GPO Federal Digital System

Architecture and Design

Deng Wu
Paul Nelson
Johnny Gee

Agenda

- System Architecture
  - Conceptual Model
  - Data Model and OAIS Implementation
  - Application Architecture
  - Application Security
  - FDsys Ingest Process Model
- Data Processing and Search
  - Collections, Packages and Granules
  - Architectural Philosophy
  - Search Features
  - Accessing Documents
- Repository Design
- Q & A

System Architecture
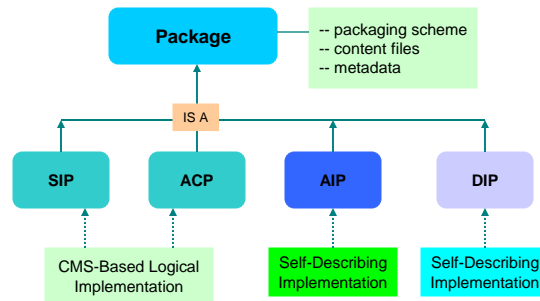
Deng Wu

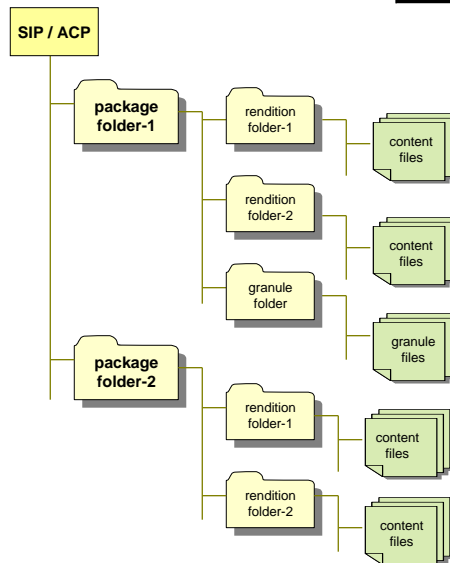## *System Architecture*

- Conceptual Model
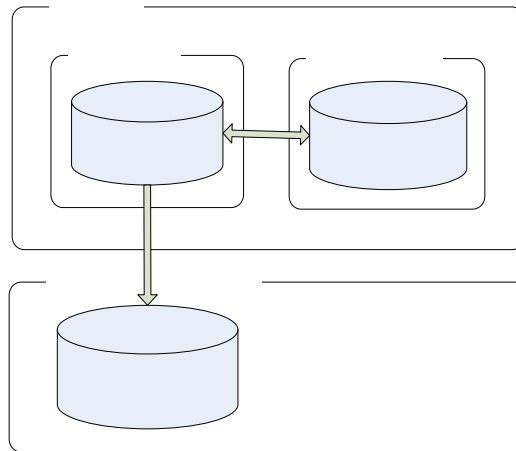
Data Model

- OAIS Implementation
  - SIP, AIP, DIP
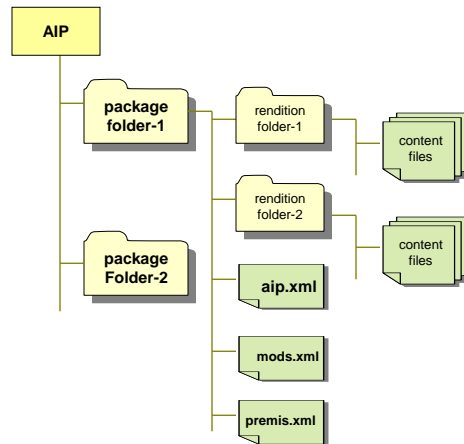 - Metadata Standards: MODS, PREMIS, METS
- FDsys Package Model

Data Model (cont.)

- SIP & ACP

Data Model (cont.)

