

“The FDsys” - The New GPO Access
November 4th, 2010
Philadelphia Chapter of the Special Libraries Association
Ashley Dahlen, Outreach Librarian, U.S. Government Printing Office

What is FDsys?

FDsys will automate the collection, management and dissemination of electronic information from all three branches of government.

- **Content Management System:** FDsys securely controls digital content throughout its lifecycle to ensure content integrity and authenticity
- **Preservation Repository:** FDsys follows archival system standards to ensure long-term preservation and access of digital content, built on OAIS model
- **Advanced Search Engine:** FDsys combines extensive metadata creation with modern search technology to ensure the highest quality search experience

Where content will come from

- Submitted directly into FDsys:
 - Provide Congressional Content Originators with an interface to electronically submit Congressional bills, letterhead, envelopes, and other ephemeral material to GPO, along with metadata related to the content
 - Currently working with Congressional Publishing Services on Congressional Content Workflows
- Harvested from official Federal Web sites
- Converted digital content from print publications:
 - Pending approval from the Joint Committee on Printing (JCP)
 - Release 2 (begins January, 2011) will work on streamlining the workflow of ingesting converted content. Pilot collection will be the Coastal Zone Information Center

Digital preservation

- Safeguard digital content and relevant metadata
- Assess the condition and needs of collections of digital information
- Meaningfully render content despite continually changing technology

Continuity of Access (COA)

- The public will have access to content should something happen to the main system. The System Continuity of Access (COA) site is operational:
 - COA failover tests successfully completed and Authority to Operate (ATO) has been granted
 - With COA in place, GPO can maintain public access to content in the event of an emergency

Continuity of Operations (COOP)

- Full system failover (all FDsys components) planned to be operational in December 2010:
 - Plans are underway to establish an offsite backup for FDsys
 - Site will provide Continuity of Operations (COOP) for data management and access

Transition tasks completed to date

- Data migration process is completed, all 40 content collections completed
 - ~100 individual publications identified on GPO web servers that are being worked on because of special features (e.g. 9/11 report)
 - Less than 1% of content remains to be processed
- Introduced significant improvements during migration
 - Automated metadata extraction process for content collections developed to support search and preservation
 - Data inconsistencies handled by parser and metadata editor
 - Digital signatures added to PDF files in content collections

Remaining tasks to complete before GPO Access sunset, December, 2010

- Key activities have been identified and a completion plan has been developed:
 - GPO Access redirects will be in place
 - Getdoc link redirects will be in place
 - Single publications found on GPO Access (“Quarantined” packages), will be processed
 - Technical planning for GPO Access shutdown
- Once completed, FDsys will assume role as GPO’s electronic system of record

Impact of sunset on you

- Libraries need to update their literature, presentations, web pages
- New promotional material will be available for Federal depository libraries to display and handout
- Learn about FDsys:
 - User Training Manual (PDF) under [About Government Publications](#)
 - Tutorials and GPO presentations available under [For more information on FDsys](#)

FDsys is making a difference

- FDsys content is available in all major search engines
- Citizens are finding new ways to interact with key public content through XML data
 - Includes sites like FedThread.org and govpulse.us
- Other government sites rely on FDsys to enrich their user experience
 - FR 2.0 (federalregister.gov) developed by OFR and GPO as an XML-based edition of the daily Federal Register
 - Regulations.gov uses FDsys metadata to power their public search, eliminating the need for OFR support
 - Science.gov is able to integrate GPO collections into the federated search gateway to government science information

Comparison of GPO Access and FDsys

	GPO Access	FDsys
Document format/renditions	Text and PDF	HTML, PDF, TXT, and XML
Metadata	Index of title	MODS (descriptive metadata), PREMIS (preservation metadata), METS (list of all files for a content package)
Authentication Note: the authoring agency must give permission to authenticate content	Available for package (entire document) for select titles Not available for granule (individual part of a document)	Available for package (entire document) Available for granule (individual part of a document, e.g. an individual FR notice)
Search technology	WAIS search technology, developed by "Thinking Machines Inc." in 1988 for indexing and searching document indexes.	FAST search engine, developed by Fast Search and Transfer, in 1999, now a Microsoft Subsidiary. It is metadata driven searching.
Search options	Search, Browse, Search across collections	Search, Browse, Advanced search, Citation search, Refine and narrow results by filters, Sort search results
URLs	Cannot generate static link, URLs are session-generated	Static URLs created
Bookmark	Can bookmark pages	Can bookmark pages & searches (will automatically re-run a search when you access the bookmark)

Bulk download

- XML format
- Collections available:
 - Federal Register
 - CFR
 - Supreme Court Decisions (1937-75) (txt. format)
 - Future plans for Public Papers of the President (starting with the Obama administration)
- Located under:
 - **Latest Resources** on the homepage (will be there for a while)
 - Next to the collection title
 - Collection title and then look under **Related Resources** (left side),
 - Some also available on on www.data.gov
- Under the **Main Directory > resources** you will see the:
 - schema (.xsd),
 - the style sheet (.xsl, to make it pretty),
 - user guide (contains a FAQ, including answers to the legal status of the XML)

- LOCKSS statement
- Note: converted content may not have graphics
- Can download all, a month, a day...
- Check out www.federalregister.com or www.govpulse.us = plan on teaching your users about authenticity, which is only available on FDsys!!!

