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
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 FDLP
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
# Search Results in CGP

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

<table>
<thead>
<tr>
<th>#</th>
<th>Title</th>
<th>Year</th>
<th>Author</th>
<th>SuDoc Number</th>
<th>Internet</th>
</tr>
</thead>
</table>
Archive-It: Search Collection=GPO
## Archived Sites Examples

<table>
<thead>
<tr>
<th>Title</th>
<th>Pool safety: simple steps to save lives.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>online resource</td>
</tr>
<tr>
<td>Internet Access</td>
<td>Archived site: <a href="http://purl.fdlp.gov/GPO/gpo34141">http://purl.fdlp.gov/GPO/gpo34141</a></td>
</tr>
<tr>
<td>Author</td>
<td>U.S. Consumer Product Safety Commission, author</td>
</tr>
<tr>
<td>Published</td>
<td>[Bethesda, Md.]: [U.S. Consumer Product Safety Commission]</td>
</tr>
<tr>
<td>SuDoc Number</td>
<td>Y 3.C 76/3.2 INTERNET/4</td>
</tr>
<tr>
<td>Item Number</td>
<td>1052-G (online)</td>
</tr>
<tr>
<td>Variation of Title</td>
<td>Simple steps to save lives</td>
</tr>
<tr>
<td></td>
<td>Steps to save lives</td>
</tr>
<tr>
<td>Edition</td>
<td>[Archived version]</td>
</tr>
<tr>
<td>Description</td>
<td>1 online resource</td>
</tr>
<tr>
<td>Content Type</td>
<td>text</td>
</tr>
<tr>
<td></td>
<td>still image</td>
</tr>
<tr>
<td>Abstract</td>
<td>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, submergence injuries and entrapments. Site features resources for parents and families, including a Kids Corner with educational activities, and recalled pool and spa products.</td>
</tr>
<tr>
<td>Description Source</td>
<td>Description based on archived web page captured Apr 18, 2013; title from title screen (Viewed Feb. 3, 2014).</td>
</tr>
</tbody>
</table>
Pool Safely: simple steps to save lives

Drowning Deaths and Injuries

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.

Learn More

End of Summer Drowning Update | Español
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 Safety's PSAs available in English and Spanish
Games

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.
United States Holocaust Memorial Museum
Available in Many Languages

**HOLOCAUST HISTORY**

- Introduction to the Holocaust
- Holocaust Encyclopedia
  - Français
  - Español
  - Italiano
  - Русский
  - Türkçe
  - Português (BR)
  - العربية
  - 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 »**
猶太人大屠殺百科全书

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了蓄意的灭绝。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”（反犹大屠杀）这一术语意指该种试图对某一人群实施的大屠杀。简略形式的“大屠杀”则专指纳粹残害整个犹太种族的行径。
Ten Records We’re Thankful to Have at the National Archives

by Mary (admin) on November 26, 2013
Thursday
November 25
1948

THANKSGIVING DINNER

Consomme
Roast Stuffed Turkey
Cranberry Sauce
Siblet Gravy
Mashed Potatoes
Broccoli
Hollandaise Sauce
Braised Celery

Tossed Salad
Mince Pie - Hard Sauce
Pumpkin Pie

Coffee
Candy
Nuts

SUPPER

Club Sandwich
Butter milk
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”
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.fdrlp.gov/docs/index.html

• FDsys
  http://www.gpo.gov/fdsys/

• Archive-It
  https://archive-it.org/home/FDLPwebarchive
**Index of Harvested Instances of One Web Site**

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

<table>
<thead>
<tr>
<th>Entered Web Address:</th>
<th>http://</th>
</tr>
</thead>
<tbody>
<tr>
<td>Look up URL</td>
<td>in general Internet Archive web collection</td>
</tr>
</tbody>
</table>

* denotes when page was updated

**Found 8 Captures between Jul 18, 2012 - Aug 26, 2013**

<table>
<thead>
<tr>
<th>Date</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td></td>
</tr>
<tr>
<td>7 pages</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug 15, 2012 *</td>
<td></td>
</tr>
<tr>
<td>Aug 15, 2012 *</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug 26, 2013 *</td>
<td></td>
</tr>
</tbody>
</table>
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

040 GPO ǂb eng ǂe rda ǂc GPO  → NO ǂe 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]
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.

