



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

James R. Jacobs [jrjacobs@stanford.edu]
Government Information Librarian
freegovinfo.info, [@freegovinfo](https://twitter.com/freegovinfo)
Society of California Archivists Annual General Meeting '13
April 12, 2013



STANFORD UNIVERSITY LIBRARIES

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?



problem statement

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 I

- Drops: Everyday Electronic Materials (EEMs)

Counterterrorism detention and interrogation activities (September 2001–October 2003) [electronic resource] / Central Intelligence Agency, Inspector General.

<p>Online </p> <ul style="list-style-type: none">• graphics8.nytimes.com• purl.stanford.edu <p>Other libraries </p> <p> Find it at other libraries via WorldCat</p>	<p>Corporate Author: United States. Central Intelligence Agency. Inspector General.</p> <p>Language: English</p> <p>Imprint: [Langley, VA : Central Intelligence Agency], 2004.</p> <p>Format:  Book 259 p. in various pagings : digital, PDF file.</p> <p>Note: 2003–7123.IG At head of title: Special review. 7 May 2004 Sections have been redacted. Title from title screen (viewed on Aug. 25, 2009). Downloaded from http://www.archive.org/download/CounterterrorismDetentionAndInterrogationActivitiesSept2001Oct2003/ciaigreport2004_text.pdf on Aug 11, 2010.</p> <p>Access: Original document classified "Top secret".</p> <p>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. 2003.</p> <p>Format: Mode of access: World Wide Web. System requirements: Adobe Reader.</p> <p>Subjects: United States. Central Intelligence Agency > Rules and practice > Evaluation. Human rights > United States. Military interrogation > Government policy > United States. Detention of persons > Government policy > United States. Torture > Government policy > United States. War on Terrorism, 2001–2009 > Prisoners and prisons. Terrorism > United States > Prevention. Prisoners > Abuse of > United States.</p>
---	--

Catkey: 8707791 | [Compare in Socrates](#) | [Librarian View](#)

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



digital strategy II

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



LOCKSS LOTS OF COPIES
KEEP STUFF SAFE

[About LOCKSS](#) [The Community](#) [News & Media](#) [Technical Support](#) [How To Join](#) [Contact Us](#)

[Home](#) > [The Community](#) > [Global & Private LOCKSS Networks](#) > Digital Federal Depository Library Program

Digital Federal Depository Library Program

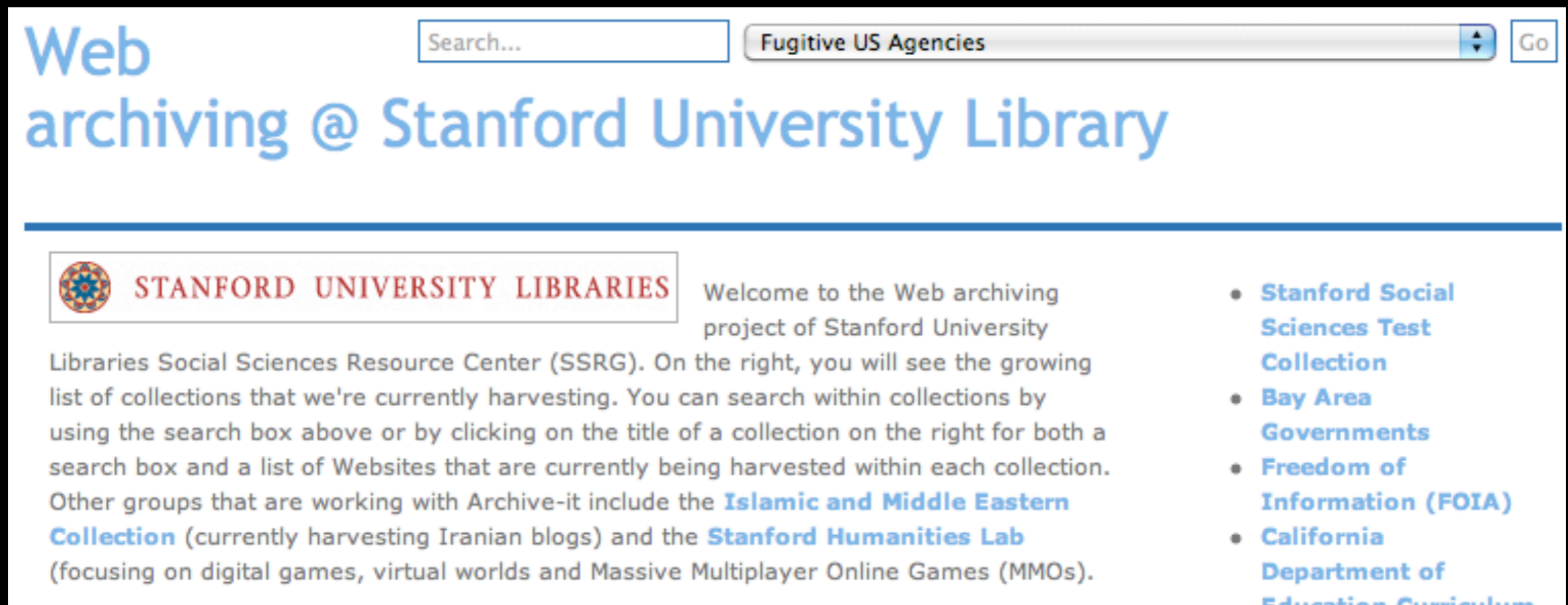
The Digital Federal Depository Library Program, or the "USDocs" private LOCKSS network replicates key aspects of the United States Federal Depository System. The content is held in geographically distributed sites and replicated many times. Citizens have oversight and responsibility for the long-term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

lockss-usdocs.stanford.edu
lockss.org




digital strategy III

- Oceans: Archive-it Web harvesting



Web archiving @ Stanford University Library

Search...

