Capturing and Preserving the Federal Web for Permanent Public Access

Dory Bower, David Walls
Depository Library Conference, May 2, 2014

Why we harvest:

The Web is dynamic: "A system characterized by constant change, activity, or progress." "Something always changing."

- New pages appear without notice
- Changing content on existing pages
- "Link Rot": Complete removal of Websites without prior notice or changes in urls that make bookmarked content inaccessible.

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

GPO
GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Progress = Complexity

1997 US Census Bureau Website:

https://web.archive.org/web/19990117073605/http://www.census.gov/epcd/www/econ97.html

Basic HTML code structure with pdf reports. Everything from a single url.

2014 US Census Bureau Website: https://www.census.gov/

HTML 5, Java, Flash Media, content served on-demand from related databases, content from more than a dozen separate web addresses make up this single site.

GPO
GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Fulfilling GPO's Mission of Permanent Public Access to Government Information

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Government "Publications"

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

FDLP Web Archive

Our commitment:

- Ownership of all harvested content
- Cataloging
- Preservation
- Collaboration

Preserving Archived Content

- WARC file format- ISO 28500:2009
- Mirror site hosting on Archive-It “Wayback”
- Ability to move content
- File agnostic ingest design of FDsys
- Widespread knowledge among peers on preserving the Web and the software that runs it.

Collaboration

Peer Institutions

Archive-It Partners

International Consortium

GPO
GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Web Archiving Options

Decision process for Government Resource

- PURL: Individual publications and less complex web sites, using Teleport software
- Archive-It: Content rich websites
- Partnership: Hard to harvest sites, database sites or real time information

GPO
GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Collection Development

Develop and build website level collection

- Pilot sites: 3 sites to begin testing workflow
- Must be within scope of FDLP
- SuDoc Y3 sites: commissions, committees, independent agencies
  - Avoid duplication of effort with other institutions
  - Not distributing through print
  - Government information disseminated through web and content not cataloged
- All shaped collection development practices

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

Collection Development

Moving forward

- Y3s almost complete
- Working with Supervisor of Collection Development and staff with their extended experience to determine needs
- Non-standard sites (fatherhood.gov, read.gov)
- Look at agencies grouped by branch to determine amount of publications issued; move from smaller to larger sites
- Special collections
  - Native American resources

Collection Development

- Nominations
  - Document Discovery
  - AskGPO
  - Listserv announcements
  - fdlpwebarchiving@gpo.gov
- What is excluded:
  - What is in FDsys
  - Duplicative effort (already in Archive-It)

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

Archive-It Workflow

- Notification to Agency
  - Full disclosure of what we are doing
  - Chosen for inclusion into FDLP
- Seed list
  - Manual review of site
- Run and QA Test Crawl
  - Review all reports
  - Determine out of scope or missing content
  - Maximize collection, avoid unnecessary content

Archive-It Workflow

- Run and QA Archive crawl
  - Review carefully in Wayback checking all functionality
- Run patch crawls
- Submit lots of questions
- Maximize user access
- Best playback possible

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

Archive-It Workflow

- Cue catalogers to create metadata and access to CGP
- Process was completely changed, lessons learned, steps to develop best practice
- More information: webinar 3 of the Library Technical Services Webcast and Webinar Series

TITLE: Archiving & Cataloging Federal Agency Web Sites: GPO's Web Archiving Project

DATE: Friday, May 16th from 2-3 pm

Access

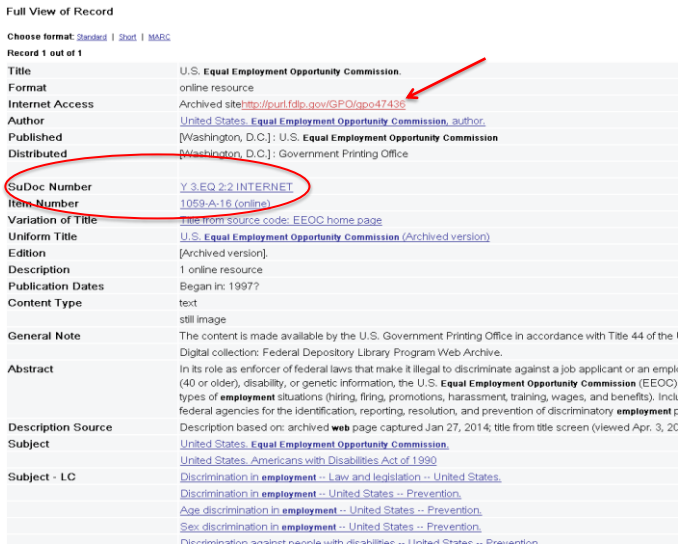
- Two locations for Access
  - Catalog of U.S. Government Publications (CGP)
    - Identifiable through "INTERNET" in SuDoc number
    - Expert search of wcat=web archiving retrieves all
  - Archive-It
    - Search for "GPO" or "FDLP"
    - Partner name is now "Federal Depository Library Program Web Archive"

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

Contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Catalog of U.S. Government Publications



