Access to Individual Harvested Sites in a Web Archive

Tracy Meehleib
DLF Fall Forum, Providence, RI
November 13th, 2008
EVENT-DRIVEN

- September 11th, 2001
- Winter Olympic Games 2002
- U.S. Congresses 107th, 108th, 109th, etc.
- Iraq War 2003-
- Papal Transition 2005
- Supreme Court Nominations 2005-2006
- Crisis in Darfur, Sudan 2006
- Egypt 2008

FORMAT/COLLECTION-DRIVEN

- Organizational Sites corresponding to Papers/Archives collected by LC’s Manuscript Division
- Sites corresponding to creators whose works are collected by/represented in LC’s P&P Division
- Legal Blawgs identified by the Law Division
Iraq War, 2003 Web Archive

The Potential Costs of War With Iraq... Is the Administration playing "Russian Roulette" with the economy?

The White House strategy of war with Iraq will begin this week, and the Bush administration has stubbornly refused to provide a substantive estimate of what it might cost. In the words of Secretary of Defense Donald Rumsfeld, it is not "a very useful exercise" to prepare cost estimates in light of the many variables.

Iraq War

Updated 02 May 2003

New Open: TCP Forum
- Go To The Publisher's Desk
- Big Selling: Air Passion - 1.
- New Success: All The preseason Has - 1.
- Site Top 10: The Now Top 10 - 0.
- MSP: The MSP will Start From Pay Boys - 0.
- BTP: The BTP is Selling - 0.
- BTP: Campaigning is ~Big60-Page Mal
t- Your Site: 1000.001.

Iraq Aftermath

Declared No Lesser a Combat Zone, There's No Peace In Alaskabomb

Refugees Are Neglected in Combat Zones

NY Times, 03 May 2003

Al Destroyed as Refugees

Albomis Are Neglected in Combat Zones

NY Times, 03 May 2003

Iraq's Other Government Challenges: Small Arms

Waffenbrief, Associated Press, 02 May 2003

Iraq Situation: Conflict, High Art Agencies

New York Times, 02 May 2003

Istanbul's Famed Art Market Cages

New York Times, 02 May 2003

The War in Iraq: Misreporting

P.M. Anderson, The Independent, 01 May 2003

Iraqi Women's Authorities

New York Times, 02 May 2003

Iraqis Wary After Violence

The New York Times, 01 May 2003

Picture: Baghdad, Iraq, 01 May 2003

Iraq's Top Sites

Iraq Image Archive

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

 Iraq War - Iraq's Other Government Challenges: Small Arms

Iraqi Women's Authorities

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

Iraqi War News

 Iraq War - Wikipedia, The Free Encyclopedia

Iraq War - Wikipedia, The Free Encyclopedia
Crisis in Darfur, Sudan 2006 Web Archive

On May 3, 2006, a commission of Sudan activists and leaders in the movement to stop genocide and slavery in Sudan embarked on a week-long fact-finding and relief mission in Southern Sudan and Darfur. Click here to see photos and find out what it was like for everyone involved.

It's a Divestment Bonanza!

We've recently received word that the state of Maine has divested from Sudan. That's already fantastic news, but to top that off, one of California's major retirement funds — CalSTRS — has voted to divest as well. About $6 billion has been kept out of the hands of the Sudanese regime thanks to these victories. Can you help keep the momentum going by starting your own divestment campaign?

Global Justice Campaign

One standard is For a Fair & Just US Policy toward Multinational Corporations, Human Rights, and Sovereignty

Arthur R. Ashe Foreign Policy Library Resources

Archival World Video Film Series Viewpoint Lecture Series Breadth: Corrison Writers' Corner

TransAfrica Forum News Articles, Commentaries

TransAfrica Forum is seeking to fill the position of President

TransAfrica Forum News, November 2006

An October Surprise: Grassroots Movement's 20th Anniversary

Foreign Policy Victory

By Dr. Sylvia Hill, Vice Chair, TransAfrica Forum

On October 2, 1986 the USA anti-apartheid
Visual Image Web Archive

Milton Rogovin
Social Documentary Photographer

Prokudin-Gorskii Images
Color photographs from the Russian Empire taken a century ago (1909-1915).

