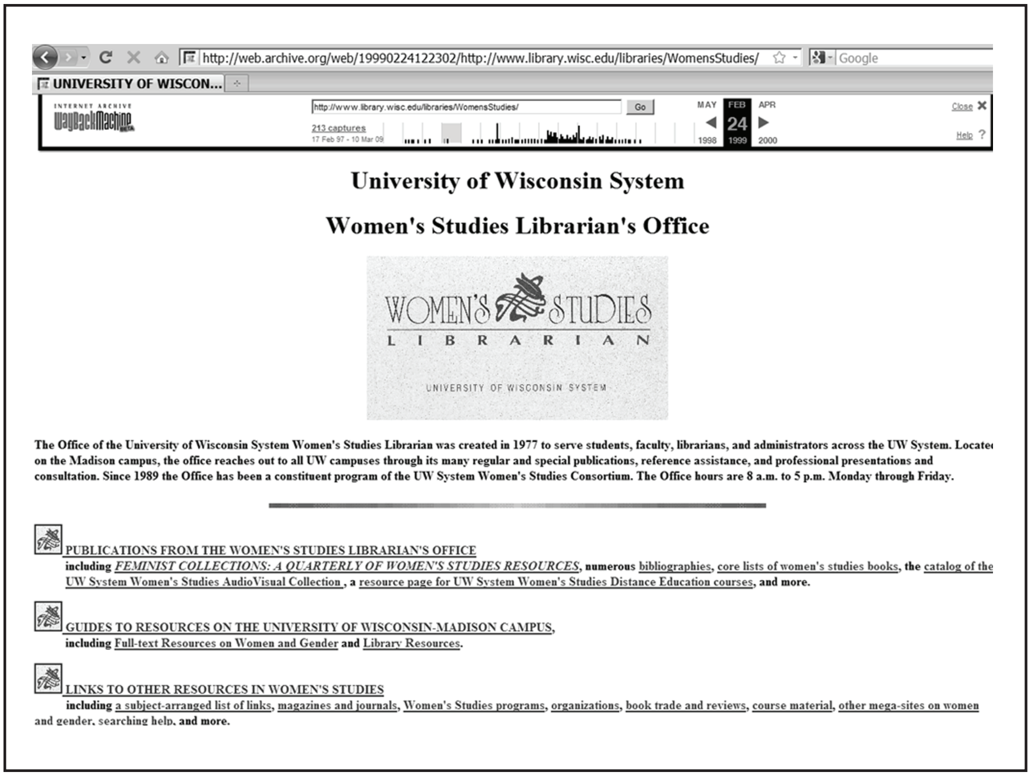
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

APR

				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

Figure 1

Figure 2


http://web.archive.org/web/19990224122302/http://www.library.wisc.edu/libraries/WomensStudies/
Google
UNIVERSITY OF WISCON...
INTERNET ARCHIVE
WaybackMachine BETA
http://www.library.wisc.edu/libraries/WomensStudies/
Go
213 captures
17 Feb 97 - 10 Mar 09
MAY
FEB
24
APR
1998
1999
2000
Close
Help
WOMEN'S STUDIES
LIBRARIAN
UNIVERSITY OF WISCONSIN SYSTEM

University of Wisconsin System

Women's Studies Librarian's Office

The Office of the University of Wisconsin System Women's Studies Librarian was created in 1977 to serve students, faculty, librarians, and administrators across the UW System. Locate on the Madison campus, the office reaches out to all UW campuses through its many regular and special publications, reference assistance, and professional presentations and consultation. Since 1989 the Office has been a constituent program of the UW System Women's Studies Consortium. The Office hours are 8 a.m. to 5 p.m. Monday through Friday.

PUBLICATIONS FROM THE WOMEN'S STUDIES LIBRARIAN'S OFFICE
including *FEMINIST COLLECTIONS: A QUARTERLY OF WOMEN'S STUDIES RESOURCES*, numerous bibliographies, core lists of women's studies books, the catalog of the UW System Women's Studies AudioVisual Collection , a resource page for UW System Women's Studies Distance Education courses, and more.

GUIDES TO RESOURCES ON THE UNIVERSITY OF WISCONSIN-MADISON CAMPUS,
including Full-text Resources on Women and Gender and Library Resources.

LINKS TO OTHER RESOURCES IN WOMEN'S STUDIES
including a subject-arranged list of links, magazines and journals, Women's Studies programs, organizations, book trade and reviews, course material, other mega-sites on women and gender, searching help, and more.

for which you'd like to see the site. If you've never used the Wayback Machine, give it a try by inserting any URL, past or present, and see what results. Here's an example using our office website: **Figure 1** shows the Wayback calendar of visits in early 1999, with a graph of visits over time, and **Figure 2** reproduces how our homepage appeared in January 1999.

