Collecting the Human Rights Web

Tessa Fallon
Web Collection Curator

Pamela Graham
Director, Center for Human Rights Documentation & Research

Alex Thurman
Web Collection Curator

SALALM LVI, May 28-June 1, 2011
Philadelphia, PA
Human Rights Web Collecting

- Selecting the web
- CUL’s Web Collecting Program / HRWA
- Permissions
- Harvesting
- Access
- Assessment
What is the HR Web?

- Reports, working papers, articles, books
- Journals, newsletters
- News releases / reports
- Documents, Testimony
- Blog Posts
- Multi-media / visual content
- Websites as objects
- Social media
Why Collect the HR Web?

- High Research Value
- Shifting models of publishing / replace print
- Expand scope and range of collecting
- Ephemeral and at-risk
The Human Rights Web Archive is an initiative by Columbia University Libraries and Information Services (CUL) and its Center for Human Rights Documentation & Research (CHRDR), with support from the Andrew W. Mellon Foundation, to select, preserve, and provide access to freely available internet resources, specifically addressing at-risk websites in the area of human rights. As expressed in the Universal Declaration of Human Rights, this concept includes such commonly recognized areas as freedom from torture, slavery, and arbitrary arrest, but also embraces social, cultural, and economic rights, freedom of movement and assembly, the right to work, and more.

The interdisciplinary, wide-ranging, and highly networked field of human rights relies extensively on web-based information to disseminate publications, reports, media, and other content. While placing material on the internet allows for widespread and rapid access to information, much of this content is at risk of disappearing within a relatively short time period. Archiving this information and facilitating its ongoing discovery and use is essential for building the documentary record of human rights practice and research.

The Human Rights Web Archive project complements Columbia University Libraries’ strength in human rights
Timeline and Goals

• Mellon Foundation Support
  – 2008 1-year planning grant
  – 2009 3-year grant

• Integrate selection, harvest, and access into mainstream collection development

• Focus on HR but explore other collecting areas
Selection Process

- Selectors @ Columbia & Cornell
- Web Collection Curators
- Open Nominations
  - Site owners
  - Public
Criteria and Priorities

- NGOs
- National Human Rights Institutes
- At-Risk + Established Sites
- Bloggers
- Multi-regional, multi-lingual
- Prioritize domains, countries w/o archiving initiatives
- Coordinate with similar web collecting programs
Nominate your site for the Human Rights Web Archive

This form is for site owners who wish to nominate their organization or site for inclusion in the Human Rights Web Archive. If you are not the owner of the site you are nominating, please use the public nomination form.

Please note that we cannot guarantee the inclusion of a nominated website in this program. We review the nominations carefully and assess the subject, technical, and copyright issues that may affect a site’s suitability for archiving. Send questions and comments to culhrweb@libraries.cul.columbia.edu. See also our FAQ.

* Required

Name of website/organization *

URL of website *

Description of website/organization *

Additional Information

Permissions *
If we select your site for inclusion in the Human Rights Web Archive, will you give Columbia University Libraries/Information Services permission to crawl your site and provide access to the archived content in the Human Rights Web Archive?

☐ Yes
☐ No
☐ Please contact us to discuss permissions
Permissions Process

• Standard permission request text in English
• Also available in Spanish, French, Arabic, Portuguese and Russian
• Two-step process
  – Permission request
  – Follow-up request & notification of intent
• Site owners can reply via e-mail or linked nomination form
Permissions Data

• All permissions correspondence archived in dedicated e-mail folder
• Permission request dates and results recorded in Archivist’s Toolkit
• Statistics
  
  Requests sent: 432
  “Yes”: 240
  No response: 167
  “No”: 1
  Pending: 24
Harvesting Websites

• Goals: To preserve important web resources and help document their evolution over time
• Tools: Archive-It web archiving service, includes open source crawler Heritrix and open source display software Wayback Machine
• Preservation format: content and structure of website is copied into a WARC (Web ARChive) file
• Data storage: WARC files stored at Internet Archive and in Columbia’s Fedora repository
Harvesting Challenges

Can’t capture
- Password protected sites
- Form and database driven content

Can sometimes capture
- Javascript
- Streaming media (Flash)
- Robots.txt exclusions
Harvesting Process

- Initial site assessment and crawl scoping
- Test crawl
- Re-scoping site to capture missing content or block unwanted content
- Actual capture
- Quality control review
- Assigned for regular re-crawling (quarterly)
Access and Assessment

• Access
  – Cataloging
    • ICR
  – Archive-It
    • Dublin Core
  – CLIO
    • MARC

• Assessment
  – Audience
  – Usability
  – Quality Review
COLLECTION: Human Rights

Partner: Columbia University Libraries
Crawling Activity: 2008 - present
Topics: Society & Culture
Videos Archived: list videos

An initiative of CUL's Center for Human Rights Documentation and Research, the Human Rights Web Archive is an effort to preserve and ensure access to freely available human rights resources created mainly by non-governmental organizations, national human rights institutions, and individuals. Funded by the Andrew W. Mellon Foundation.