Greg Lynn Form
1847 Lincoln Boulevard
Venice, CA 90291
USA
+1 (310) 881-2600 tel
+1 (310) 881-9209 fax
email
Library of Congress Web Archives

Election 2000  800
Election 2002  4000
Election 2004  1945
Election 2006  2098
Election 2008  2000*

107th Congress  579
108th Congress  583
109th Congress  580*
110th Congress  580*

September 11, 2001  2300
Winter Olympics 2002  62
Iraq War 2003-  231
Papal Transition 2005  192
Crisis In Darfur, Sudan 2006  218
Visual Images  17
Organizational Sites, Manuscript Division  30
U.S. Supreme Court Nominations 2005-2006  281
Legal Blawgs 2007  90
Egypt 2008  30
Web Archives Processing Workflow

- Identify and select sites

- Create a “seed list” of sites to be crawled, determine how frequently they will be crawled and submit to IA

- IA captures selected web sites as **W/ARC files**

- Create catalogers’ list and a **MODS template** for metadata extraction and submit them to IA

- IA extracts metadata from archived web sites (**W/ARC files**) into the MODS template
Web Archives Processing Workflow

- Metadata extraction results in a preliminary MODS record for each archived site

- Enhance record, reviewing & revising some values if needed (title, language, abstract, keywords) and adding some values (LCSH headings—subjects and sometimes names)

- Register item-level handles

- Load MODS records onto server, index, generate item-level search/browse

- Create collection-level record in ILS and register collection-level handle
W/ARC Files

- ARC file format used by Internet Archive to store web archives since 1996

- Access to archived web sites in ARC files depends on large-scale indexing of ARC files

- ARC file indexing can only support access by URL and date

- **WARC** has since been developed as an extension to ARC and is now an ISO standard, it carries a little more metadata, but access to web sites in WARC files is still very limited

- As tools are developed to support WARC files, WARC files will be preferred to ARC files for storing web archives
NutchWAX Keyword Indexing & Search

• **NutchWAX** is a web archive search technology based on **Nutch** an open source web search software—designed to improve access to W/ARC files

• **NutchWAX** ("**Nutch** + **Web Archive eXtensions**") can search keyword indexes of ARC files—so it extends more basic access (by URL and Date) to include keyword access

• However, building/rebuilding indexes for each archive is still cumbersome and expensive

• Building/rebuilding comprehensive indexes that include more than one web archive is even more cumbersome and expensive

• And even with NutchWAX’s keyword access, archived sites are not searchable/browseable/integratable with other web archives or library resources
Why Provide Site Level Access to these Sites?

- Access limitations of **W/ARC files** and **NutchWAX ("Nutch + Web Archive eXtensions")**
- Use of controlled vocabularies
- Leverage subject cataloging & language expertise to quickly and substantially enhance subject access
- Resources become highly integratable with other library resources at the item level
- Better precision and recall
- Persistent IDs/handles allow for stable citations and digital scholarship at site-level
- Leverage use of existing search/browse systems
How Do We Provide Site-Level Access to these Sites?

• Boilerplate as much relevant archive-level and site-level metadata as is possible into the MODS template

• Extract as much useful metadata as is possible from archived web sites W/ARC files (using a perl script or other method that grabs the metadata from meta tags in the W/ARC files)—titles, dates, file types, abstracts, subject keywords, etc.

