



# Cooperative Archiving: Event Harvesting in Perspective

Abbie Grotke, The Library of Congress



# Benefits of Cooperative Archiving

- React more quickly to rapidly unfolding events
- Archive more content
- Variety of partners bring their own expertise to table: subject, technical, languages
- Learn from others, share with others



# Examples

- September 11 Web Archive
- Hurricanes Katrina and Rita
- End of Term Government Archive (2008)
- Earthquake in Haiti
- Jasmine Revolution
- North Africa and Middle East (Arab Spring)
- Japanese Earthquake
- Olympics 2012

# End of Term 2008 Project

## Focus: US Government Websites

- Institutions: Library of Congress (LoC), California Digital Library (CDL), University of North Texas (UNT), Government Printing Office (GPO), Internet Archive (IA)
- Crawled: August 2008 – August 2009
- 2500+ seeds
- Bookend snapshots, weekly, monthly, quarterly crawls performed pre and post elections, pre and post inauguration
- ~25 TBs

### News from the Library of Congress

**Contact:** Jennifer Gavin, Public Affairs Office (202) 707-1940; Abbie Grotke, Office of Strategic Initiatives (202) 707-2833

**August 14, 2008**

#### **Library Partnership Preserves End-of-Term Government Web Sites**

The Library of Congress, the California Digital Library, the University of North Texas Libraries, the Internet Archive and the U.S. Government Printing Office today announced a collaborative project to preserve public United States Government web sites at the end of the current presidential administration ending January 19, 2009. This harvest is intended to document federal agencies' online archive during the transition of government and to enhance the existing collections of the five partner institutions.

As part of this collaboration, the Internet Archive will undertake a comprehensive crawl of the .gov domain. The Library of Congress has been preserving congressional Web sites on a monthly basis since December 2003 and will focus on development of this archive for the project. The University of North Texas and the California Digital Library will focus on in-depth crawls of specific government agencies. The project will also call upon government information specialists -- including librarians, political and social science researchers, and academics -- to assist in the selection and prioritization of web sites to be included in the collection, as well as identifying the frequency and depth of the act of collecting. The Government Printing Office will lend expertise to the curation process along with libraries in its Federal Depository Library Program. A tool has been designed by the project team and developed by the University of North Texas to facilitate the collaborative work of these specialists, and will be made available to participants in August 2008.

"Digital government information is considered at-risk, with an estimated life span of 44 days for a web site. This collection will provide an historical record of value to the American people," said Director of Program Management Martha Anderson of the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP).



# UNT Nomination Tool

End of Term Presidential Harvest 2008

[Other Projects](#) | [Help](#)

[Project Home](#)   [About This Project](#)

Number of URLs Nominated: 4622  
Number of Nominators: 29

☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆

### Search by URL

Search for an existing URL in the system.

☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆☆

### Browse URLs

#### COM

---

0 1 2 3 4 5 6 7 8 9

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

#### GOV

---

0 1 2 3 4 5 6 7 8 9

A



Jan 2, 2009



Jan 15, 2009

## End of Term Harvest U.S. Federal Government Websites 2008-2009



Jan 22, 2009



Jan 24, 2009

[Home](#)

[Project Partners](#)

[Search](#)

[Browse](#)

**Results:** 40 Items

Sorted by: [relevance](#) | [title](#)

Explore over 4,000 sites in this archive.

Browse by: [Facet](#) | [Title](#) | [URL](#)

Page 1 [2](#) [Next](#)

### Site lookup:

[Go](#)

Look up a site by keywords in the title, description or URL. To search the full text of sites, use the [Search](#) tab.

### Government branch

- [Executive](#) (21)
- [Legislative](#) (10)
- [Independent agency](#) (5)
- [Judicial](#) (1)

### URL segment

The following text appears in a section of the site's URL. This can help you narrow results to particular agencies.

- [nasa](#) (7)
- [senate](#) (6)
- [gsfc](#) (3)
- [house](#) (3)



**Title:** NIST Centennial Homepage

**Archival URL:** <http://crawls-wm.us.archive.org/eot08/20080916020757/100.nist.gov/>

**Coverage:** September 16, 2008 - August 14, 2009

[\[Show details\]](#) [\[Display site\]](#)



**Title:** Visualization of Remote Sensing Data

**Archival URL:** <http://crawls-wm.us.archive.org/eot08/20080916005400/rsd.gsfc.nasa.gov/>

**Coverage:** September 16, 2008 - August 14, 2009

**Description:** The Vis/RSD website is a showcase for stunning visualizations of satellite data by NASA's Goddard Laboratory for Atmospheres.