The Internet Archive also offers a subscription service called "Archive-It" (**http://archive-it.org/**), currently employed by some 160 partner institutions,[1] through which partners can archive their own material. Virtually all of those libraries, historical societies, schools, museums, and NGOs have archived material on women. All this material becomes part of the Internet Archive's Wayback Machine, but at this time one needs to know the URL in order to use the Wayback Machine to find it. The Archive-It site instead provides entrée by words and phrases or by institution. It is also possible to search by collections within each institution. Putting in "women" as a general search across all institutions is about as useful as Googling "women." What does one do with, in this case, some *ninety million* hits? Well, there's one way to bring the number down considerably to guaranteed "aboutness." Just above the Archive-It result is a link to "Catalog Metadata Results," of which there are only about 200. These are, in effect, the items mounted with "women" in their subject description. At the top of the list is the Center for Egyptian Women's Legal Assistance (CEWLA, **http://www.cewla.org/**), which is archived as part of the Columbia University Libraries' Human Rights Web Archive. Thus far, CEWLA has been archived three times, in February, March, and June 2011. Many of the rest of the Catalog Metadata hits are also for women's organizations being archived by the Human Rights Web Archive. Another contributor is the IT History Society (**http://www.ithistory.org/**), a "world-wide group of over 500 members working together to assist in and promote the documentation, preservation, cataloging, and researching of Information Technology (IT) history," which archives "Past Notable Women of Computing and Mathematics." Other sites retrieved through this search include the Alabama Women's Hall of Fame, collected by the Alabama State Archives; the Maquila Women's Association Homepage and the Women's Studies Institute Twitter Feed, both collected by the University of Texas, San Antonio; Virginia Press Women, Inc., collected by the Library of Virginia; and Urban Outfitters Women's Apparel (!), part of a Teen Consumerism Collection from Miramonte High School. Trying out some other search methods on the Archive-It site, I found a Women's Ordination Web Archive from Marquette University; a collection of the Radical Women/Freedom Socialist Party, from the California Digital Library; "Ludology.org: women," a gaming blog collected by the Stanford Humanities Lab; and "Working Out Her Destiny: Women's History in Virginia 1600–2004," an exhibit of historical images and documents, collected by the Library of Virginia.

Important E-Archiving Projects Wholly on Women

Some projects have from the outset been wholly devoted to women, and do not use Archive-It software. Schlesinger Library, Radcliffe Institute, Harvard University, has two of them[2:]

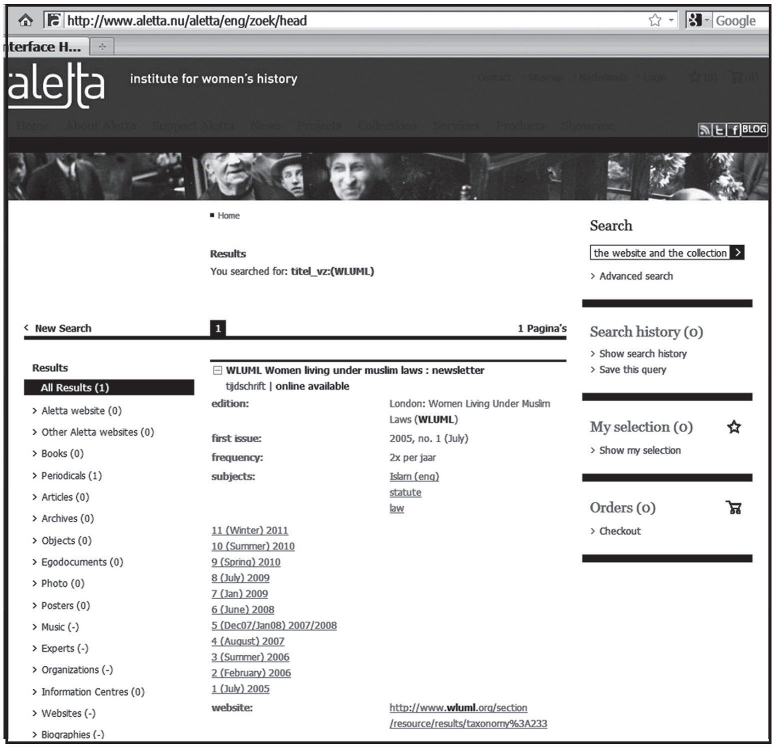
(1) **Blogs: Capturing Women's Voices** (**http://nrs.harvard.edu/urn-3:RAD.SCHL.WAX:2222628**) consists of a sample of about twenty blogs selected by the library that "illuminate the lives of African-American and Latina women, lesbians, and women grappling with health and reproductive issues, and typically reflect their engagement with politics, their personal lives and philosophies, and their work lives." Currently there are links to periodic archived versions of thirteen of these blogs, including *A Chronic Dose: A Chronic Illness Blog*, *Latina Liz*, and *SistersTalk: A Lesbian Blog with Liberal Tendencies*. *SistersTalk* may be of interest to Wisconsin readers in particular, as it was written by a woman in Beloit, WI.[3]