Enter Search Terms Here...
GO Advanced Search

Prefer to search by URL? Enter URLs Here...
http://

- 1. Non-governmental organizations (NGOs) - Amnesty International sites
    Canadian English-speaking section of Amnesty International. For francophone section, see http://www.amniste.ca/
    Hungary section of Amnesty International. In Hungarian.
    Ireland section of Amnesty International. In English.
Dublin Core

http://www.abuelas.org.ar/ Edit
Updated: May 13, 2011 2:24 PM by fallont

Seed Operations
Human Rights

Metadata
Settings
Comments

Back to Seed Management Previous Next Save Save And Return To List Save and Show Previous Save and Show Next

Dublin Core

Title
Abuelas de Plaza de Mayo

Grab Title

Creator
Asociación de Abuelas de Plaza de Mayo

Subject
Children of women prisoners
Disappeared persons
Human rights
Missing children

Description
Argentina-focused human rights organization based in Argentina which aims to locate the children who disappeared as a result of the political repression

Publisher
Asociación de Abuelas de Plaza de Mayo

Contributor

Date
2008-05-21 00:00:00

Type
text

Format
txt/html

Identifier
http://www.abuelas.org.ar/

Source

Metadata

You can add seed level metadata below using the Dublin Core Element Set. Information you include in the title and description fields will appear on your collection page on www.archive-it.org.

You can export seed metadata for your collection via an XML feed. Click the "XML" link located on the upper right of your collection management page. To export metadata for all your collections, use the "XML" link located on the home screen.

If you wish to have one of your seeds not be displayed on the public www.archive-it.org website or you would like to change seed frequency, click the "Settings" tab.
000 01645cei a22004333a 4500
001 7038393
005 20110518163250.0
006 m|||d|||d|||
007 c|n|n|n|n|n|n|n
008 090219cuuuw5999ag u wsslii|||l 2spa d
035 _ |a (OCoLC)507380043
035 _ |a (OCoLC)ocn507388343
035 _ |a (NNC)7038393
040 _ |a NNC |c NNC
043 _ |a s-ag--
050 _ |a JC599 A7
110 2 _ |a Asociación de Abuelas de Plaza de Mayo.
245 10 |a Abuelas de Plaza de Mayo [h [electronic resource].
260 _ |a Argentina : |b Abuelas de Plaza de Mayo.
500 _ |a Viewed on March 6, 2009.
500 _ |a Preserved by Columbia University Libraries' Web Resources Collection Program.
520 _ |a "Grandmothers of May Square is a civil non governmental organization which is not affiliated to any political party. Its aim is to locate the children who disappeared as a result of the political repressive methodology implemented by the military dictatorship and to restore them to their legitimate families."
546 _ |a In Spanish; some text also in English and Italian.
610 20 |a Asociación de Abuelas de Plaza de Mayo.
650 _0 |a Human rights |z Argentina.
650 _0 |a Children |z Argentina.
650 _0 |a Disappeared persons |z Argentina.
650 _0 |a Human rights advocacy |z Argentina.
653 _0 |a Non-governmental organizations
655 _0 |a Web sites.
852 _ |x pmg16
856 40 |u http://www.columbia.edu/cgi-bin/curl/resolve?ci=7038393 |z Current site
856 40 |u http://www.columbia.edu/cgi-bin/curl/resolve?ci=7038393.002 |z Archived site
900 _ |a AUTH
920 40 |u http://www.abuelas.org.ar/ |z Current site
948 1 _ |a 20050305 |b o |c at2186 |d OSMC
965 _ |a 965nportal
965 _ |a 965nportal
Human Rights Portal
(under development)
Assessment

• Audience
  – Academic community
  – NGOs
  – Independent researchers
  – Government employees
  – Human rights defenders
Usability & Quality Review

• Interface
  – User testing of tools and functions of Human Rights Portal
  – Familiarizing users with web archives

• Content
  – Quality of archived sites
  – Preserving integrity of archived materials
Other issues to consider...

- Privacy
- Security
- Collecting scope
- Collaboration
- Long-term preservation