[\[Show details\]](#) [\[Display site\]](#)



**Title:** none provided

**Archival URL:** <http://crawls-wm.us.archive.org/eot08/20080916022923/aip.arsusda.gov/>

**Coverage:** September 16, 2008 - August 26, 2009

[\[Show details\]](#) [\[Display site\]](#)



**Title:** NASA Langley Atmospheric Science Data Center (Distributed Active Archive Center)

# End of Term 2012

## Focus: U.S. Government Websites and US National Elections 2012

- Institutions: Library of Congress (LOC), National Archives and Records Administration (NARA), California Digital Library (CDL), Harvard Libraries, Harvard Kennedy School (HKS) Library and Knowledge Services, University of North Texas (UNT), Government Printing Office (GPO), Internet Archive (IA)
- In planning stage
- Crawl Start: Fall, 2011



# What's Coming for 2012...

- Outreach to gov doc experts to help seed 2012 end of term collection
- Outreach to researchers whose specialty is in analysing elections to address candidates for federal offices







# Jasmine Revolution and Middle East Project Statistics

## Jasmine Revolution - Tunisia 2011

- Institutions: Library of Congress, Bibliothèque nationale de France (BnF)
- Started: Jan 19, 2011
- 513 seeds
- daily, weekly, monthly, bi-monthly crawling
- 17.7 million urls, 1.7 TB
- status: winding down (currently just bi-monthly crawling, with a few seeds at weekly)

## North Africa & the Middle East 2011

- Institutions: Library of Congress, BnF, British Library, American University in Cairo, Stanford University
- Started: Jan 27, 2011
- 2,020 seeds
- daily, weekly, monthly, bi-monthly crawling
- status: ongoing
- 35.5 million urls, 2 TB



Click here to watch a  
live stream of the  
latest developments  
in Egypt



News In Depth Programmes Video Blogs Business Weather Sport Watch Live

عربي

Africa Americas Asia-Pacific Central & South Asia Europe Middle East

Search



## Giant protest to kick off in Egypt

Thousands begin to assemble in  
downtown Cairo for the "million-  
man-march" aimed at forcing president  
Mubarak to resign.

Last Modified: 01 Feb 2011 07:04 GMT

Middle East

[Read More](#)

- Live blog Feb 1 - Egypt protests
- Live Messages from Egypt
- New cabinet sworn in
- Chaos reigns at Cairo airport
- Egyptian financial crisis looms
- Al Jazeera equipment seized

## Live Stream



ANGER





**NATIONAL DIGITAL  
INFORMATION INFRASTRUCTURE  
AND PRESERVATION PROGRAM**



# ALJAZEERA

latest developments  
in Egypt



News

In Depth

Programmes

Video

Blogs

Business

Weather

Sport

Watch Live

عربي

Africa Americas Asia-Pacific Central & South Asia Europe Middle East

Search



## Total internet blackout in Egypt

Country's last working internet service provider has been disconnected, completing a complete shut down of web access.

Last Modified: 01 Feb 2011 01:56 GMT

Middle East

[Read More](#)

- [Army vows not to use force](#)
- [Egypt under Israel's spotlight](#)
- [Egyptians urge march of millions](#)

## Live Stream





### Baheyya: Egypt Analysis and Whimsy بهية

Commentary on Egyptian politics and culture by an Egyptian citizen with a room of her own.

SATURDAY, FEBRUARY 12, 2011

**Fin du Régime**



عاجل  
سلامة: الرئيس مبارك قرر التخلي عن منصب رئيس الجمهورية ويكلف المجلس الاعلى للقوات المسلحة

In the end, they leave, with hollow eyes and a few plain words. Stripped of their ill-gotten power, they are miserable, ashen, and base. All of the rhetoric they spewed lingers like a bad smell, soon to evaporate in the fresh air of freedom. "The Egyptian people still need to develop a culture of democracy. Their grievances are economic, not political. The ruling party won a sweeping victory. The extremists are going to take over. The government supports limited income

**About Me**

My Photo **Name:** Baheyya بهية  
Baheyya is an Egyptian female name that has come to stand in for Egypt itself. The symbolism of course is the handwork of the gifted duo of Shaykh Imam Eissa and Ahmad Fu'ad Nigm in their haunting song, "Masr yamma, ya Baheyya." I make no foolish claims to represent Egypt or all Egyptians, I just like the name.