(2) In the case of **SL Sites: Archived Websites from Schlesinger Library Collections** (**http://nrs.harvard.edu/urn-3:RAD.SCHL:3922990**), Harvard is archiving websites related to the paper collections housed at Schlesinger, including the Boston Women's Health Book Collective (*Our Bodies, Ourselves*), Judy Chicago, and Holly Near.

Some Web archiving projects are not designed to do periodic captures of sites, yet fulfill a vital preservation role nonetheless. Such projects capture issues of online periodicals at time of issuance. If you've ever needed back issues of a web-based magazine or newsletter and found only very current issues on the website (or worse, that the magazine has ceased and no issues were available), you will appreciate such preservation efforts. Located in Amsterdam, **Aletta, Institute for Women's History** (formerly the International Archives for the Women's Movement), is home to one such project, which preserves digital periodicals (**http://www.aletta.nu/aletta/eng/collections/tijdschrift_digitaal**). Aletta staff create or save PDFs of hundreds of women's e-magazines and newsletters published online or received by the library through email. Each issue saved is linked from the Aletta catalog record for the title (along with a link to the publication's homepage, if one exists) and is accessible to any user of the catalog. This collection is especially valuable for organizational newsletters — classic ephemeral "grey literature" on the border between library and archival material — which are primary sources for studying the organizations themselves as well as activist interests during a particular period.

See It, Tell It, Change It!, a newsletter from the Third Wave Foundation in the mid-2000s, is one such title that has come and gone online. The organization continues to exist, but communicates more recently through a blog, and the older newsletters are no longer on the Foundation's website. Aletta has them. Issues of *See It, Tell It, Change It!* happen to be retrievable as well through the Wayback Machine, but one can't rely on the Machine to have all back issues of everything, and it is spotty for newsletters from organizations. Sometimes the Web crawler didn't burrow far enough into the website to find them (e.g., the newsletter from AWID [Association for Women's Rights in Development]); in other cases, the newsletters were locked behind "members only" walls. Some were never posted on websites because they were only sent through email (e.g., *Pink Link*, a Dutch gay and lesbian publication). It's also much faster to find the back issues in the Aletta collection than it is to poke around through the old versions of an organization's website using the Wayback Machine.

Figure 3 shows the Aletta catalog record for another online newsletter, *WLUML: Women Living Under Muslim Laws*. Note that each archived issue is separately linked to its file on the Aletta server, and that Aletta also provides a link to the original, organizational home for the newsletter.



Results
You searched for: **titel_vz:(WLUML)**

⊟ **WLUML Women living under muslim laws : newsletter**
tijdschrift | **online available**

edition:	London: Women Living Under Muslim Laws (**WLUML**)
first issue:	2005, no. 1 (July)
frequency:	2x per jaar
subjects:	Islam (eng) statute law

11 (Winter) 2011
10 (Summer) 2010
9 (Spring) 2010
8 (July) 2009
7 (Jan) 2009
6 (June) 2008
5 (Dec07/Jan08) 2007/2008
4 (August) 2007
3 (Summer) 2006
2 (February) 2006
1 (July) 2005