• Leverage LC subject cataloging & language expertise and controlled vocabularies to add subject access
Overview of MODS Record Data Elements

<table>
<thead>
<tr>
<th>Data Element</th>
<th>Description</th>
</tr>
</thead>
</table>
| Title                     | Extracted from W/ARC file/HTML title tag  
                          | - Cataloger uses if viable, otherwise supplies |
| Alternative Title         | Cataloger supplies if another useful and different title displays on piece |
| Name Personal             | Included for some archives, when relevant, cataloger supplies |
| Name Corporate            | Included for some archives, when relevant, cataloger supplies |
| Type of Resource          | Boilerplate “text” |
| Genre                     | Boilerplate “Web site” |
| Origin Info               | Extracted from W/ARC file – first/last dates captured YYYMMDD(iso8601) |
| Language                  | Boilerplate in if known (iso639-2b code)  
                          | - Cataloger can supply additional languages |
| Physical Description      | Extracted from W/ARC file/MIME type, e.g., text/css, image/jpeg |
| Abstract                  | Extracted from W/ARC file/META name=description content  
                          | - Cataloger can edit/enhance |
| Subject/Keywords          | Extracted from W/ARC file/META name=keywords content  
                          | - Cataloger can edit/enhance |
| Subject/LCSH              | Cataloger supplies |
| Collection Title/PID      | Boilerplate, collection title & collection PID/handle |
| Identifier                | Boilerplate, variant of handle, e.g, hdl:loc.natlib/mrva0000.0000 |
| Note                      | Extracted from W/ARC file, resolves to URL for active site |
| Location/Usage            | Boilerplate item-level PID/handle  
                          | - PID is registered to resolve to archived Web site URL |
| Access Condition          | Boilerplate rights info/permissions info – imported from OSI records |
| Record Info               | Boilerplate record creation date  
                          | - Boilerplate record identifier, handle suffix mrva0000.0000 |
Archive size: 218 sites

Harvest info: 1 phase, multiple captures

Frequency: Varies--weekly to monthly crawls for each site

Metadata: 1 collection-level MARC record, with collection level PID
218 item-level MODS records, with item-level PIDs

LCSH: 1 boilerplate LCSH heading
Unlimited specific LCSH headings at site level—these are selected by cataloger from a list of about 20 LCSH terms that relate to the content in the archive
Catalogers’ List for Darfur, 2006 Web Archive

Wayback Machine links to seeds:

--- Live/Proxy links to the live Web. To enable proxy viewing:
        manually enter the proxy host and port informat:
        Proxy host: loc-wm.archive.org
        Port number: 9000

--- Wayback 1.4 links to Wayback Machine 1.4.

--- WARC links to WARC deployment.

- [Live/Proxy] [Wayback 1.4] [WARC] http://aagainstou.blogspot.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://arabist.net/arabawy/
- [Live/Proxy] [Wayback 1.4] [WARC] http://egyworkers.blogspot.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://ghosty1313.blogspot.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://masgon.maktoobblog.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://masradam.blogspot.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://tadamonnasr.wordpress.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://wa7damasrya.blogspot.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.6-april.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.al-araiby.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.alwasifi.org/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.alwasatparty.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.dakahilakhwani.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.egymas.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.egytiangreens.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.el3amal.net/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.elbehira.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.gharbionline.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.harakamasria.org/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.ikhwanonline.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.imbh.net/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.keeepourgas.blogspot.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.manala.net/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.mdedalghala.org/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.nowabikhwani.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.saveegyptfront.org/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.sharkiaonline.com/
- [Live/Proxy] [Wayback 1.4] [WARC] http://www.tortureinegypt.net/
- [Live/Proxy] [Wayback 1.4] [WARC] http://ymeeneshomal.blogspot.com/
Crisis in Darfur, Sudan, Web Archive, 2006 Resource Page

Each date listed below represents one capture of the site. Clicking any date will display the web site record as it appeared at that time. Following other links from that page will return the closest matches available from this Web Archive, however, some linked pages may not be available in the Archive. Please use the browser back button to return to this page if you wish to visit other archived dates.

7 archived captures for http://www.afrika.no/

* denotes when site was updated.

