

Help! I'm an Accidental Government Information Librarian

presents

Fugitive hunters

Community-Based
Digital Collection Development
of
Born-Digital
Government Information

with James A. Jacobs and James R. Jacobs

Part 1

Introduction, background,
terms, scale, focus

Notes, links, slides, etc. on FGI

FreeGovInfo.info/fugitives

Born-Digital U.S. Federal Government Information: Preservation and Access

- You can find links to the forum, the report, my speaker notes and more at freegovinfo.info/fugitives

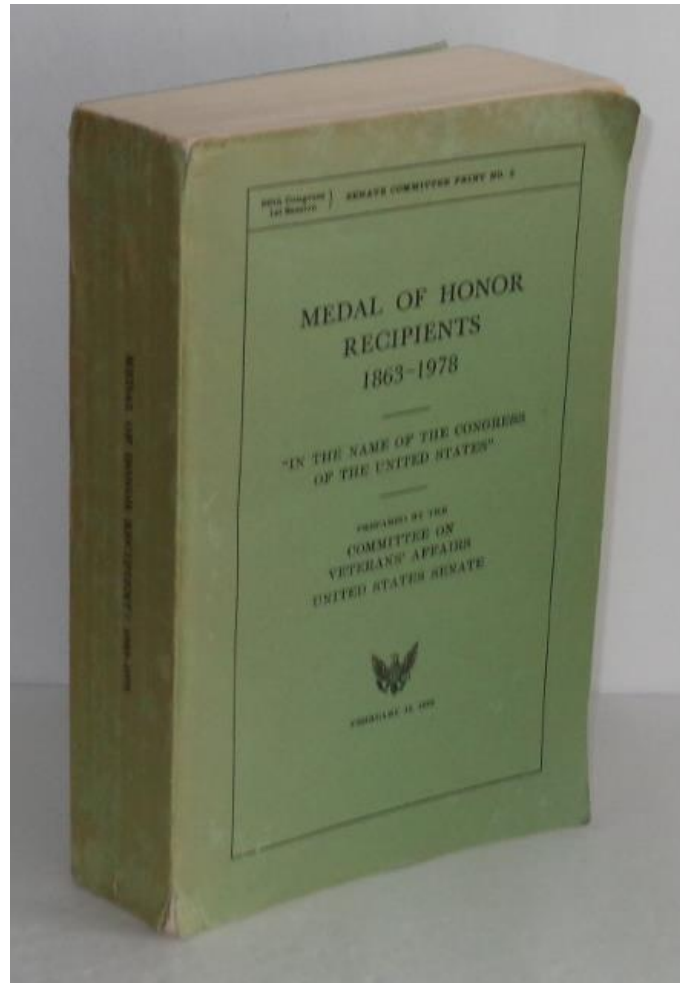
What is a Fugitive?

Any “document” that, according to the [44 USC](#) definition, *should* be part of the depository system, but that is not in the system.

Government Document: *“Informational matter” published at Government expense or as required by law*

Example

(not a new problem)



a “Committee Print”

Why a Fugitive?

- Administrative errors (GPO)
- Agency non-compliance with requirements (desktop publishing)
- Agencies publish with private sector and place publication under copyright (privatization)
- Agencies use a Title 44 exemption for publications “which must necessarily be sold in order to be self-sustaining”
- Agencies distribute information via agency web sites without notifying the FDLP

From: DiMario, Michael F. [Prepared Statement of Public Printer Before The Subcommittee On Legislative Branch Appropriations Committee On Appropriations U.S. Senate On Appropriations Estimates For Fiscal Year 1998.](#) (Jun. 5, 1997)

Examples

from DiMario (1997)

- Scientific and technical documents
(70% of tech docs: 55,000 in 1996)
- *Journal of the National Cancer Institute* (Oxford University Press)
- Court decisions
- Federal Election Commission financial disclosure statements
- CRS reports

Examples

from Bower (1989)

- *Air quality benefits of alternative fuels.*
- *Greenhouse effect, sea level rise and coastal wetlands.*
- *Reagan administration regulatory achievements*
- *A Report to the Secretary on the homeless and emergency shelters*
- *Report, Task Force on Women in the Military.*

Bower, Cynthia [Federal Fugitives, DND, and other Aberrants: a Cosmology.](#)
DttP v17 n3 (Sep 1989).

