



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

**GPO Library Technical Services
Webcast and Webinar Series**

**Presented by GPO, Library Services &
Content Management, Library Technical
Services**

1



Presenters



Fang Gao



Liselle Drake



Agenda

- Web Archiving Project overview
- Metadata creation to provide access



Our Mission

Keeping America Informed:

to produce, protect, preserve and distribute
the official publications of the Federal
Government.



Web Archiving for the FDLDP





Web Archiving for the FDLP

- Manually harvest web versions of tangible documents from agencies' websites
- Capture full websites using Heritrix when the content is within scope
- Classify and catalog all harvested content



Overview of Web Archiving Project

- Began partnership in August 2011 with Archive-It
- Teamed with IIPC partners for guidance
- Developed a workflow to incorporate the use of Heritrix to capture entire websites
- Created metadata using RDA standard



Timeline

- Aug. 2011, a web Archiving Task Force formed in Library Services and Content Management (LSCM)
- Oct. 2012, outcome of work presented at the Fall DLC conference
- April 2013, LSCM update on harvesting, during Preservation Week
- May 2014, presentation given at DLC Conference
- As of May 12, 2014, 50 archived sites made available for access



Acquire Web Content by

- Identifying web versions of tangible documents acquired through the acquisitions workflow
- Doing outreach to identify fugitive agency publications
- Developing formal partnerships for formats not easily harvested



Steps for Archiving

- Scope determination
- Run test crawls
- Contact agencies
- Quality Review
- Make changes
- Run crawl
- Quality Control on “Wayback”



Access and Searching

- Catalog of U.S. Government Publications (CGP)
 - Expert search mode type:
wcat=web archiving
- Archive-It web site
 - “Find Collection by name” type: GPO



Expert Search: WCAT=Web Archiving

CATALOG OF U.S. GOVERNMENT PUBLICATIONS

Catalogs to Search:

[Congressional Serial Set](#) [Historic Shelflist](#) [Serials](#)
[Congressional Publications](#) [Internet Publications](#)

My Options:

[Bookshelf](#) [Results list](#)
[Preferences](#) [Previous Searches](#)

[Catalog of U.S. Government Publications Home Page](#) > National Bibliography of U.S. Government Publications

Expert Search: Full Catalog

Enter command language phrase





Search Results in CGP

Records 1 - 10 of 41 (maximum sort is 2000 records)

Jump to #

Previous Page

#	Title	Year	Author	SuDoc Number	Internet
1	OFR blog : the official blog of the Federal Register 2.0 project.	2011	United States.Office of the Federal Register,	AE 2.102:INTERNET	http://purl.fdlp.g
2	SaferProducts.gov	2010	U.S. Consumer Product Safety Commission,	Y 3.C 76/3:2 INTERNET/3	http://purl.fdlp.g
3	United States Election Assistance Commission.	2010	United States.Election Assistance Commission,	Y 3.EL 2/5:2 INTERNET	http://purl.fdlp.g
4	NARAtions : the blog of the United States National Archives.	2009	United States.National Archives and Records Administration,	AE 1.102:INTERNET	http://purl.fdlp.g
5	U.S. Nuclear Waste Technical Review Board.	2001	United States.Nuclear Waste Technical Review Board,	Y 3:N 88/2:INTERNET	http://purl.fdlp.g
3	The U.S. Office of Government Ethics.	200u	United States.Office of Government Ethics,	Y 3.ET 3:2 INTERNET	http://purl.fdlp.g



Archive-It: Search Collection=GPO

Archive-It - Web Archiving Services for Libraries and Archives - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Archive-It - Web A... x GPO Entire CGP Catalo... x Index of /gpo46349 x United States Holo... x Home | FoodSafet... x Inbox (6,551) - fa... x 王秀芬 美声唱法... x +

https://archive-it.org

Most Visited Bookmark on Delicious My Delicious Create PURL National Finance Cent... The Official TSP Home... GPO U.S. Government Prin... Welcome to BENEFED... >>

Search dictionaryBoss™ Define Word of the Day Spell Check Translate More >>

ARCHIVE-IT HOME EXPLORE LEARN MORE CONTACT US

A web archiving service to harvest and preserve digital collections a service of the Internet Archive

Welcome to Archive-It!
Attend a live informational webinar and demo to learn more about the service