E-CFR

- Office of the Federal Register (OFR) and GPO are committed to adding the e-CFR to FDsys
- It is expanding to be an archive of a “daily snapshot” of the CFR
- E-CFR will be available until the new version is available
- A vision document from the OFR is due in November

Future Options for Users:

- Create an account, add on alert services (like RSS feeds), create your personal list of frequently used tools, help features, etc.
- “Shopping cart” or “my folder” to save search results
- Save your search queries
- Tailoring of search results:
 - Highlight your search terms
 - Analyze search results
- Enabling of mobile devices
- Search API
- Linking between documents (i.e. a Public Law that references the U.S. Code, will link to the U.S. Code citation).

Authentication

Content authentication will ensure:

- Content has been approved by, contributed by, or harvested from an official source
- Content has been verified by GPO to be complete and unaltered

What your user needs to see to know that the document is official and unchanged



The Blue Ribbon icon indicates that the certification is valid. It affirms that that the document has not been modified since it was certified and the digital signature is valid. This will only appear if the document was certified authentic by GPO and if the user’s computer can verify with GeoTrust that the signature is valid.

What is going on in the back end (the simple explanation)

Step One: GPO receives the content package from the agency or congress, processes it, applies a digital signature, affirming that the content has not changed since it was authenticated and disseminated by GPO, and then uploads the content into FDsys.



When users open up the PDF, they will see the GPO Seal of Authenticity on the document

Step Two: The User opens up the PDF file in FDsys, a saved version, email, etc. Their computer recognizes the certificate, checks the document to see if any changes were made, and if none have been, then it reports that the GPO signature and certificate are valid.

For more information, please visit [About Government Publications, Introduction, Authentication of Information](#)

Authentication icons

Adobe V8	Adobe V9	Significance	Status	Cause
		Certified	Verified. Certification signature.	Document has not changed or only contains permitted changes.
		Valid, with changes	Verified. Approval signature.	Permitted but signed changes exist.
		Unknown	Check has not executed.	Integrity check has not executed. This icon may appear if a user is not connected to the Internet or if a digital certificate is not available on a user's computer.
		Invalid	Signer's certificate expired, or revoked at the time of signing.	Illegal changes made, document corrupted, or policy restrictions violated.

***Hint:** You can always click on a symbol if you don't know what it means!

A note about the GPO Seal of Authenticity and printing...

The GPO Seal of Authenticity will be visible in a print out. However, the concept of “authentication” only applies to the digital version – NOT the print out. **GPO cannot guarantee the integrity of a print out of an authenticated online GPO product.**

This is analogous to the print realm. A print volume from GPO is an official government publication. A photocopy of an official print version is only a copy of an official version.

Authentication and Emailing of PDFs

Your users can save and email an authenticated PDF. The PDF will remain authenticated. This will make it difficult to mark up, edit, or use any of the commenting features of Adobe.

How to use mark-up, editing, and commenting features of Adobe PDFs

Summary of steps: extract the pages, save as a new file, and highlight and make notes on the new file.

- 1) Save the file to your computer.
 - 2) You may need to change the file name to something else so that you don't overwrite it the next time you save a document. You will need to make sure you keep the .pdf file extension when you save it.
 - 3) Open the file in Adobe Acrobat Professional. If you have both Adobe Professional and Adobe Reader on your computer, you may need to click the file with the right mouse button and select "open with Adobe Professional" from the drop down menu.
 - 4) Click "Document" in the toolbar across the top of the Adobe window, and select "Extract pages..." from the menu.
 - 5) You should get a dialog box that says "Extract pages" at the top.
 - a) The first line of the dialog box gives you places to type in the pages you want to extract. For example, with the enrolled version of HR 1, the dialog box opened and said "From: 1 To: 1 of 407" as a default.
 - b) You will want to extract all the pages and save it as a new file, so leave the "1" in the first box and changing the number in the second box to "407" so that the line reads "From: 1 To: 407: of 407." Leave the boxes next to "Delete pages after extracting" and "Extract pages as separate files" unchecked. Then click OK.
 - 6) The top of the Adobe Acrobat Professional window will state "Pages from ____ .pdf." where ____ is the name of the file you saved it as.
 - 7) The GPO authentication logo no longer appears at the top of the first page. Click "File" in the toolbar at the top of the Adobe Acrobat Professional window and select "Save As..." to save the new file. You may want to give it a new file name so that you have the original file you downloaded as a backup.
 - 8) You should be able to highlight and make comments on this new file.
-

Search Options in FDsys

- Basic search
- Field Operators in Basic Search Box
- Advanced search
- Collection search
- Browse search
- Citation search

The Basic Search:

A basic search searches the full text of content and select metadata elements that are most important to each collection

Type a search term or terms into the search box. Hit **Search**. Use the filters on the left side to narrow your search results.