website: http://www.wluml.org/section/resource/results/taxonomy%3A233

Figure 3

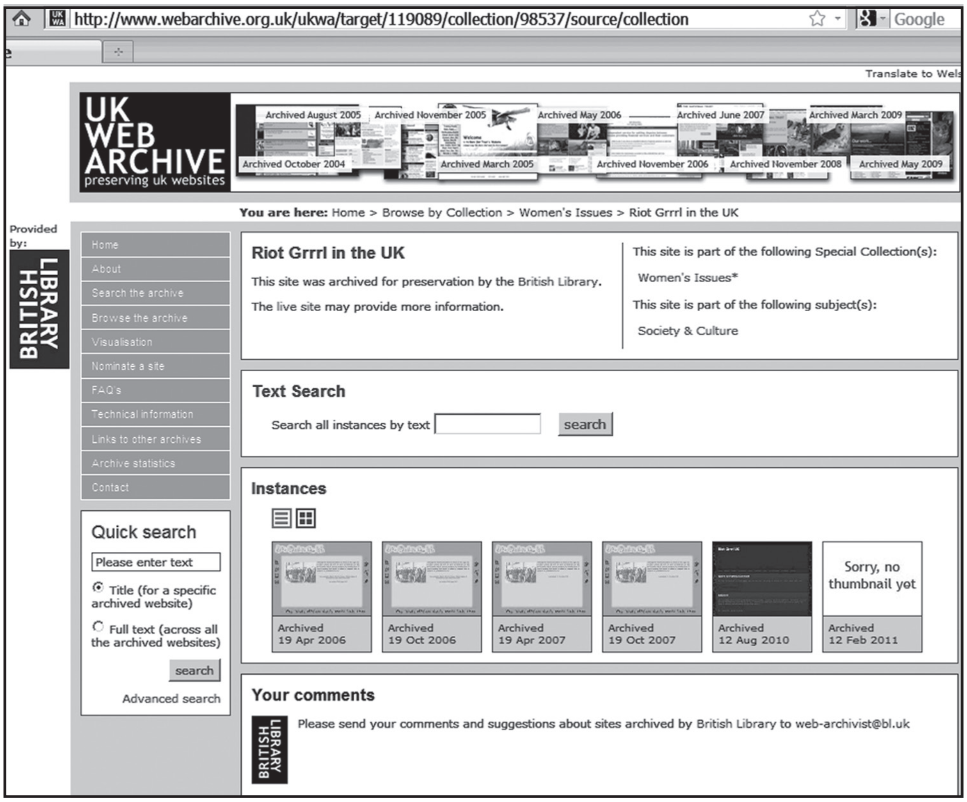
LOCKSS and Portico

Another byproduct of the availability of electronic versions of periodicals (and, increasingly, books) — and of the high cost of storing print materials — is that libraries have become less afraid to discard their print runs. At this point most still keep a "last copy," perhaps stored offsite and in collaboration with other institutions, but many now question whether it is necessary even to do that. Maybe one print run per region of the U.S. or per other country should suffice "for old time's sake," if there's a trusted cover-to-cover electronic version available. Enter projects that ensure that trusted e-copies are stored in perpetuity and attention is paid to format migration, as needed, with advances in technology (anyone try to use a floppy disc lately?).

LOCKSS (Lots of Copies Keep Stuff Safe), the related project CLOCKSS,[4] and **Portico** are initiatives to assure permanence somewhere of electronic versions of periodicals and more. **LOCKSS** (**http://lockss.stanford.edu/**) is an international nonprofit alliance of libraries using digital preservation open-source software developed at Stanford University. With permission from publishers, LOCKSS member libraries sign up for titles for which they are willing to take archiving responsibility; but rather than have the files reside on just one server, the archived content is replicated across the network of LOCKSS members and is accessible to libraries subscribing to the particular title. To date, 7,100 e-journal titles from approximately 470 publishers are part of the Global LOCKSS Network. Women-focused titles in LOCKSS include *Camera Obscura, Meridians, Journal of International Women's Studies, Journal of Women's History, Gender Issues, Gender, Place & Culture, Women in Management Review, WSQ: Women's Studies Quarterly, Violence Against Women, Nashim: A Journal of Jewish Women's Studies and Gender, Journal of Mideast Women's Studies,* and *Feminism & Psychology*.[5]

The nonprofit **Portico** (**http://www.portico.org/digital-preservation/**), launched in 2002 by JSTOR with a grant from the Andrew W. Mellon Foundation, has a somewhat different model. Instead of a network that links copies mounted on library servers, Portico houses all e-copies itself and licenses its files to libraries. It is only invoked when a "trigger" event occurs, such when a publisher ceases publishing a title or goes out of business entirely. Portico currently has 129 publishers, over 12,000 e-journal titles, some 103,000 e-books, and 46 databases.[6] *Journal of Women and Minorities in Science, Journal of Midwifery & Women's Health, Women's Health Issues, Women's Studies International Forum, Journal of Women & Aging, Journal of Women, Politics, & Policy, Gender & History*, and *Gender & Language* are all Portico journal titles in women's studies. *Defining Gender* and *Women in the National Archives* are two databases preserved by Portico.

