What is to be done?
Archive-it and the digital FDLP

James R. Jacobs [jrjacobs@stanford.edu]
Government Information Librarian
freegovinfo.info, @freegovinfo
Society of California Archivists Annual General Meeting ’13
April 12, 2013
agenda

• Problem statement: digital govt information
• Digital strategies for collecting the Web
• Ways to use Archive-it
• Serendipitous collaboration -- Fugitive documents
• What is to be done?
Government agencies across the board are publishing online with no clear methodology or mandate for long-term digital preservation. Memory organizations – archives, museums, libraries – are taking up the challenge of Web harvesting/preservation but in a largely scattershot way and usually with a very limited scope of responsibility.

23 state archives – missing many of the most populous states inc CA, NY, PA, IL, CT, FL, etc.

Only a few federal govt agencies like OSTI, Library of Congress and more recently GPO(!)

Almost ½ (98 of 245) of all archive-it partners are colleges and universities preserving their own sites.

Subject: Government -- 19 Collecting Organizations; 1051 Collections; 112,246 Sites [http://sn.im/govt-sites-archived]
digital strategy 1

- Drops: Everyday Electronic Materials (EEMs)


Language: English
Imprint: [Langley, VA : Central Intelligence Agency], 2004.
Format: Book
Note: 259 p. in various pagings : digital, PDF file.

Access: Original document classified "Top secret".

Summary: In January 2003, the DDO informed OIG that he had received allegations that Agency personnel had used unauthorized interrogation techniques with a detainee, Abd Al Rahim Al-Nashiri, at another foreign site, and requested that OIG investigate. Separately, OIG received information that some employees were concerned that certain covert Agency activities at an overseas detention and interrogation site might involve violations of human rights. In January 2003, OIG initiated a review of Agency counterterrorism detention and interrogation activities. This Review covers the period September 2001 to mid-October 2003.


searchworks.stanford.edu/view/8707791
snipurl.com/eems-report
digital strategy II

- Waterfalls: LOCKSS-USDOCS: harvest/preserve Fdsys.gov

[Image of LOCKSS website]

digital Federal Depository Library Program

The Digital Federal Depository Library Program, or the “USDocs” private LOCKSS network replicates key aspects of the United States Federal Depository System. The content is held in geographically distributed sites and replicated many times. Citizens have oversight and responsibility for the long-term care and maintenance of the content. All these

lockss-usdocs.stanford.edu
lockss.org
digital strategy III

- Oceans: Archive-it Web harvesting

archive-it.org/home/SSRG
Ways to use Archive-it

• Targeting document types (FOIA and CRS Reports)
• Targeting subject domains (Climate change and environmental policy)
• Serendipitous/collaborative harvesting (Fugitive documents)
Serendipitous collection

FDLP

Fugitive documents

GPO

lostdocs.freegovinfo.info

Fugitive documents archive-it collection
http://www.archive-it.org/collections/2361

Created using diagram.ly
Fugitive seeds

URL

http://naca.central.cranfield.ac.uk/
http://www.albany.edu/sourcebook/
http://www.albany.edu/sourcebook/index.html
http://www.census.gov/manufacturing/cir/index.html
http://www.census.gov/prod/www/abs/cfr.html
http://www.epa.gov/
http://www.gao.gov/
http://www.globalchange.gov/
http://www.nauticalcharts.noaa.gov/
http://www.nbii.gov/
http://www.noaa.gov/deepwaterhorizon/
http://www.state.gov/s/l/treaty/tias/index.htm
Government agencies across the board are publishing online with no clear methodology or mandate for long-term digital preservation. Memory organizations – archives, museums, libraries – are taking up the challenge of Web harvesting/preservation but in a largely scattershot way and usually with a very limited scope of responsibility.

23 state archives – missing many of the most populous states inc CA, NY, PA, IL, CT, FL, etc.

Only a few federal govt agencies like OSTI, Library of Congress and more recently GPO(!)

Almost ½ (92 of 223) of all archive-it partners are colleges and universities preserving their own sites.

Subject: Government -- 18 Collecting Organizations; 992 Collections; 4395 Sites [http://sn.im/govt-sites-archived]
many hands make light work

Courtesy of Flickr user Cindy47452 by-nc-sa cc license [http://sn.im/amishbarn]