Contact Us to sign up for an upcoming session:
Mar 04 2014, 11:30 AM PST
Mar 18 2014, 11:30 AM PDT

Explore Collections GPO Search Show All Collections

North Carolina State Government Web Site Archive
By North Carolina State Archives and State Library of North Carolina
The North Carolina State Government Web Site

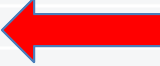
Arizona State Agencies
By Arizona State Library, Archives, and Public Records
The Arizona State Agencies collection contains content from the websites of

University of Iowa Archives
By University of Iowa Libraries
The University of Iowa Archives regularly captures the web site of the Iowa Writers' Workshop, the nation's premier creative

start A... 3 3 4 2 U... EN 10:16 AM



Archived Sites Examples

Title	Pool safety : simple steps to save lives.
Format	online resource
Internet Access	Archived site: http://purl.fdlp.gov/GPO/gpo34141 
Author	U.S. Consumer Product Safety Commission, author.
Published	[Bethesda, Md.] : [U.S. Consumer Product Safety Commission]
Distributed	[Washington, D.C.] : Government Printing Office
SuDoc Number	Y 3.C 76/3:2 INTERNET/4
Item Number	1062-C (online)
Variation of Title	Simple steps to save lives Steps to save lives
Edition	[Archived version].
Description	1 online resource
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 US Code 2013- Digital collection: Federal Depository Library Program Web Archive.
Abstract	Pool Safety reflects the U.S. Consumer Product Safety Commission's public education campaign to educate consumers, government officials, and industries of best practices to improve pool and spa safety to prevent child drownings, submersion injuries and entrapments. Site features resources for parents and families, including a Kid's Corner with educational activities, and recalled pool and spa products.
Description Source	Description based on: archived web page captured Apr 18, 2013; title from title screen (viewed Feb. 3, 2014).



Pool Safely: simple steps to save lives

You are viewing an archived web page, collected at the request of [Government Printing Office](#) using [Archive-It](#). This page was captured on 12:13:37 Apr 18, 2013, and is part of the [United States Consumer Product Safety Commission](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Videos Metadata](#)

The screenshot shows the Pool Safely website homepage. At the top left is the logo "Pool SAFELY" with the tagline "simple steps save lives". To the right is a navigation menu with links: Home | CPSC Home | About CPSC | RSS Feeds | Site Map | Contact Us. Below the logo is a search bar and buttons for "THE P&SS ACT" and "EN ESPAÑOL". A section titled "Recalled Pool & Spa Products" is partially visible. A central navigation bar contains buttons for: POOL & SPA SAFETY, FOR PARENTS & FAMILIES, FOR OPERATORS & INDUSTRY, FOR STATE & LOCAL OFFICIALS, FOR PARTNER ORGANIZATIONS, NEWS & MEDIA RESOURCES, and SHARE YOUR STORIES. The main content area features a graphic titled "Drowning Deaths and Injuries" showing a pyramid of children by age group: < 1 Year Old (2%), 1-3 Years Old (66%), 4 Years Old (11%), 5-9 Years Old (17%), and 10-14 Years Old (4%). Text next to the graphic states: "An average of 390 children ages 0-14 die in pool and spa drownings each year. About 5,200 more are treated for injuries in emergency rooms." Below this is a "Learn More" link and a link for "End of Summer Drowning Update | Español". To the right is a "Connect With Us" section with social media icons for Twitter, Facebook, YouTube, and Email. Below that is a "Kids' Corner" section with a cartoon cat character and a "2 ft" depth marker. At the bottom, there are buttons for "NEWS", "EVENTS", "TWEETS", and "VIDEOS". A date "November 5th, 2012" is visible on the left, and a section titled "Pool & Spa Safety Educational Videos" is on the right.



Videos

Pool Safely Videos



Drain Covers and Your Safety



Lifeguard Drain Cover Training
Got It Covered



Simple Steps to Safer Pools



Simple Steps Saves Lives Series



VGB Compliance Guides Series
Introduction



Pool Safely's PSAs available
in English and Spanish



Games

The Adventures of Splish and Splash

**THE ADVENTURES OF
SPLISH AND SPLASH**

Welcome to "The Adventures of Splish & Splash!" These child-friendly, interactive games were created to help families teach children about pool and spa safety.