<table>
<thead>
<tr>
<th>Search Results for Jan 1, 2006 - Dec 31, 2006</th>
</tr>
</thead>
</table>

home >> previous page >> Crisis in Darfur, Sudan, Web Archive, 2006 Resource Page

The Library of Congress >> More Online Collections
Africa at large: Rich nations are offside in Geneva
Source: Business Day (South Africa), by Brendan Williams, 07.06.2006
The past week’s Secretariat Olympics in Geneva saw the United States more superstates than did the somber drama of Geneva, where a meeting of 60 of the world’s trade ministers failed to agree on the stalled Doha Round of the World Trade Organisation (WTO). Indeed, the virtual quasiregional rivalry of the French delegation over their Brazilian counterparts is perhaps representative of theelles of multilateral trade negotiations. Northern protectionist interests have always played an adroit game, skillfully outflanking the legitimate developmental ambitions of the south. And while the Europeans put up a brave fight in the tournament, most states — as in past trade negotiations, notably the Uruguay Round — were soon hooked, cut and sent packing.
read more 

Africa at large: Aid For The Poor, Not For the Consultants
Source: Business Day (South Africa), by Nevisha Idene, 07.06.2006
Johannesburg — No less than a quarter of annual development aid — about 20 billion dollars — is being used by donor countries to fund technical assistance of sometimes dubious worth, says ActionAid Internationally in a new report.
read more 

Uganda: Museveni amnesty to Kony illegal- ICC
Source: Monitor (Uganda), by Frank Mpekha, 07.06.2006
THE International Criminal Court has reminded Uganda of her obligation to arrest Lord’s Resistance Army rebel leader Joseph Kony.
read more 

Uganda: Donors warn of more aid cuts
Source: Monitor (Uganda), by Elias Mbirikuna and Hussein Began, 07.06.2006
The government should prepare for more announcements of cuts in budget support, sources within the diplomatic community in Kampala have warned.
read more 

Kenya: President’s office perceived most corrupt
Source: Standard (Kenya), by Alex Kigege, 07.06.2006
It’s no official. The Office of the President is perceived to be the most corrupt in the land. A new survey by a government think tank, whose chief executive is the President’s close associate, has arrived after stunning shock. 
read more
MODS Subject Heading List - Darfur, 2006 Web Archive

Additional subject headings (catalogers may choose additional headings from among those listed below):

1. 151 #a Sudan #x History #y Civil War, 1983-2005
   <subject authority="lcsf">
   <geographic>Sudan</geographic>
   <topic>History</topic>
   <temporal>Civil War, 1983-2005</temporal>
   </subject>

2. 151 #a Sudan #x Politics and government #y 1985-
   <subject authority="lcsf">
   <geographic>Sudan</geographic>
   <topic>Politics</topic>
   <temporal>1985-</temporal>
   </subject>

3. 151 #a Sudan #x Economic conditions #y 1983-
   <subject authority="lcsf">
   <geographic>Sudan</geographic>
   <topic>Economic</topic>
   <temporal>1983-</temporal>
   </subject>

4. 151 #a Sudan #x Civilization
   <subject authority="lcsf">
   <geographic>Sudan</geographic>
   <topic>Civilization</topic>
   </subject>