Figure 4



http://www.webarchive.org.uk/ukwa/target/119089/collection/98537/source/collection

You are here: Home > Browse by Collection > Women's Issues > Riot Grrrl in the UK

Riot Grrrl in the UK

This site was archived for preservation by the British Library.

The live site may provide more information.

This site is part of the following Special Collection(s):

Women's Issues*

This site is part of the following subject(s):

Society & Culture

Text Search

Search all instances by text

Instances

Archived 19 Apr 2006

Archived 19 Oct 2006

Archived 19 Apr 2007

Archived 19 Oct 2007

Archived 12 Aug 2010

Sorry, no thumbnail yet

Archived 12 Feb 2011

Your comments

Please send your comments and suggestions about sites archived by British Library to web-archivist@bl.uk

National E-Archiving Activities

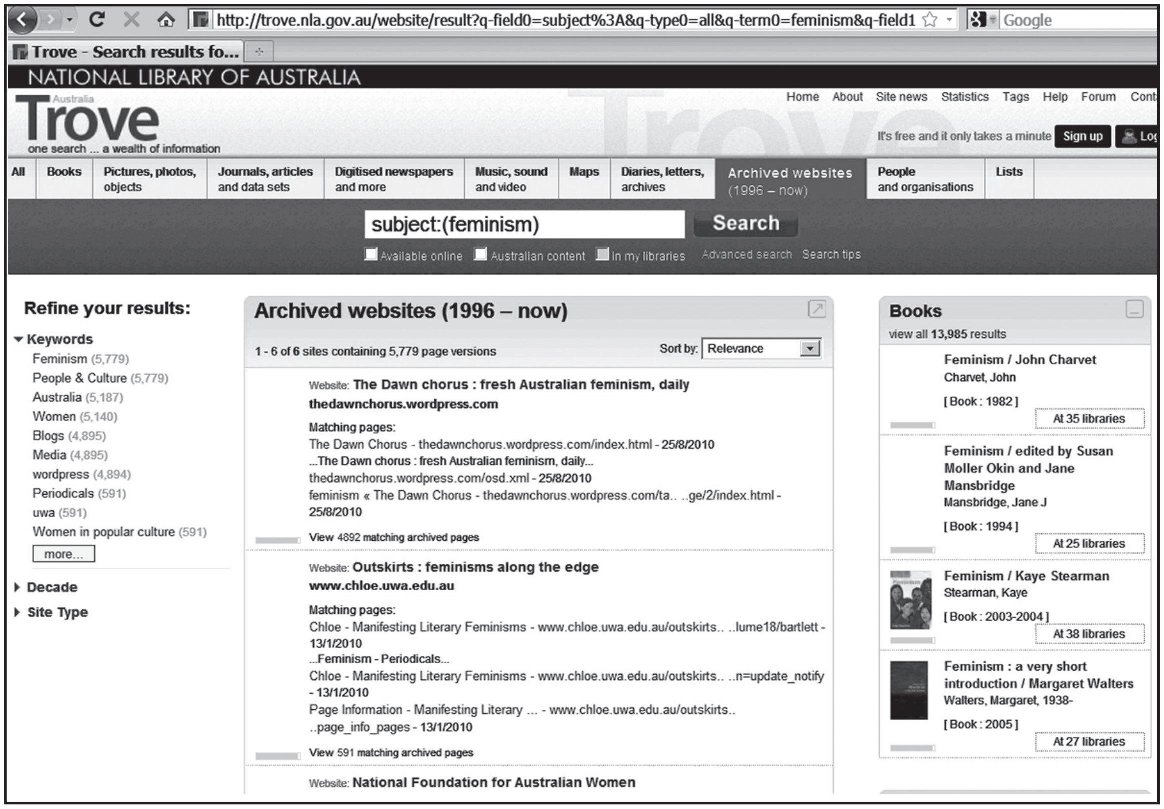
Several countries, including the U.K., Australia, and Canada, have invested in Web archiving at the national level. The **UK Web Archive** (**http://www.webarchive.org.uk/ukwa/**), provided by the British Library in partnership with several other libraries in Britain, has been capturing websites since 2004. Its aim is to preserve sites that "publish research, that reflect the diversity of lives, interests and activities throughout the UK, and demonstrate web innovation. This includes "grey literature" sites: those that carry briefings, reports, policy statements, and other ephemeral but significant forms of information.[7] It is not necessary to know URLs in order to search the UK