You will meet Splish, a smart, sensible, adult-like character who knows about pool and spa safety, and Splash, an impulsive childlike character, who learns about safety through choices and consequences

Our objective is to show children that going to pools and spas requires

- The permission and accompaniment of parents or caregivers;
- The need to practice safe behavior and obey parents, caregivers and lifeguards and to pay attention to rules.

The videos are not meant to teach swimming but to show that it can be done with careful practice under the attentive watch of a parent or caregiver.

**GAME 1
SAFETY AT A
HOME POOL**

**GAME 2
SAFETY AT THE
PUBLIC POOL**

**GAME 3
SPLASH LEARNS
HOW TO SWIM**

Simple Steps Save Lives

Residential Pool & Spa Owners

Getting Involved

Education & Training

Resources for Parents & Families

Kids' Corner

The Adventures of Splish and Splash

GAME 1: Safety at a Home Pool

GAME 2: Safety at the Public Pool

GAME 3: Splash Learns How to Swim

FAQs

Connect With Us





United States Holocaust Memorial Museum

You are viewing an archived web page, collected at the request of [Government Printing Office](#) using [Archive-It](#). This page was captured on 12:59:34 八月 26, 2013, and is part of the [United States Holocaust Memorial Museum](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#)

UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM



Entire Site ▾

Search

SEARCH

MUSEUM EDUCATION RESEARCH HISTORY REMEMBRANCE GENOCIDE SUPPORT CONNECT

DONATE



You can help families learn what happened to loved ones during the Holocaust. **Start today.**

**WORLD
MEMORY
PROJECT** THE POWER OF TRUTH IS IN YOUR HANDS

1 2 3 4



United States Holocaust Memorial Museum Available in Many Languages

HOLOCAUST HISTORY

Introduction to the Holocaust
Holocaust Encyclopedia

Français

Español

Italiano

Русский

Türkçe

Português (BR)

عربي

فارسی

اردو

Ελληνικά

Bahasa Indonesia

简体中文

한국어

FEATURED



Alfred Rosenberg Diary Is Discovered

Missing for years, the diary of the Nazi leader will provide new insight into Holocaust history.

[MORE »](#)



CBS Sunday Morning Features Museum's 20th Anniversary

Segment highlights two decades of preserving memory and inspiring action.

[MORE »](#)



Some Were Neighbors

A new Museum exhibition and website explore the collaboration and complicity of countless individuals in Nazi Germany's persecution and murder of Jews.

[MORE »](#)



United States Holocaust Memorial Museum Available in Chinese

UNITED STATES
HOLOCAUST
MEMORIAL
MUSEUM

THE POWER OF TRUTH
20
YEARS
2013

犹太人大屠杀百科全书

有关词汇的注释

在第二次世界大战期间，平民的死亡人数达到了前所未有的数字。尽管许多人丧生于军事行动中，但还是有数以百万计的无辜受害者遭受到蓄意的灭绝。纳粹德国（由阿道夫·希特勒领导）在贯彻其种族主义政策的过程中，迫害并残杀了许多族群。而犹太人则是最大的受害族群。纳粹国家和其欧洲盟友实施了一系列有组织的种族灭绝屠杀，将近六百万犹太男女和儿童遭到了杀害。

“Holocaust”（反犹太屠杀）这一术语意指该种试图对某一人种实施的大屠杀。简略形式的“大屠杀”则专指纳粹残害整个犹太种族的行径。



NARAtions: the Blog of the United States National Archives

You are viewing an archived web page, collected at the request of [Government Printing Office](#) using [Archive-It](#). This page was captured on 16:19:43 一月 13, 2014, and is part of the [NARA blogs](#) collection. The information on this web page may be out of date. See [All versions](#) of this archived page. [Metadata](#)