5. 150 #a Public opinion
   <subject authority="lcsf">
   <topic>Public opinion</topic>
   </subject>
<mods xmlns="http://www.loc.gov/mods/v3" version="3.2">
  <titleInfo>
    <title>afrika.no: The Norwegian Council for Africa</title>
    <genre>Web site</genre>
    <originInfo>
      <dateCaptured encoding="iso8601" point="start">20060717</dateCaptured>
      <dateCaptured encoding="iso8601" point="end">20061120</dateCaptured>
    </originInfo>
    <language>
      <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
      <languageTerm authority="iso639-2b" type="code">nor</languageTerm>
    </language>
    <abstract>
    </abstract>
    <subject authority="keyword">
      <topic>afrika, africa, culture, development, economy, education, environment, health, politics, travel</topic>
    </subject>
    <subject authority="lcsh">
      <geographic>Sudan</geographic>
      <topic>History</topic>
      <temporal>Darfur Conflict, 2003-</temporal>
    </subject>
    <subject authority="lcsh">
      <topic>International relief</topic>
    </subject>
    <subject authority="lcsh">
      <geographic>Sudan</geographic>
      <topic>Economic conditions</topic>
      <temporal>1983-</temporal>
    </subject>
  </titleInfo>
  <relatedItem type="host">
    <titleInfo>
      <title>Crisis in Darfur, Sudan Web Archive, 2006</title>
      <location>
        http://hdl.loc.gov/loc.natlib/collnatlib.00000011
      </location>
    </titleInfo>
    <identifier>hdl:loc.natlib/mrva0011.0037</identifier>
    <note type="system details">www.afrika.no</note>
    <location>
    </location>
    <accessCondition>Access restricted to on-site users at the Library of Congress</accessCondition>
    <recordInfo>
      <recordCreationDate encoding="iso8601">20070516</recordCreationDate>
      <recordIdentifier source="dlc">mrva0011.0037</recordIdentifier>
    </recordInfo>
  </relatedItem>
</mods>
Displayed MODS Record - Darfur, 2006 Web Archive

**Title:** afrika.no: The Norwegian Council for Africa

**Abstract:** afrika.no - The Index on Africa and Africa News Update. Features news on and links to all countries in Africa. With sections on Culture, Development, Economy, Education, Environment, Health, Human Rights, News and Politics. By the Norwegian Council for Africa.

**Date Captured:** July 17, 2006 - November 20, 2006

[Archived Site]

**Subject(s):** Sudan--History--Darfur Conflict, 2003-
International relief
Sudan--Economic conditions--1983-

**Language(s):** English
Norwegian

**Genre:** Web site

**Access Condition:** Access restricted to on-site users at the Library of Congress.

**URL at time of capture:** www.afrika.no/

**Citation ID:** http://hdl.loc.gov/loc.natlib/mrva0011.0037

**Record ID:** mrva0011.0037

**Collection Title:** Crisis in Darfur, Sudan Web Archive, 2006
Web Archives Available:
- United States 107th Congress Web Archive
- United States 108th Congress Web Archive
- Crisis in Darfur, Sudan, Web Archive, 2006
- Library of Congress Manuscript Division Archive of Organizational Web Sites
- Papal Transition 2005 Web Archive
- September 11, 2001 Web Archive
- United States Election 2000 Web Archive
- United States Election 2002 Web Archive
- Visual Image Web Sites Archive

The Library of Congress Web Archives (LCWA) is composed of collections of archived web sites selected by subject specialists to represent web-based information on a designated topic. It is part of a continuing effort by the Library to evaluate, select, collect, catalog, provide access to, and preserve digital materials for future generations of researchers. The early development project for Web archives was called MINERVA.
Collection Overview

Citing Resources - Copyright Information

Scope: Described by the United Nations as the world’s worst humanitarian crisis, the current conflict in Sudan’s western region of Darfur began in early 2003, when rebels attacked government targets and then expanded to the widespread killing of civilians and the destruction of villages. These attacks have led to more than two million people being displaced and thousands killed. More than 10,000 aid workers, both Sudanese and international, are actively involved in assisting victims of this man-made disaster.

In the spring of 2004, the conflict began to receive broad international press coverage; scholars and development specialists produced in-depth studies and reports. Governments issued position papers and sent representatives to Sudan to see the situation firsthand. New organizations, each with Web sites, have been formed because of the Darfur crisis. Reminiscent of those seen during the 1970's and 1980's which opposed the oppressive apartheid system in South Africa, campaigns have begun urging U.S. states and universities to divest from companies doing business in Sudan and asking the public to petition its government officials to take action. Web sites and blogs have replaced many of the pamphlets, flyers, and newsletters used to disseminate information in an earlier era. A Web Archive of the Crisis in Darfur, Sudan, will preserve the documentation of this humanitarian crisis for future historians, and will include sites of key organizations, a sampling of news reports, and the responses of government, international organizations and the general public in the U.S. and worldwide.