[View my complete profile](#)



**CROWDVOICE**

## TRACKING VOICES OF PROTEST

**EMERGENCY LAW AND POLICE BRUTALITY IN EGYPT**  
نقطة مسنودة للهيئة الشعبية في المن

**PROTESTS IN TUNISIA**  
**RIOTS IN ALGERIA**

**MOROCCAN OCCUPATION OF WESTERN SAHARA**


**VIEW MORE VOICES**

Find a Voice



**YouTube** Search Browse Upload Create Account Sign In

**Tahrir Caravan**  
EricGeoffrey 3 videos Subscribe



0:32 / 3:54 360p 2,661 views

Uploaded by EricGeoffrey on Feb 13, 2011  
I wanted to celebrate and honor the spirit, bravery, and determination of the

**Suggestions**

- DiGiorno® Pizza & Cookies by DiGiornoPizzas 154,744 views Promoted Video
- Protesters in Tahrir Square break into song by ScarceClips 623,024 views
- Egypt explodes in celebration as Mubarak steppi... by RussiaToday 90,406 views
- Egypt Revolution - Wave Your Flag by zaynabon 88,459 views
- Egypt Slide Show w/ Music by Dizzanya 37 views
- Congratulations To The People of Egypt by mibor07

**جبهة التغيير الشعبية**

الحرية له ثمن لا مر يعملون ارواحهم على كفه  
**FREEDOM IS ONLY FOR THOSE WHO ARE READY TO DIE**

Monday, 14 March 2011  
الجان الشعبية للمراقبة الاستثنائية

تحديث:  
صفحة على الفيس بوك... يالا اشركوا في الجان الشعبية للمراقبة الاستثنائية  
هو مين اللي حيراقب الاستفتاء؟  
هل منصوب على ان الاستفتاء ده عليه رقابة قضائية؟  
طب انا ما اعرفش.. ولو فيه رقابة احنا ما نعلمش بقنا لحد ابدأ  
والله لو جابوا لنا الدفتاتور براتب.. مش حاشق في حد  
احنا عايزين نعمل لجان شعبية لمراقبة الاستفتاء  
كل لجنة فيها على الاقل اثنين يراقبوا عملية الاستفتاء، من اول الصبح بدرى قوي، من اول ما الصناديق تيجي وهي فاضية، وطول عملية  
الاستفتاء وبتر تصوير العملية ده، وتفضل لإزالة الصناديق. كلك قيس وهي ليم لحد عملية الفرز

facebook

Adresse électronique

Mot de passe

Connexion

Garder ma session active

[Mot de passe oublié ?](#)

Inscription

Comité de Solidarité avec la Lutte du Peuple Egyptien est sur Facebook

Inscrivez-vous sur Facebook pour communiquer avec Comité de Solidarité avec la Lutte du Peuple Egyptien.



Comité de Solidarité avec la Lutte du Peuple Egyptien  J'aime

Mur Infos Photos Discussions Évènements Vidéo

Comité de Solidarité avec la Lutte du Peuple Egyptien + autres

Comité de Solidarité avec la Lutte du Peuple Egyptien uniquement

Seulement les autres

Informations

À propos de :  
Comité de Solidarité avec la Lutte du Peuple Egyptien /  
<http://egyptesolidarite.wordpress.com>

1 218 personnes aiment ça



Marylène Dany



Beatrice Nicolas



Hajer Bouden



Mhammed Henniche



Hichem Jaïche Achour



Peter Snowdon



Comité de Solidarité avec la Lutte du Peuple Egyptien



Manifestation de solidarité à la lutte du peuple égyptien

 Il y a 7 minutes · J'aime · Commenter

 Claire Boulette, Zenab Sémia Choug, Julie Cros et 7 autres personnes aiment ça.

 Mohàa D'unatremonde Meme pas sa me toutes les photo vous deconez lool  
Il y a 13 heures · Signaler



Comité de Solidarité avec la Lutte du Peuple Egyptien

[https://secure.avaaz.org/en/democracy\\_for\\_egypt\\_/?copy](https://secure.avaaz.org/en/democracy_for_egypt_/?copy)

Avaaz - STAND WITH THE PEOPLE OF EGYPT

# Managing Scope...

- The Web challenges our state and national boundaries and policies
- What's in? What's out?
- Need to define consistent selection and scoping criteria while the project and events develop.
- Tools such as the UNT Nomination Tool can help organize many people working on a project and the many URLs
- Who should take care of what is *in between* institutional boundaries - or *everywhere*?
- What is the risk? What is the value?