NATIONAL ARCHIVES

RSS + Share | | 0

SEARCH

NARAtions: The Blog of the United States National Archives

Home > Blogs > NARAtions

SEARCH

Home

About NARAtions

About the Bloggers

Comment and Posting Policy



Ten Records We're Thankful to Have at the National Archives

by Mary (admin) on November 26, 2013

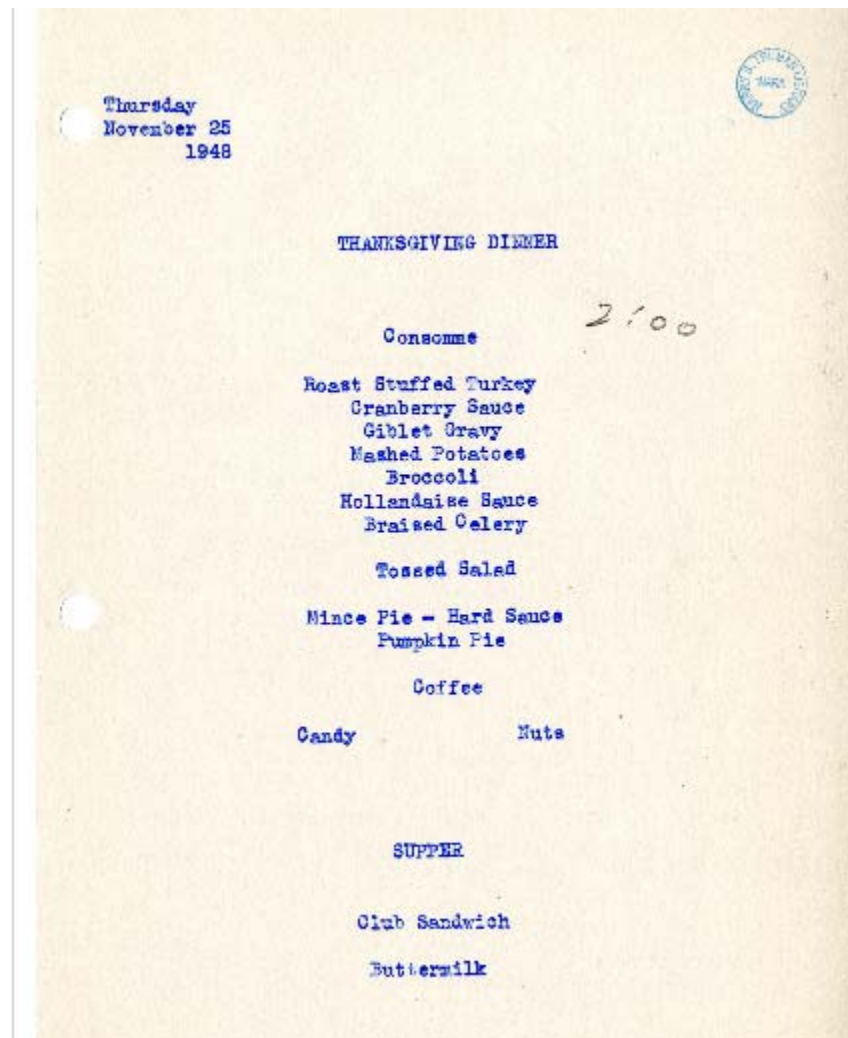
SUBSCRIBE TO EMAIL UPDATES

Your email:

22



White House Thanksgiving Menu, 11/25/1948. NARA ID 6882375





Metadata Creation



Cataloging Approach to the FDLP Web Archive Collection

- FDLP features, including PURLs and SuDocs
- RDA features, including Mode of issuance decision
- Project-specific features



GPO Introductory Webinar on RDA

An Overview of RDA and Its Implementation at GPO

<http://login.icohere.com/public/topics.cfm?cseq=1172>

Select from list under “FDLP Webinars”



PURL and Wayback URL

Archived site:

<http://purl.fdlp.gov/GPO/gpo34076>

URL:

http://wayback.archive-it.org/3287/*

<http://www.csb.gov/>



Review of GPO's Archiving Venues

- GPO's Permanent Server

<http://purl.fdlp.gov/docs/index.html>

- FDsys

<http://www.gpo.gov/fdsys/>

- Archive-It

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

CGP



Index of Harvested Instances of One Web Site



United States Holocaust Memorial Museum Web Archive (Government Printing Office)

Enter Web Address: All

Searched for <http://www.ushmm.org/>
[Look up URL](#) in general Internet Archive web collection

* denotes when page was updated

Found 8 Captures between Jul 18, 2012 - Aug 26, 2013	
2012	2013
7 pages	1 page
Jul 18, 2012 * Aug 15, 2012 * Aug 15, 2012 *	Aug 26, 2013 *



1st Cataloging Decision: Mode of Issuance

RDA 2.13: Modes of Issuance

single unit

multipart monograph

serial

integrating resource



1st Cataloging Decision: Mode of Issuance

Harvested instances of an updating web site
are cataloged as

an integrating resource.