Web Archive, as it supports text searching of the site title or URL. Of more interest to women's studies is the fact that this archive has pulled together significant items into special collections, including one for "women's issues"[8] that is maintained in collaboration with the Women's Library, London Metropolitan University. There are currently more than 300 websites from women's organizations and campaigns, as well as research reports, women-focused government publications and statistics, blogs, and e-zines. Diverse examples include the Bedford Centre for the History of Women (eleven captures since April 2006), a blog called *Domestic Sluttery* (three captures since July 2010), the Older Feminist Network (seven captures since June 2006), and the governmental Women's National Commission (eleven captures since October 2005). Each entry has a link to what was the live site at the time of first capture, and most of these remain active at this time. The real value, though, for archival purposes, is preservation of what was — documentation of both how the site evolved over time and, if the live site disappears, its existence in the first place. An example is the blog *Riot Grrl in the UK*, which appears to have stopped being updated as of 2007 (see **Figure 4**).

Australia's effort is called **PANDORA**, an acronym for Preserving and Accessing Networked Documentary Resources of Australia (**http://pandora.nla.gov.au/**). It was started in 1996 by the National Library of Australia and has grown to include nine collaborating Australian libraries and cultural organizations. Like the UK project, PANDORA is selective, collecting "materials that document the cultural, social, political life and activities of the Australian community and intellectual and expressive activities of Australians."[9]

There are several ways to search PANDORA material; I recommend using the advanced search at **http://trove.nla.gov.au/website?q&adv=y**, where you can search by subject, title or keyword. The subject "feminism" has six results: *The Dawn Chorus: Fresh Australian Feminism Daily* (archived in July 2009 and again in August 2010), *Outskirts: Feminism Along the Edge* (23 issues of this periodical archived since 1996), and the National Foundation for Australian Women (annual captures since 2005), all visible in **Figure 5**; plus Women's Rights Action Australia (annual captures from 2004), the blog *Zero at the Bone* (captured once in November 2009), and the website of an Australian Broadcasting Company journalist named Virginia Haussegger (annual

Figure 5



http://trove.nla.gov.au/website/result?q-field0=subject%3A&q-type0=all&q-term0=feminism&q-field1

Trove - Search results fo...

subject:(feminism)

□ Available online □ Australian content □ In my libraries

Archived websites (1996 – now)

1 - 6 of 6 sites containing 5,779 page versions

Sort by: Relevance

Website: **The Dawn chorus : fresh Australian feminism, daily**
thedawnchorus.wordpress.com

Matching pages:
The Dawn Chorus - thedawnchorus.wordpress.com/index.html - **25/8/2010**
...**The Dawn chorus : fresh Australian feminism, daily**...
thedawnchorus.wordpress.com/osd.xml - **25/8/2010**
feminism « The Dawn Chorus - thedawnchorus.wordpress.com/ta.. ..ge/2/index.html - **25/8/2010**

View 4892 matching archived pages

Website: **Outskirts : feminisms along the edge**
www.chloe.uwa.edu.au

Matching pages:
Chloe - Manifesting Literary Feminisms - www.chloe.uwa.edu.au/outskirts.. ..lume18/bartlett - **13/1/2010**
...**Feminism - Periodicals**...
Chloe - Manifesting Literary Feminisms - www.chloe.uwa.edu.au/outskirts.. ..n=update_notify - **13/1/2010**
Page Information - Manifesting Literary ... - www.chloe.uwa.edu.au/outskirts.. ..page_info_pages - **13/1/2010**

View 591 matching archived pages

Website: **National Foundation for Australian Women**

Refine your results:

Keywords
Feminism (5,779)
People & Culture (5,779)
Australia (5,187)
Women (5,140)
Blogs (4,895)
Media (4,895)
wordpress (4,894)
Periodicals (591)
uwa (591)
Women in popular culture (591)

Books
view all **13,985** results

Feminism / John Charvet
Charvet, John
[Book : 1982]
At 35 libraries

Feminism / edited by Susan Moller Okin and Jane Mansbridge
Mansbridge, Jane J
[Book : 1994]
At 25 libraries

Feminism / Kaye Stearman
Stearman, Kaye
[Book : 2003-2004]
At 38 libraries

Feminism : a very short introduction / Margaret Walters
Walters, Margaret, 1938-
[Book : 2005]
At 27 libraries