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/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 10 Foodsafety.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.


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.


77508 |i Current version: |t WWW.FoodSafety.gov |w (OCoLC)44286075
Reciprocal Linkage to Archived Version

<table>
<thead>
<tr>
<th>NUC</th>
<th>Fields</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>245</td>
<td>3 0</td>
<td><a href="http://WWW.FoodSafety.gov">WWW.FoodSafety.gov</a></td>
</tr>
<tr>
<td>246</td>
<td>3 0</td>
<td>FoodSafety.gov</td>
</tr>
<tr>
<td>246</td>
<td>3 0</td>
<td>Gateway to government food safety information</td>
</tr>
<tr>
<td>260</td>
<td>1</td>
<td>[Washington, D.C.] : #b FDA's Center for Food Safety and Applied Nutrition</td>
</tr>
<tr>
<td>362</td>
<td>1</td>
<td>Began in 1990s.</td>
</tr>
<tr>
<td>538</td>
<td>1</td>
<td>Mode of access: World Wide Web.</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>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.</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>Title from title screen (viewed Apr. 29, 2009).</td>
</tr>
<tr>
<td>650</td>
<td>1</td>
<td>Food #x Safety regulations #z United States.</td>
</tr>
<tr>
<td>650</td>
<td>1</td>
<td>Food #z United States #x Safety measures.</td>
</tr>
<tr>
<td>650</td>
<td>1</td>
<td>Food additives #x Government policy #z United States.</td>
</tr>
<tr>
<td>650</td>
<td>1</td>
<td>Food handling #x Government policy #z United States.</td>
</tr>
<tr>
<td>710</td>
<td>2</td>
<td>Center for Food Safety and Applied Nutrition (U.S.)</td>
</tr>
<tr>
<td>775</td>
<td>8</td>
<td>Archived version: #a United States. Department of Health and Human Services. #t Foodsafety.gov #w (DLC) 20132300062 #w (OCoLC)864393010</td>
</tr>
<tr>
<td>856</td>
<td>4 0</td>
<td><a href="http://www.foodsafety.gov">http://www.foodsafety.gov</a></td>
</tr>
</tbody>
</table>
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 rdaccontent
336 |a still image |2 rdaccontent
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

<table>
<thead>
<tr>
<th>Name</th>
<th>Artist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abigail Powers Fillmore</td>
<td>Unknown Artist</td>
</tr>
<tr>
<td>Angelica Singleton Van Buren</td>
<td>Henry Inman</td>
</tr>
<tr>
<td>Anna Eleanor Roosevelt Roosevelt</td>
<td>Douglas Chandor</td>
</tr>
<tr>
<td>Barbara Pierce Bush</td>
<td>Herbert E. Abrams</td>
</tr>
<tr>
<td>Barbara Pierce Bush</td>
<td>Chas Fagan</td>
</tr>
<tr>
<td>Caroline Scott Harrison</td>
<td>Daniel Huntington</td>
</tr>
<tr>
<td>Claudia (Lady Bird) Taylor Johnson</td>
<td>Elizabeth Shoumatoff</td>
</tr>
<tr>
<td>Dolley Payne Madison</td>
<td>Gilbert Stuart</td>
</tr>
<tr>
<td>Dolley Payne Todd Madison</td>
<td>Mary E. E. Cults, attr.</td>
</tr>
<tr>
<td>Edith Bolling Galt Wilson</td>
<td>Adolfo Müller-Ury</td>
</tr>
</tbody>
</table>
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.
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