Full View of Record

Choose format: Standard | Short | MARC

Record 1 out of 1

Title	U.S. **Equal Employment Opportunity Commission.**
Format	online resource
Internet Access	Archived site http://purl.fdlp.gov/GPO/gpo47436
Author	United States. **Equal Employment Opportunity Commission**, author.
Published	[Washington, D.C.] : U.S. **Equal Employment Opportunity Commission**
Distributed	[Washington, D.C.] : Government Printing Office
SuDoc Number	Y 3.EQ 2:2 INTERNET
Item Number	1059-A-16 (online)
Variation of Title	Title from source code: EEOC home page
Uniform Title	U.S. **Equal Employment Opportunity Commission** (Archived version)
Edition	[Archived version].
Description	1 online resource
Publication Dates	Began in: 1997?
Content Type	text
	still image
General Note	The content is made available by the U.S. Government Printing Office in accordance with Title 44 of the
	Digital collection: Federal Depository Library Program Web Archive.
Abstract	In its role as enforcer of federal laws that make it illegal to discriminate against a job applicant or an empl (40 or older), disability, or genetic information, the U.S. **Equal Employment Opportunity Commission** (EEOC) types of **employment** situations (hiring, firing, promotions, harassment, training, wages, and benefits). Inclu federal agencies for the identification, reporting, resolution, and prevention of discriminatory **employment** p
Description Source	Description based on: archived **web** page captured Jan 27, 2014; title from title screen (viewed Apr. 3, 20
Subject	United States. **Equal Employment Opportunity Commission.**
	United States. Americans with Disabilities Act of 1990
Subject - LC	Discrimination in **employment** -- Law and legislation -- United States.
	Discrimination in **employment** -- United States -- Prevention.
	Age discrimination in **employment** -- United States -- Prevention.
	Sex discrimination in **employment** -- United States -- Prevention.
	Discrimination against people with disabilities -- United States -- Prevention.

Contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

You are viewing an archived web page, collected at the request of Federal Depository Library Program Web Archive using Archive-It. This page was captured on 13:45:13 Jan 27, 2014, and is part of the U.S. Equal Employment Opportunity Commission collection. The information on this web page may be out of date. See All versions of this archived page. Metadata Enable QA

Español | Other Languages

U.S. Equal Employment Opportunity Commission

Enter search terms... Search

Home | About EEOC | Employees & Applicants | Employers | Federal Agencies | Contact Us

1 2 3 4

PERFORMANCE RESULTS

EEOC and OSC Renew Agreement Strengthening Enforcement of Federal Sector EEO Laws

Approved in FY 2013 under Strategic Plan Performance and Performance Measures

EEOC Issues FY 2013 Performance Report

Agency Limits Pending Inventory Increase; Secures Record Monetary and Injunctive Relief for Victims of Workplace Discrimination

Learn More

What You Should Know

EEOC and Employment Opportunities for Disabled Veterans

[Archived Topics]

How to ...

- File a charge of employment discrimination
- Check on the status of a charge
- Submit a Freedom of Information Act request
- Find help for small businesses
- Get training and technical

Newsroom

EEOC Sues Genesis Healthcare

Sedona Staffing of San Diego Settlement

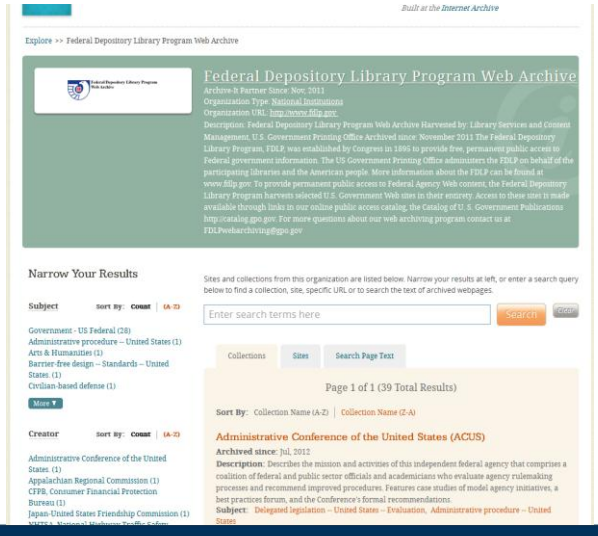
The EEOC settled charges of discrimination filed against Sedona Staffing of San Diego (Huyssen, Inc.) and the Sedona Group. If you applied for a job

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

GPO any specialized services contact GPO today.

Archive-It

https://archive-it.org/home/FDLPwebarchive



GPO any specialized services contact GPO today.

GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

Resources and Developments

- Adding to staff resources
- Staff who complete all steps from A – Z to streamline workflow and expedite completion of work
- Developing workflow for regular frequency of crawls

U.S. GOVERNMENT PRINTING OFFICE www.gpo.gov | www.fdsys.gov | http://www.fdlp.gov/ | http://bensguide.gpo.gov | http://bookstore.gpo.gov

GPO
GOVERNMENT INFORMATION AT YOUR FINGERTIPS
FEDERAL DEPOSITORY LIBRARIES

FDLP Web Archiving Team

- Mark Ames
- Dory Bower
- Liselle Drake
- Valerie Furino
- Fang Gao
- Stacey Kinsel
- Sunny Sanders
- Todd Scudiere
- David Walls
- Ben Webb

fdlpwebarchiving@gpo.gov