Not a new problem. Past Strategies.

- **Institutional.** The Library of Congress ran a service called DocEx from 1946-2004. Libraries could subscribe to this service and receive fugitives tracked down in D.C. (See [Thomas Shaw, 1966](#))
- **Technical.** GPO microfiches copies of documents printed elsewhere.
- **Individual.** Librarians captured fugitives by scouring agency newsletters and press releases and local newspapers and by cultivating agency contacts.
- **Legal.** Attempts to revise Title 44

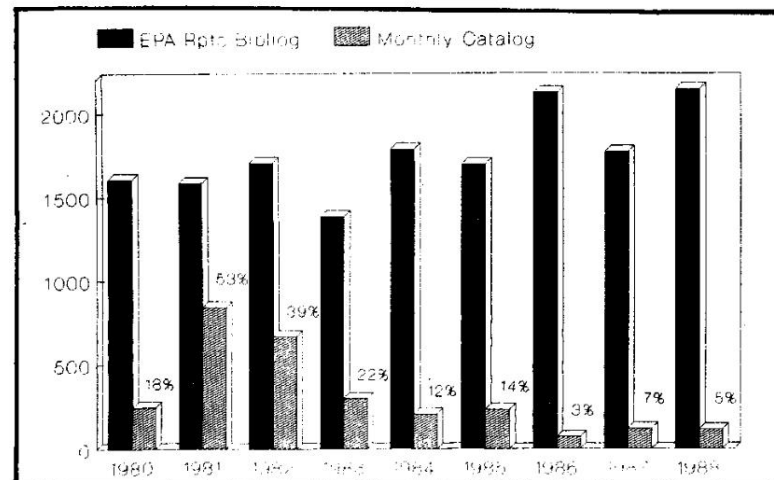
Scope

Bower (1980s)

Cynthia Bower study

- “No one knows”
- fugitive documents outnumber depository documents by an average ratio of **eight to one**
- **43%** of docs in American Statistics Index

Figure 3. Comparison of EPA Reports Cited in *EPA Publications Bibliography* With Those Cited in the *Monthly Catalog*, 1980-1988