[gross domestic product](#)

Basic Search Options:

*default operator is AND (in GPO Access it was OR)

- Normal Boolean operators: and, or, not (or “-” sign)
- Exact phrase: use “double quotes”
- Find the words: (in any order)
- Proximity: near/#
["Health care" near/10 Obama](#)
- Ordered proximity: before/#
[\(Bill Clinton\) before/50 \(Hillary Clinton\)](#)
- Truncation/wild cards: * to replace multiple characters, ? to replace a single character
 - [congress*](#) for congressional or congressmen
 - [capitali?e](#) for capitalize or capitalize
- “+” sign adds weight to a term:
[autism cancer diabetes](#) vs. [autism +++++cancer diabetes](#)

Using Field Operators in a Basic Search Box:

Note: Field operators vary for each collection. To identify which operators are used for an individual collection, click on [About Government Publications](#) and drill down to “[What’s Available](#).” If you use a field operator, your search results will exclude publications or collections that do not use that field operator.

[lastaction:approved](#) (only searches History of Bills, will not search Public Laws)

[dateheld:2005-05-26](#) (will search hearings, bills, CR, FR, but not all collections)

[zipcodes:20401](#) (only searches the Congressional Directory)

[state:Illinois](#) (only searches Congressional Directory)

[serialnum:109](#) (searches congressional reports and hearings)

Committees

[committee:"Natural Resources" and congress:111 and recovery](#)

Search for a Member

[member:feingold](#)

Limit to Appropriations hearings

[isappropriation:true](#)

[agency:transportation](#)

Limiting by date or congress:

[publishdate:2010](#)

[publishdate:\(2008-10-27\)](#)

[congress:range\(108, 110\)](#) will get you the 108th congress up through the 110th congress.

[congress:range\(,105\)](#) everything up to and including the 105th congress

[congress:range\(105,\)](#) 105th congress and everything after

Collection search examples:

Congressional Record collection

[collection:crec](#)

Congressional Record documents from the 111th Congress that reference H.R. 3200

[collection:crec congress:111 billscitation:"h.r. 3200"](#)

Numbered Congressional Amendments in the Congressional Record

[collection:crec congress:111 amendnumber:145](#)

Search over the Daily Digest Summary and Daily Digest Category, which provide descriptive text about information in the Congressional Record

[ddsummary:\(reid amendment and murray amendment\)](#)

Amendments in the Congressional Record (Look for bill in Associated Legislation navigator)

[crtype:\(samendmentssub or samendtext or hamendments\)](#)

Search types of Congressional Record documents including tributes and memorials

[crtype:\(tributeto or inmemoryof or retirement or recognizing or honoring or passingof\)](#)

Resumption of Congressional Activity in Congressional Record

[crtype:\(ddresumefinal or ddresumeongoing\)](#)

Search Congressional Bills in the 109th Congress, where Mikulski was present and the topic was the Chesapeake Bay watershed restoration, but not anything with forestry.

[collection:bills member:mikulski title:"chesapeake bay" \(watershed near/5 restoration\) and congress:109 not forestry](#)

Search a particular table in the Economic Indicators
collection:econi and tablecategory:(money, credit, and security markets)

Citation search examples:

References to public laws (but not the law itself)
plawcitation:"public law 111-5"

The actual public law
citation:"public law 111-5"

To find references to Public Law mentioned in the text of any collection other than the actual document itself
refcitation:"public law 111-5"

All sections in CFR Title 23 Part 140 from the 2009 edition. Note: Click on more information for the option to download the XML or PDF for the entire part.
collection:cfr cfrtitlenum:23 cfrpartnum:140 publishdate:2009

All documents with the citation 23 CFR 140.902 or that reference 23 CFR 140. Check out the Federal Register documents that reference this CFR part (CFR references in the FR are usually just to the part.)
citation:"23 CFR 140.902" or refcitation:"23 CFR 140"

Search for RIN in FR and CREC
rinnumber:2120-AA64

Search for any document that is or references H.R. 4473.
citation:"H.R. 4473" or refcitation:"H.R. 4473"

Advanced Search:

Limit by Date: fill in only what you want!

For example, MM/DD/YYYY to MM/DD/YYYY or YYYY to YYYY

Limit by collection(s)

Search various fields

You can add more search fields by clicking on [Add More Search Criteria](#) below the search box.

Collection Search:

Identify collection to search from the FDsys homepage. Browse or "drill down" to find what you want.

Browse Search:

- By collection

- By date (only 1 publication per date, if you want the granules or parts of a publication, then you have to do a basic search). You can search by pre-set time frames such as 24 hours, or by specifying a time period.
- By congressional committee

Citation Search:

Use the drop down box to select the collection, and then input your citation information in the fields provided.

Help tools:

- Help (in the top menu bar)
- AskGPO: top of FDsys page
- FDsys tutorials: [For more information on FDsys](#), under **Reference Material**
- Context specific help available in "[More Information About...](#)"
- [About Government Publications](#) > **What's Available?**

Neat tricks & odd stuff:

- Saving your search: bookmark the URL or Add to Favorites
- Misspell a term to show "did you mean?"
- If your user does not know about what a publication is, refer them to the [About Government Publications](#) (link on the left side)