Google
NATIONAL LIBRARY OF AUSTRALIA
Australia
Trove
one search ... a wealth of information
Home
About
Site news
Statistics
Tags
Help
Forum
Cont
It's free and it only takes a minute
Sign up
Log
All
Books
Pictures, photos, objects
Journals, articles and data sets
Digitised newspapers and more
Music, sound and video
Maps
Diaries, letters, archives
Archived websites (1996 – now)
People and organisations
Lists
Search
Advanced search
Search tips
more...
Decade
Site Type

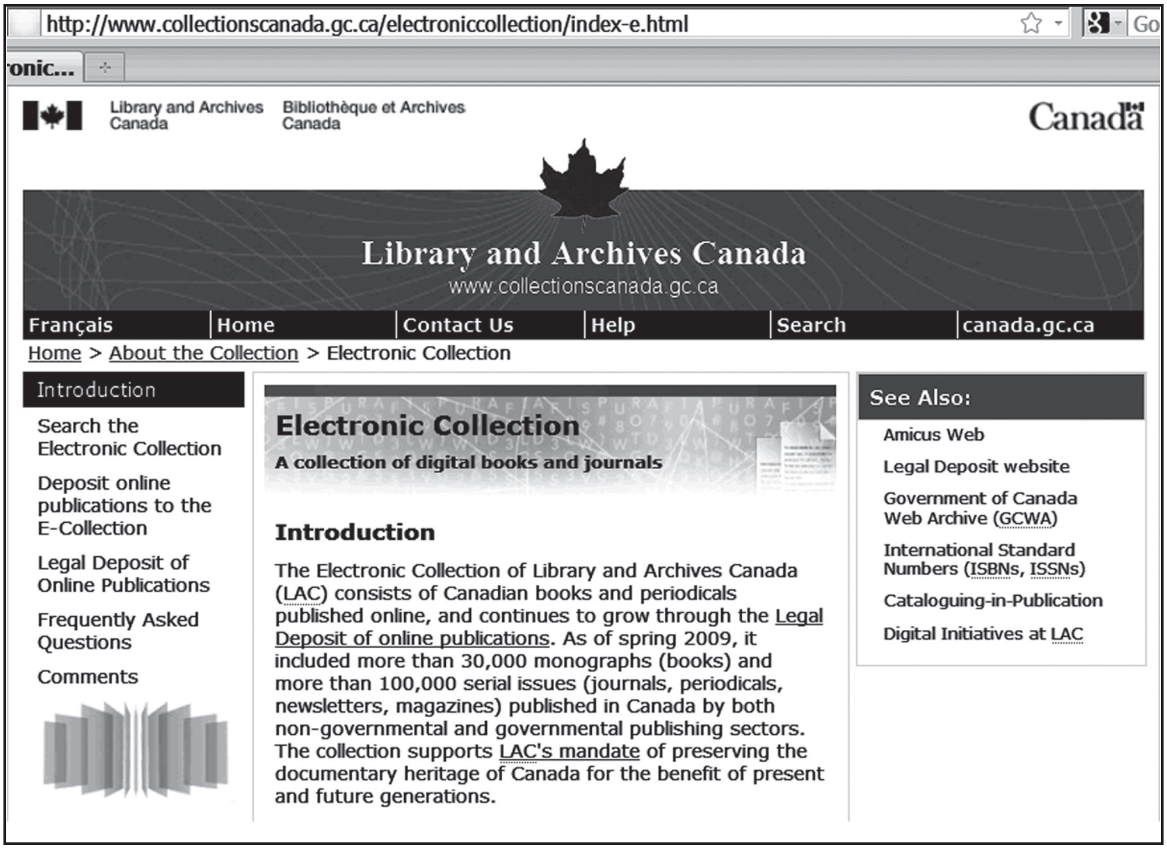
captures since 2010). A title search for “feminism” raises the number of results to 185 (12,678 page versions captured), while a keyword search for that concept ups the number to about 800 sites (more than 200,000 page versions captured.)

The Electronic Collection from **Library and Archives Canada** (**LAC**, at **http://www.collectionscanada.gc.ca/electroniccollection/index-e.html**) is not doing a period crawl for Canadian sites, blogs, and tweets. Instead it consists of Canadian e-books and e-periodicals collected and archived once. It also includes everything in the Web domain of the Federal Government of Canada (**http://www.collectionscanada.gc.ca/webarchives/index-e.html**). LAC’s general mandate is “preserving the documentary heritage of Canada for the benefit of present and future generations.”[10] See **Figure 6**.