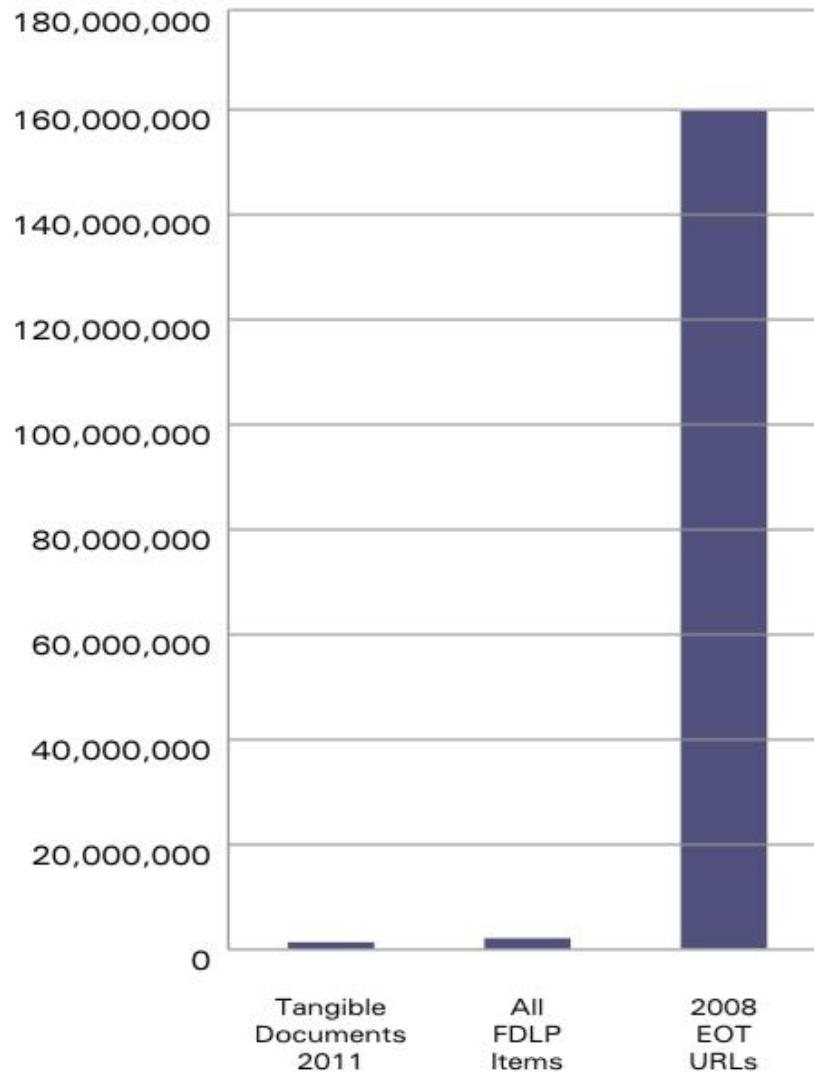


Scope

- About 50% of the universe of Federal printing
- 78% of NIH's publications

Baldwin, Gil. [Fugitive Documents - On the Loose or On the Run.](#)
Administrative Notes Vol. 24, no. 10 (August 15, 2003)

Scope



What is in FDsys?

- **Congress.** More than half of the [FDsys Collections](#) are explicitly Congressional
- **Courts.** Opinions of some, but not all, federal appellate, district, and bankruptcy courts, dating back to 2004
- **Executive.** Big gaps. (Apparently only 56 of 246 “government authors” in Fdsys are agencies)

Summary

- “Fugitives” are not a new problem.
- The number of fugitives is very, very large.
- The information content of fugitives is important
- There is no fool-proof strategy/solution to the fugitive problem, yet.


So...

In absence of
a universal solution,
every library can contribute
to
a loosely-coupled,
decentralized
strategy/solution.

Technical Notes

1. "Page" vs. "Title" vs. "File"

Example: Keystone XL Pipeline Project [Final Supplemental Environmental Impact Statement \(SEIS\)](#)



The screenshot shows a web browser window displaying the U.S. Department of State website for the Keystone XL Pipeline Project. The page title is "Final Supplemental Environmental Impact Statement (SEIS)". The header features the U.S. Department of State logo and the text "U.S. DEPARTMENT OF STATE KEYSTONE XL PIPELINE PROJECT". The main content area is titled "Final Supplemental Environmental Impact Statement (SEIS)" and includes a breadcrumb trail "Home > Final SEIS". A note states: "Note: If you have trouble viewing any of the PDFs on this site, please download the latest version of [Adobe Reader](#) or reload the latest version if it is already installed on your computer." The page lists several documents for download:

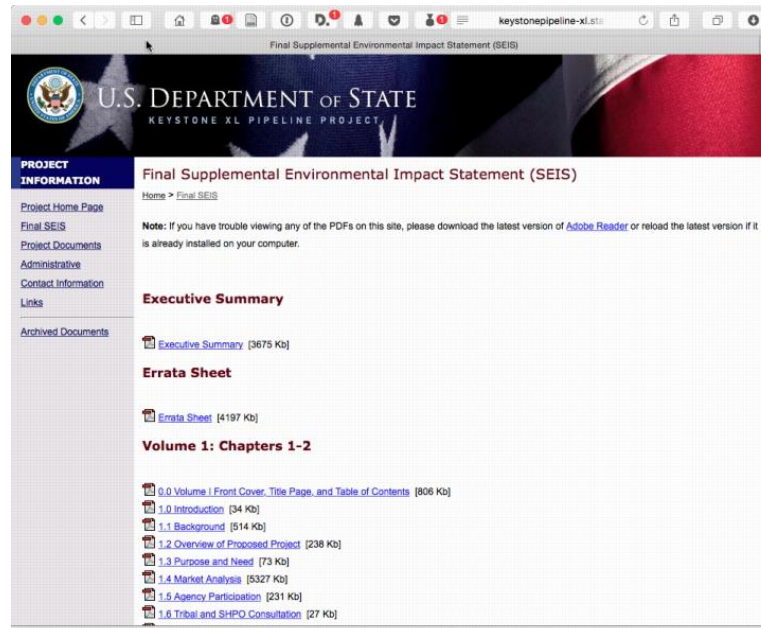
- Executive Summary**
 - [Executive Summary](#) [3675 Kb]
- Errata Sheet**
 - [Errata Sheet](#) [4197 Kb]
- Volume 1: Chapters 1-2**
 - [0.0 Volume I Front Cover, Title Page, and Table of Contents](#) [806 Kb]
 - [1.0 Introduction](#) [34 Kb]
 - [1.1 Background](#) [514 Kb]
 - [1.2 Overview of Proposed Project](#) [238 Kb]
 - [1.3 Purpose and Need](#) [73 Kb]
 - [1.4 Market Analysis](#) [5327 Kb]
 - [1.5 Agency Participation](#) [231 Kb]
 - [1.6 Tribal and SHPO Consultation](#) [27 Kb]

Technical Notes

1. “Page” vs. “Title” vs. “File”

To preserve this “Page,” you would need to preserve 13 Files (URLs).

- The HTML file that is this page at its URL, plus the other files that comprise the contents of this page:
 - 7 images files
 - 3 javascript files
 - 2 css style sheet files



Technical Notes

1. “Page” vs. “Title” vs. “File”

To preserve this “**Title**,” you would need to preserve

- 94 PDF **files** that comprise all the 11 volumes and all the parts of the Environmental Impact Statement.

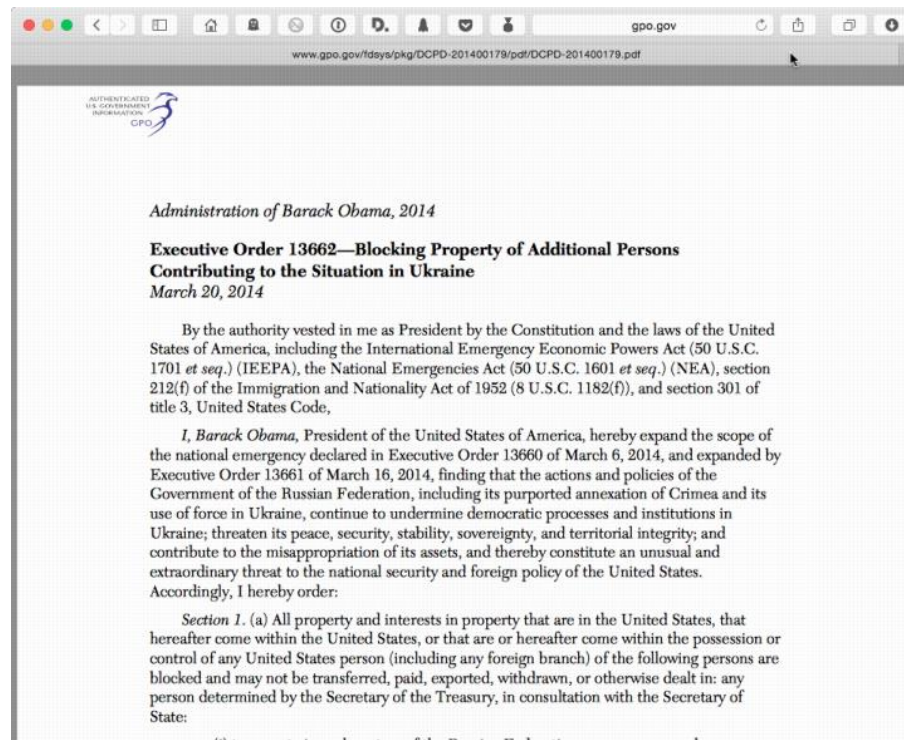


Technical Notes

2. Which version, edition, or copy?

Example: [Executive Order 13662](#) (Blocking Property of Additional Persons Contributing to the Situation in Ukraine)