FDLP Web Archive Collection Record

MARC Coding for fixed fields and 040 for IRs in RDA

Continuing Resources		<u>Rec stat</u>	c	<u>Entered</u>	20140203	<u>Replaced</u>	20140219115140.0
<u>Type</u>	a	<u>ELvl</u>		<u>Srce</u>	c	<u>GPub</u>	f
<u>Ctrl</u>		<u>Lang</u>	eng	<u>BLvl</u>	i	<u>Form</u>	o
<u>Conf</u>	0	<u>Freq</u>		<u>MRec</u>		<u>Ctry</u>	dcu
<u>S/L</u>	2	<u>Orig</u>		<u>Regl</u>	x	<u>Alph</u>	
<u>EntW</u>		<u>Desc</u>	i	<u>SrTp</u>	w	<u>Cont</u>	
<u>DtSt</u>	c	<u>Dates</u>	20uu				9999

040 GPO #b eng #e **rda** #c GPO

→NO #e pn



FDLP Web Archive Collection Record

074 |a 1070-K-01 (online)

0860 |a Y 3.H 74:2 INTERNET

1102 |a United States Holocaust Memorial Museum, |e author.

24510 |a United States Holocaust Memorial Museum.

➔ **250 |a [Archived version].**

264 1 |a [Washington, DC] : |b [United States Holocaust Memorial Museum]

264 2 |a [Washington, D.C.] : |b Government Printing Office

300 |a 1 online resource

500 |a The content is made available by the U.S. Government Printing Office in accordance with Title 44 of the US Code, 2012-

★ **500 |a Digital collection: Federal Depository Library Program Web Archive.**

520 |a A complete orientation to the services, collections, symposia, exhibitions, and publications ...

546 |a Text in English, French, Spanish, Italian, Russian, Turkish, Portuguese, Arabic, Farsi, Urdu, ...

588 |a Description based on: archived web page captured Aug 26, 2013; title from title screen ...

61020 |a United States Holocaust Memorial Museum.

650 0 |a Holocaust, Jewish (1939-1945) |x Museums |z Washington (D.C.).

650 0 |a Holocaust survivors |x Services for |z United States.

650 0 |a Holocaust survivors' families |x Services for |z United States.

650 0 |a Genocide |x Prevention.

★ **7102 |a Federal Depository Library Program, |e collector.**

7101 |a United States. |b Government Printing Office, |e distributor.

➔ **85640 |z Archived site: |u <http://purl.fdlp.gov/GPO/gpo27784>**



SuDoc Construction for Works in FDLP Web Archive

TD 8.2:INTERNET	NHTSA
TD 8.2:INTERNET/2	Distraction.gov
TD 8.2:INTERNET/3	NHTSA EMS
TD 8.2:INTERNET/4	911.gov
TD 8.2:INTERNET/5	Safercar.gov
Y 3.C 76/3:2 INTERNET	U.S. Consumer Product Safety Comm.
Y 3.C 76/3:2 INTERNET/2	ATVSafety.gov
Y 3.C 76/3:2 INTERNET/3	SaferProducts.gov
Y 3.C 76/3:2 INTERNET/4	Pool safely



FDLP Web Archive Collection Record

1101 |a United States. |b Department of Health and Human Services, |e author.

245 10Foodsafety.gov : ‡b your gateway to Federal food safety information.

246 1 Food safety information

247 10 www.foodsafety.gov ‡f <-2009?>

250 [Archived version].

362 1 Began in 1990s.

588 Description based on: archived web page captured Aug 22, 2013.

7101 |a United States. |b Department of Agriculture, |e author.

7101 |a United States. |b Food and Drug Administration, |e author.

7102 |a Centers for Disease Control and Prevention (U.S.), |e author.

7102 |a National Institutes of Health (U.S.), |e author.

7101 |a United States. |b White House Office, |e sponsoring body.

7102 |a Federal Depository Library Program, |e collector.

7101 |a United States. |b Government Printing Office, |e distributor.