Archived e-periodicals on women include *Women’Space* (several issues of this magazine on women and the Internet, until it ceased in 2000), the *Health Newsletter* of the Native Women’s Association of Canada (two issues and a supplement, 2009 and 2010), *Women in Judaism* (eight issues, 1997–2007 — this journal is still publishing at University of Toronto; the archiving seems to be a bit behind), and numerous publications from Status of Women Canada. Examples of e-books include *Mother’s Voices: What Women Say About Pregnancy, Childbirth, and Early Motherhood* (Public Health Agency of Canada, 2009) and *Reality Check: How Rape Mythology in the Legal System Undermines the Equality Rights of Women Who are Sexual Assault Survivors* (by Kathryn Penwill, Action Ontarienne Contre la Violence Faite Aux Femmes, 2002). Books and periodicals published in Canada must be deposited with LAC according to the Library and Archives of Canada Act, which was extended to online publications in 2007; thus, this is a rich, ongoing source for Web archiving.

Web archiving projects differ in size (Wayback Machine compared to the Schlesinger Blogs collection), location (global compared to exclusively national efforts), and purpose (crawling and preserving “born digital” websites vs. “dark storage” of academic journals). But all share a desire for a goodly segment of e-productivity to be available for future generations. To date, quite a bit of material on women has been preserved, but mostly by-the-by rather than by

Figure 6


http://www.collectionscanada.gc.ca/electroniccollection/index-e.html
ronic...
Library and Archives Canada
Bibliothèque et Archives Canada
Canada
Library and Archives Canada
www.collectionscanada.gc.ca
Français
Home
Contact Us
Help
Search
canada.gc.ca
Home > About the Collection > Electronic Collection
Introduction
Search the Electronic Collection
Deposit online publications to the E-Collection
Legal Deposit of Online Publications
Frequently Asked Questions
Comments
Electronic Collection
A collection of digital books and journals
See Also:
Amicus Web
Legal Deposit website
Government of Canada Web Archive (GCWA)
International Standard Numbers (ISBNs, ISSNs)
Cataloguing-in-Publication
Digital Initiatives at LAC

Introduction

The Electronic Collection of Library and Archives Canada (LAC) consists of Canadian books and periodicals published online, and continues to grow through the Legal Deposit of online publications. As of spring 2009, it included more than 30,000 monographs (books) and more than 100,000 serial issues (journals, periodicals, newsletters, magazines) published in Canada by both non-governmental and governmental publishing sectors. The collection supports LAC's mandate of preserving the documentary heritage of Canada for the benefit of present and future generations.

design. Women's studies scholars and their librarian and archivist allies may want to be more proactive and recommend publications and sites that should be preserved.

Notes

1. The Wisconsin Historical Society and the Stem Cell Research Archives Project at University of Wisconsin–Madison are the two current partners in Wisconsin. Most of the partners are U.S.-based. For a full list of partners, see **http://www.archive-it.org/public/partners.html**, accessed July 26, 2011.

2. Both Schlesinger projects use WAX: Web Archive Collection Service, a system developed at Harvard. The components of WAX include the Heritrix Web crawler, the Internet Archive's Wayback index and rendering tool, the Nutchwax indexing tool, and a scheduling tool called Quartz. See **http://hul.harvard.edu/ois/systems/wax/**.

3. Now a Twitter feed at sistertalk.net/blog; no longer crawled by the Schlesinger Project.

4. CLOCKSS (Controlled LOCKSS, at **http://www.clockss.org/clockss/Home**, accessed July 26, 2011) has aspects of both LOCKSS and Portico. As with LOCKSS, the copies are stored decentralized on library servers. Like Portico, CLOCKSS is a "dark" archive, only tapped when trigger events occur.

5. **http://lockss.stanford.edu/lockss/Publishers_and_Titles**, accessed July 26, 2011.

6. **http://www.portico.org/digital-preservation/the-archive-content-access/archive-facts-figures/**, accessed July 26, 2011.

7. **http://www.webarchive.org.uk/ukwa/info/about**, accessed July 26, 2011.

8. **http://www.webarchive.org.uk/ukwa/collection/98537/page/1/source/collection**, accessed July 26, 2011.

9. **http://pandora.nla.gov.au/overview.html**, accessed July 26, 2011.

10. **http://www.collectionscanada.gc.ca/electroniccollection/index-e.html**, accessed July 26, 2011. The simple "Search All" box on the homepage for Library and Archives Canada (**http://www.collectionscanada.gc.ca/index-e.html**) can be used to find all editions held by LAC. E-collection items are labeled [**electronic resource**]. Put title phrases in quotation marks.

[*Phyllis Holman Weisbard is the women's studies librarian for the University of Wisconsin System and the co-editor of* Feminist Collections.]