- The GPO Authenticated version:



Technical Notes

2. Which version, edition, or copy?

Example: [Executive Order 13662](#) (Blocking Property of Additional Persons Contributing to the Situation in Ukraine)

- A total of ten different URLs have (apparently) the same information and metadata about the content.

The collage consists of ten overlapping browser window screenshots. The top-left screenshot shows the White House website with the title 'Executive Order 13662—Blocking Property of Additional Persons Contributing to the Situation in Ukraine' and the date 'March 20, 2014'. The top-middle screenshot shows the White House website with the title 'Executive Order -- Blocking Property of Additional Persons Contributing to the Situation in Ukraine'. The top-right screenshot shows the Presidential Documents website with the title 'Executive Order 13661 of March 16, 2014' and the subtitle 'Blocking Property of Additional Persons Contributing to the Situation in Ukraine'. The middle-left screenshot shows the Federal Register website with the title 'Blocking Property of Additional Persons Contributing to the Situation in Ukraine' and the subtitle 'A Presidential Document by the Executive Office of the President on 03/19/2014'. The middle-right screenshot shows the Federal Register website with the title 'Blocking Property of Additional Persons Contributing to the Situation in Ukraine' and the subtitle 'A Presidential Document by the Executive Office of the President on 03/19/2014'. The bottom-left screenshot shows the Presidential Documents website with the title 'Executive Order 13662—Blocking Property of Additional Persons Contributing to the Situation in Ukraine'. The bottom-middle screenshot shows the Presidential Documents website with the title 'Executive Order 13661 of March 16, 2014' and the subtitle 'Blocking Property of Additional Persons Contributing to the Situation in Ukraine'. The bottom-right screenshot shows the Presidential Documents website with the title 'Executive Order 13661 of March 16, 2014' and the subtitle 'Blocking Property of Additional Persons Contributing to the Situation in Ukraine'. The text of the executive order is consistent across all versions, starting with 'I, Barack Obama, President of the United States of America, hereby expand the scope of the national emergency...' and detailing the blocking of property of additional persons contributing to the situation in Ukraine.

Technical Notes

3. dot-gov, dot-mil, or beyond...?



The screenshot shows a webpage with a blue header and a background image of the White House dome. The main heading is 'Current Third-Party Pages'. Below it is a text box explaining that the list includes public-facing White House accounts on third-party websites, with a note that non-public accounts are used for testing and internal purposes. The page is divided into two main sections: 'Social Networks' and 'Current Third-Party Pages'. The 'Social Networks' section has two columns: 'Sites' and 'Accounts'. The 'Current Third-Party Pages' section has a sub-section for 'Social Networks'.

Current Third-Party Pages

Below is the current list of public-facing White House accounts on third-party websites. The White House may use non-public accounts for testing and other internal uses.

Social Networks

Sites	Accounts
-------	----------


Current Third-Party Pages






- Social Networks







ABCs of Fugitive hunting



Needles...

STANFORD UNIVERSITY LIBRARIES  My Account

searchWorks catalog     

 Back to results  Previous 5 of 134  Next  Cite  Send to 

] Increasing analytic capacity of state and local law enforcement agencies moving beyond data analysis to create a vision for change

LANGUAGE English.

PUBLICATION Washington, D.C. : BJA, June 2012.

PHYSICAL DESCRIPTION 1 online resource (20 pages) : color illustrations

available online

www.bja.gov
url.stanford.edu


Contributors

CONTRIBUTOR
[United States. Bureau of Justice Assistance](#), issuing body.