77508 ‡i Current version: ‡t WWW.FoodSafety.gov ‡w (OCoLC)44286075



Reciprocal Linkage to Archived Version

245	0	0	WWW.FoodSafety.gov #h [electronic resource] : #b gateway to government food safety information.
246	3	0	FoodSafety.gov
246	3	0	Gateway to government food safety information
260			[Washington, D.C.] : #b FDA's Center for Food Safety and Applied Nutrition
362	1		Began in 1990s.
538			Mode of access: World Wide Web.
500			A gateway website that provides links to selected government food safety-related agencies. Topics include: news and safety alerts, consumer advice, kids, teens and educators, report illness and product complaints, foodborne pathogens, federal and state agencies, and selected highlights.
500			Title from title screen (viewed Apr. 29, 2009).
650	0		Food #x Safety regulations #z United States .
650	0		Food #z United States #x Safety measures .
650	0		Food additives #x Government policy #z United States .
650	0		Food handling #x Government policy #z United States .
710	2		Center for Food Safety and Applied Nutrition (U.S.)
775	0	8	#i Archived version: #a United States. Department of Health and Human Services. #t Foodsafety.gov #w (DLC) 2013230062 #w (OCoLC)864393010
856	4	0	#u http://www.foodsafety.gov



Archiving Decisions about Content

0860 |a Y 3.H 62/4:2 INTERNET/2

1102|a White House Historical Association, |e author.

24514|a The White House Historical Association. |p Resource & research library.

250 [Archived version].

300|a 1 online resource

336|a text |2 rdacontent ←

336|a still image |2 rdacontent ←

337|a computer |2 rdamedia ←

338|a online resource |2 rdacarrier ←

520|a Features a browsable database of images of portraits, landscapes, historical scenes, sculpture, and other works of art from **The White House Art Collection**. ...

The three hundred images presented currently comprise ...

650 0|a White House (Washington, D.C.) |v Pictorial works.

650 0|a Presidents |z United States |v Pictorial works.

650 0|a Presidents' spouses |z United States |v Pictorial works.

74002 |a White House art collection.



Example of Still Image Library



Abigail Powers Fillmore
Unknown Artist

[View details](#) 
[Add to Lightbox](#) 



Angelica Singleton Van Buren
Henry Inman

[View details](#) 
[Add to Lightbox](#) 



Anna Eleanor Roosevelt Roosevelt
Douglas Chandor

[View details](#) 
[Add to Lightbox](#) 



Barbara Pierce Bush
Herbert E. Abrams

[View details](#) 
[Add to Lightbox](#) 



Barbara Pierce Bush
Chas Fagan

[View details](#) 
[Add to Lightbox](#) 



Caroline Scott Harrison
Daniel Huntington



Claudia (Lady Bird) Taylor Johnson
Elizabeth Shoumatoff



Dolley Payne Madison
Gilbert Stuart



Dolley Payne Todd Madison
Mary E. E. Cutts, attr.



Edith Bolling Galt Wilson
Adolfo Müller-Ury



Decision-making for Authorized Access Point

1101 |a United States. |b Election Assistance Commission, |e author.

24500 |a United States Election Assistance Commission.

5050 |a Resources for voters -- Voting system testing and certification -- Election management resources -- Payments and grants -- Research and Data.

520 |a ... serves as a national clearinghouse ... Presents: resources for voting registration for all citizens, including military and overseas voters, and voters with disabilities and language assistance needs; an annual election calendar that tracks election dates, absentee ballot deadlines, voter registration deadlines and early voting dates for each state and the District of Columbia; voter guides and glossaries of election terminology in multiple languages; events and advisories about voting technology; and, a digital election resource library.

546 |a Text in English, Chinese, Japanese, Korean, Spanish, Tagalog, and Vietnamese; Voter's Guide to Federal Elections additionally in Cherokee, Dakota, Navajo, and Yupik.

61010 |a United States. |b Election Assistance Commission.

61010 |a United States. |t Help America Vote Act of 2002

650 0|a Voting |z United States.

650 0|a Elections |z United States.

650 0|a Voter registration |z United States.

73002 |a Voter's guide to federal elections.



Moving Forward

- Archiving treatment
- Frequency of crawls
- Bibliographic treatment
- SuDoc number construction
- Simultaneous cataloging of archived and live websites



Requests / Feedback

Web site recommendations for FDLP collection:

<http://www.gpo.gov/askgpo/>

[Select “Ask a question”]

Questions for Web Archiving Team:

fdlpwebarchiving@gpo.gov