This collection is part of a continuing effort by the Library of Congress to evaluate, select, collect, catalog, provide access to, and preserve digital materials for future generations of researchers.

Collection Period: March 20, 2006 - November 20, 2006
Browse by Subject

- Disaster relief
- Disinvestment
- Genocide
- Human rights
- Humanitarian intervention
- International relief
- Peacekeeping forces
- Public opinion
- Refugees
- Sexual abuse victims
- Sudan—Civilization
- Sudan—Economic conditions—1983-
- Sudan—History—Civil War, 1983-2005
- Sudan—History—Civil War, 1985-
- Sudan—History—Darfur Conflict, 2003-
- Sudan—Politics and government—1985-
Crisis in Darfur, Sudan, Web archive, 2006

Relevance: ★★★★★

LC Control No.: 2007700187

LCCN Permalink: http://lccn.loc.gov/2007700187

Type of Material: Book (Print, Microform, Electronic, etc.)

Main Title: Crisis in Darfur, Sudan, Web archive, 2006 [electronic resource].

Published/Created: Washington, D.C. : Library of Congress, 2007-

Related Names: Library of Congress.

Access Advisory: Descriptions of the archived Web sites are searchable without restriction; access to some archived sites is restricted to on-site users of the Library of Congress.

Summary: Selective collection of 216 Web sites, archived from March 20, 2006 to Nov. 20, 2006, relating to the humanitarian crisis in Darfur, Sudan. The archived sites are international in scope, and include those related to organizations involved in human rights, refugees, disaster relief, etc., government sites, research and public opinion group sites, sites of religious organizations, press organization sites involved in reporting on the Darfur crisis, and other miscellaneous sites.

Notes: Title from overview page, as viewed on Feb. 16, 2008.
"Library of Congress Web Archives, Minerva."

Subjects: Disaster relief --Sudan --Darfur.
Atrocities --Sudan --Darfur.
Disinvestment --Sudan --Darfur.
Genocide --Sudan --Darfur.
Human rights --Sudan --Darfur.
Humanitarian intervention --Sudan --Darfur.
International relief --Sudan --Darfur.
Peacekeeping forces --Sudan --Darfur.
Public opinion --Sudan --Darfur.
Refugees --Sudan --Darfur.
Sexual abuse victims --Sudan --Darfur.
Web archives.
Sudan --History --Darfur Conflict, 2003-
Sudan --History --Darfur Conflict, 2003--Atrocities.
Sudan --History --Darfur Conflict, 2003--Civil relief.
Google Search – Item in Darfur, 2006 Web Archive

afrika.no - Sudan: Darfur's lifeline radio (Comment)
afrika.no - Sudan: Darfur's lifeline radio (Comment)
afrika.no - The Institute for Africa and African Studies. Website for news and links to all countries in Africa. With sections on Culture, Development, ...
www.afrika.no/Details/13699.html - 24k - Cached - Similar pages

afrika.no - Darfur and Somalia: Two peace efforts differ... in Darfur, peace efforts have taken a multilateral approach, and with Sudan saying..., Norway
Phone: +47 23983312 Fax: +47 23983331 Email: afrika@afrika.no ...
www.afrika.no/Details/14370.html - 29k - Cached - Similar pages
More results from www.afrika.no...

afrika.no - The Norwegian Council for Africa (Library of Congress) ...
URL, at time of capture, www.afrika.no/ Record ID: mnr0011.0037. Collection Title: Crisis in Darfur, Sudan Web Archive, 2006 ...
lreveclean.gov.gl/afrika/nr0011.0037 - 7k - Cached - Similar pages

Search Results (All Collections, Library of Congress Web Archive) ...
5, Africafocus home page, From Crisis in Darfur, Sudan Web Archive, 2006 [ bibliographic information]; 5, afrika.no - The Norwegian Council for Africa ...
lreveclean.gov.gl/afrika/searchAll.php?collection=mnr0011.0037...History-Darfur%20Conflict%202003...
- 10k - Cached - Similar pages
More results fromreveclean.gov.gl...