Subjects

SUBJECT
[Law enforcement](#) > [United States](#) > [Data processing](#).
[Data mining in law enforcement](#).

more options

 [Find it at other libraries via WorldCat](#)

...or Haystacks



data analysis law enforcement



Web

News

Images

Shopping

Videos

More ▾

Search tools

About 3,200,000 results (0.29 seconds)

Predictive Analytics - appuri.com

Ad www.appuri.com/ ▾

Try a 30-day Free Trial to stop losing your SaaS customers



Police Chief Magazine - View Article

www.policechiefmagazine.org/magazine/index.cfm?fuseaction... ▾

Connecting the Dots: **Data Mining** and **Predictive Analytics** in **Law Enforcement** and **Intelligence Analysis**. Colleen McCue, Ph.D., Program Manager, Crime ...

Data Analysis Tools: Law Enforcement - Bureau of Justice ...

www.bjs.gov/index.cfm?ty=datp&tid=7 ▾ Bureau of Justice Statistics ▾

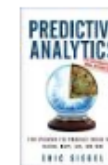
Nov 26, 2013 - **Data Analysis Tools: Law Enforcement**. Most of the spreadsheet files are in .csv format which can be easily read by most spreadsheet and ...

How 'Big Data' is helping law enforcement - PoliceOne

www.policeone.com/police.../Data.../6396543-How-Big-Data-... ▾ Police One ▾

Aug 20, 2013 - **Police** agencies are using **Big Data** to make connections and detect patterns ... IBM supercomputers help **law enforcement** gather, **analyze**, and ...

Shop for data analysis law enfo
Google



Predictive Analytics:
Click, Buy, Lie...