 **STANFORD UNIVERSITY LIBRARIES**

Welcome to the Web archiving project of Stanford University Libraries Social Sciences Resource Center (SSRG). On the right, you will see the growing list of collections that we're currently harvesting. You can search within collections by using the search box above or by clicking on the title of a collection on the right for both a search box and a list of Websites that are currently being harvested within each collection. Other groups that are working with Archive-it include the **Islamic and Middle Eastern Collection** (currently harvesting Iranian blogs) and the **Stanford Humanities Lab** (focusing on digital games, virtual worlds and Massive Multiplayer Online Games (MMOs)).

- **Stanford Social Sciences Test Collection**
- **Bay Area Governments**
- **Freedom of Information (FOIA)**
- **California Department of Education Curriculum**

archive-it.org/home/SSRG

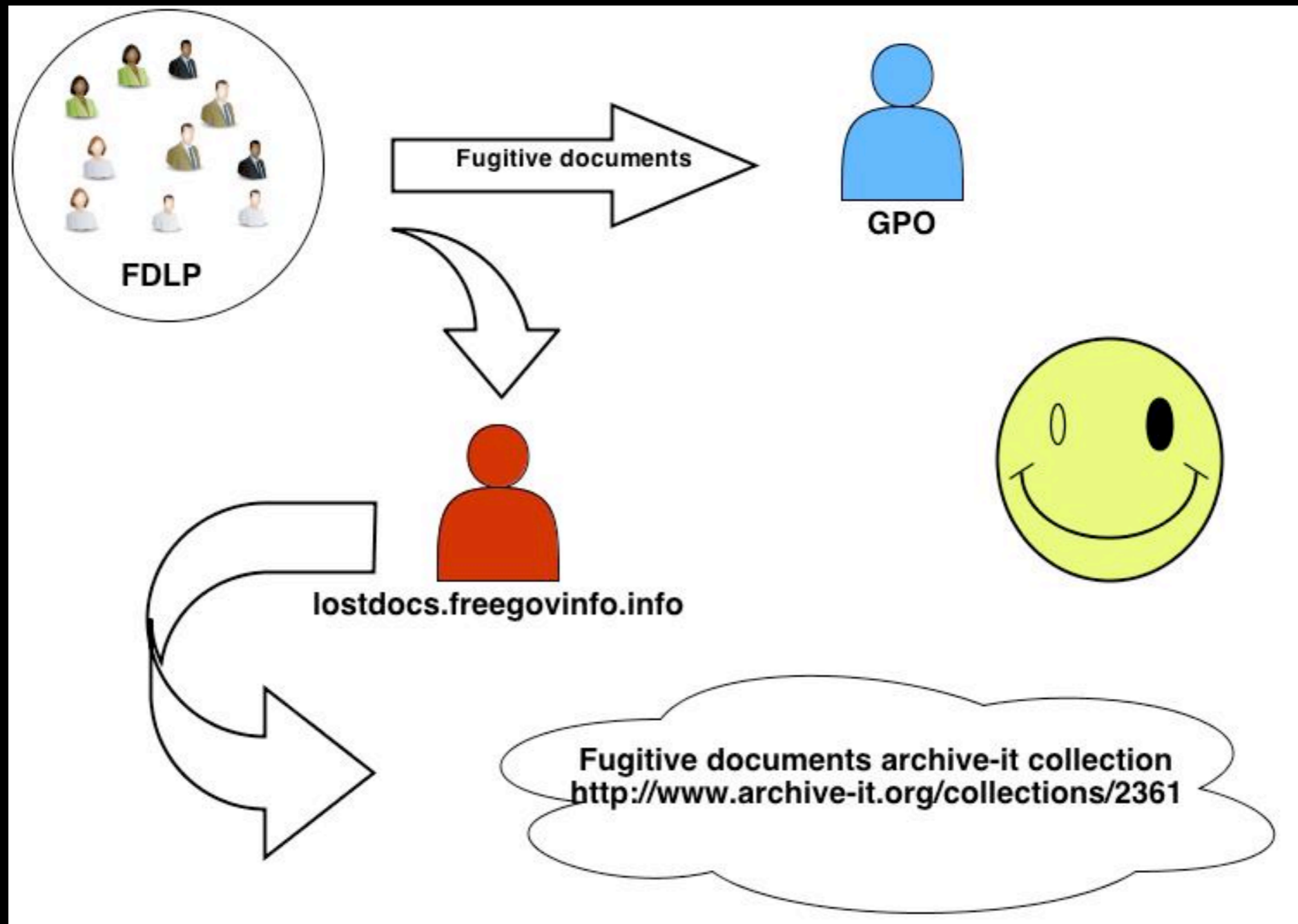


Ways to use Archive-it

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



Serendipitous collection



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/cffr.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>

<http://www.ustr.gov/about-us/press-office/reports-and-publications/>



problem statement

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>]



STANFORD UNIVERSITY LIBRARIES