Sud Afrika: No shoes
Sud Afrika: "Pray that God may open to us a door for the word, to declare the mystery of...
Disculpen que no se puedan ver las fotos en esta pagina. ...
gabyaslazar.blogspot.com/2007/05/no-shoes.html - 59k - Cached - Similar pages

Afrika News - Darfur: Is this ceasefire for real?
Written by Ernest Waditu. Editor-in-chief, Afrika News ... they were monitoring a paper ceasefire that had no resonance on the sun-baked ground in Darfur ...
afrikane.ws/index.php?itemid=38&id=43&option=com_content&task=view - 22k -
Cached - Similar pages

Afrika News - Legendary NBA Player Appeals for Help in Darfur
Mar 2, 2007 ... Written by Ernest Waditu, Editor-in-chief, Afrika News ... Why not Darfur, he wondered? He urged Americans to write letters to their ...
afrikane.ws/index.php?option=com_content&task=view&id=51&Itemid=35 - 25k -
Cached - Similar pages
More results from afrikane.ws ...

DARFUR: THE FIGHT FOR PEACE
by M DARFUR - Related articles
LC Web Archives – Levels of Access

- ILS OPAC
  MARC COLLECTION-LEVEL RECORD

- LUCENE SEARCH INTERFACE
  ARCHIVE-LEVEL HOMEPAGE & MODS RECORDS SEARCH/BROWSE
  - 107th Congress
  - 108th Congress
  - Election 2002
  - Election 2004
  - September 11, 2001
  - Olympics 2002
  - IraqWar 2003
  - Papal Transition 2005
  - Crisis In Darfur 2006
  - Egypt 2008
  - Legal Blawgs

- INTERNET SEARCH ENGINES

- MODS ITEM-LEVEL RECORDS

- W/ARC FILES ARCHIVED WEB SITES

- NUTCHWAX
  NUTCHWAX INDEXES
Results - Pros

• Archived resources are searchable and indexable along with other library collections and online resources

• Item-level and collection-level subject access and controlled vocabularies make these resources highly integratable at the item level and collection-level

• Site-level access facilitates searching and browsing within and across web archives—ability to find, refind & cite resources

• Good use and reuse of extracted and human-created metadata—friendly environment in which traditional catalogers learn XML and MODS—project benefits from specialized subject cataloger expertise

• Flexible and sustainable infrastructure for making web archives available for digital scholarship—stable/citable persistent IDS at the site level and the collection level
Results - Cons

- Scalability—approach works well with archives of up to 2,000 sites, but hasn’t been tested w/much larger archives

- Project investment is basically the same for each archive—whether it’s 100 sites or 2000 sites--project setup still requires template creation, metadata extraction, LCSH analysis at archive level, handle registration, etc.—so essentially the same amount of resources regardless of archive size
Future Considerations

• MODS tools—need for a flexible MODS input/editing form that would hide boilerplate and extracted metadata that the cataloger does not need to see—we have experimented w/XMLSPY’s Authentic and XForms, but we lose flexibility w/regard to parsed subjects with both of these

• Future plans to integrate the NutchWAX component to provide more comprehensive keyword access to W/ARC files—this will complement existing collection and site-level access

• Experiment tag cloud generators to increase subject keyword access
Tag Cloud Generated from Archived Web Site
Darfur, 2006 Web Archive
THAT’S ALL FOLKS

tmee@loc.gov