\$16.39 - Barnes & Nc
Always Ships Free or

Ads

Excel Data Analysis

www.tableausoftware.com/Excel ▾

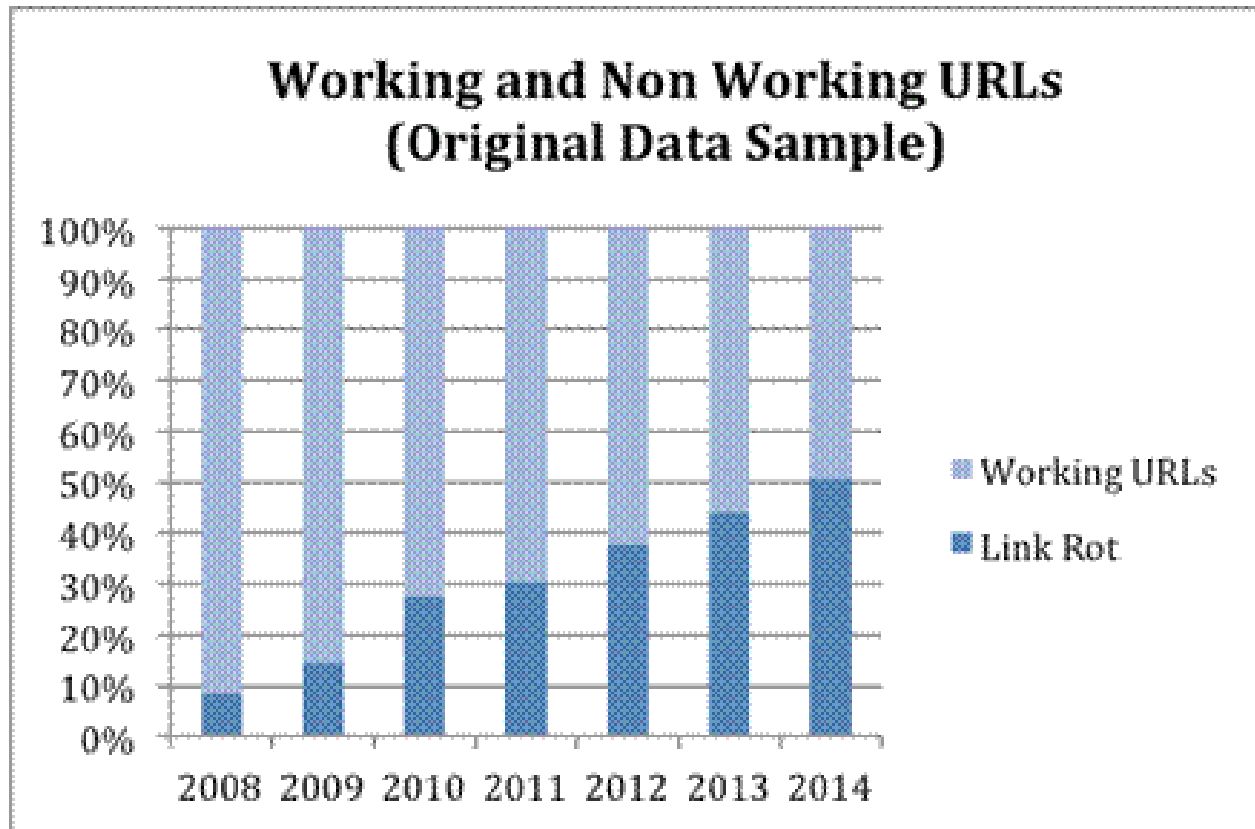
3.7 ★★★★★ advertiser rating

Visualize Your Excel **Data** for More
Analysis & Insights. Free Trial!

Data Analytics Tools

www.teradata.com/DataDriven ▾

Link rot



2014 Link Rot Report, Chesapeake Digital Preservation Group.
<http://bit.ly/2014-link-rot-report>

Drops, oceans & reservoirs



EEMs



Web Archiving



Flickr photo by Elle Is One
NonCommercial-ShareAlike
Commons license

photo by Black.Dots. CC BY-NC-ND 2.0.
<http://bit.ly/hetchhetchy-reservoir>

picture by discordia1967. That's actually me
falls in Kauai :-)

EEMs

- Everyday Electronic Materials
- Serendipitous fugitive collection
- Currently tracking 8 agencies (from lostdocs.freegovinfo.info) (HT Laura Lind!)
- Collecting the Web 1 drop at a time!
- CNI report on EEMs bit.ly/cni-eems

Increasing analytic capacity of state and local law enforcement agencies : moving beyond data analysis to create a vision for change

LANGUAGE

English.

PUBLICATION

Washington, D.C. : BJA, June 2012.

PHYSICAL DESCRIPTION

1 online resource (20 pages) : color illustrations

Available online

www.bja.gov
purl.stanford.edu

More options

[Find it at other libraries via WorldCat](#)

Contributors

CONTRIBUTOR

[United States. Bureau of Justice Assistance](#), issuing body.

Subjects

SUBJECT

[Law enforcement](#) > [United States](#) > [Data processing](#).

[Data mining in law enforcement](#).

Bibliographic information

PUBLICATION DATE

2012

Harvesting the ocean

Harvesting Websites since 2007 with Archive-it:

- Fugitive agency sites (GAO, EPA, NACA/NASA, Sourcebook of Criminal Justice Statistics, NOAA Deepwater Horizon Archive, Keystone EIS, FRUS, NBII etc)
- FOIA reading rooms and FOIA'd documents
- CRS Reports
- <https://archive-it.org/organizations/159>

Creating FDLP reservoirs



Flickr photo by Black.Dots. CC BY-NC-ND 2.0. <http://bit.ly/hetchhetchy-reservoir>

Creating FDLP reservoirs (II)

Things everyone can/should do:

- Keep track of your favorite agency's publications/data. Make sure those urls are in WayBack Machine!
- Submit fugitives to GPO's lostdocs form (and lostdocs.freegovinfo.info!)
- Save documents to local Web servers and/or upload to the Internet Archive ("seeding the cloud")
- Build Web-harvested collections that your local community wants/needs
- Join the "Everyday Electronic Materials" Zotero group and help us test out a newer, faster, more automatic fugitive document workflow!! (zotero.org/groups/everyday_electronic_materials)
- #FDLP IRC Channel to discuss fugitives and digital collection development

This is YOUR FDLP. Participate!!

Colophon

“...let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.”

— Thomas Jefferson, February 18, 1791