- AIP

```
AIP
 └── package
     folder-1 ──── rendition
                   folder-1 ──── content
                                 files
                   rendition
                   folder-2 ──── content
                                 files
 └── package
     Folder-2 ──── aip.xml

                   mods.xml

                   premis.xml
```
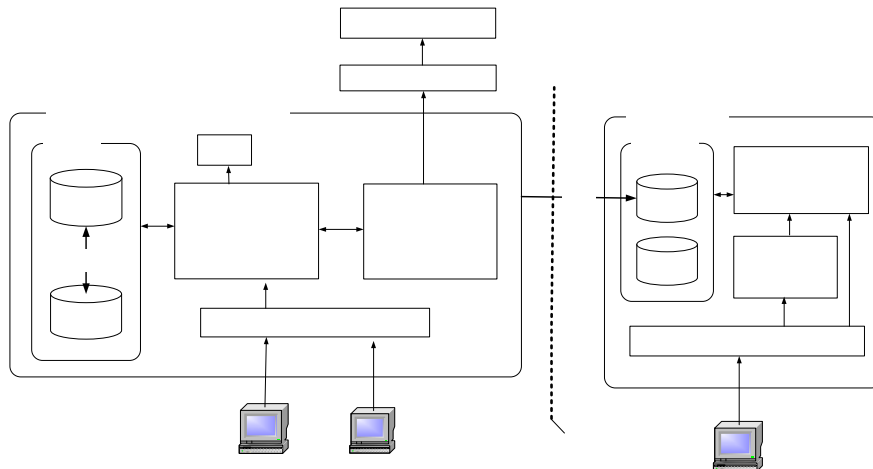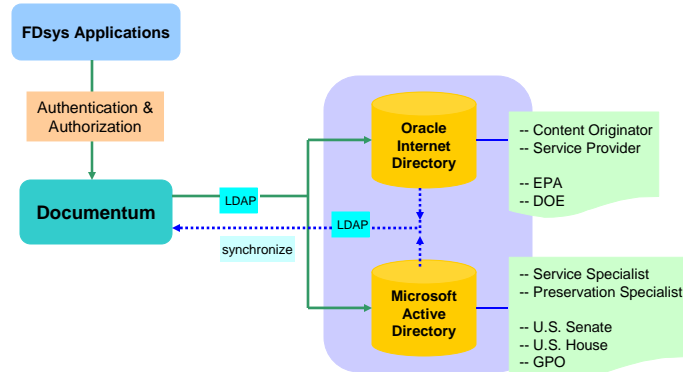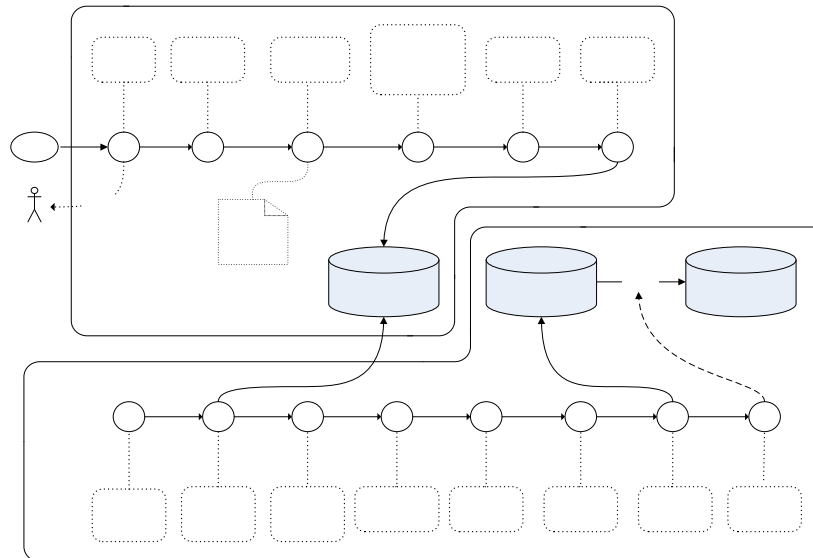
Application Architecture

Application Security

- Roles and groups to enforce application security control
- User access privileges depend on user's roles and groups
  - Content Originator, Service Specialist, Preservation Specialist, etc.
- Users, roles and groups all managed in LDAP

FDsys Applications

Authentication & Authorization

Documentum — LDAP

synchronize — LDAP

Oracle Internet Directory
- -- Content Originator
- -- Service Provider
- -- EPA
- -- DOE

Microsoft Active Directory
- -- Service Specialist
- -- Preservation Specialist
- -- U.S. Senate
- -- U.S. House
- -- GPO

FDsys Ingest Process Model

Data Processing
and
Search

Paul Nelson

Collections, Packages, and Granules

What is a collection?

- Within FDsys:
  - A group of documents which are all *processed* the same
  - Identified with a "Processing Code"
  - Examples: Federal Register, Congressional Bills, Congressional Record, Public and Private Laws

- To the public:
  - A group of documents which logically belong together
  - Identified with a "Collection Code"
  - Examples: Federal Register, Congressional Bills, Congressional Record, Public and Private Laws

---

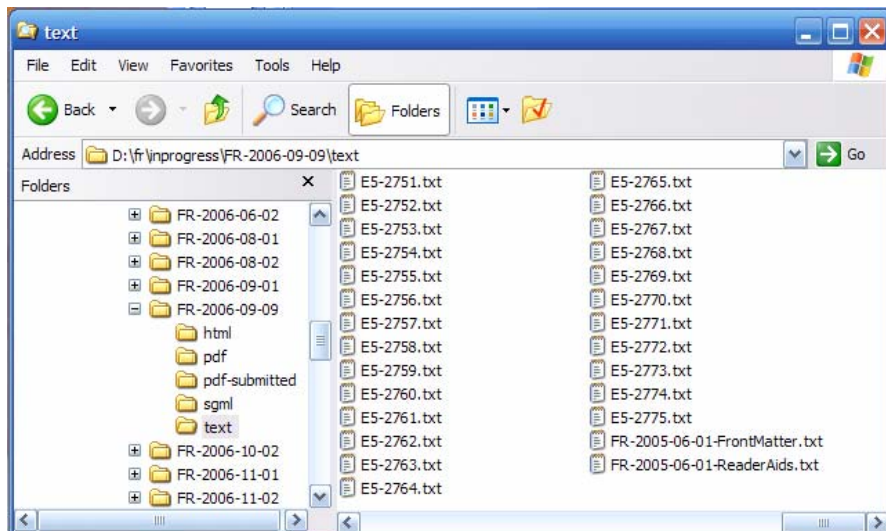## Processing Codes and Collection Codes

- Metadata is used to determine the collection code at index time
- Example Where the Two are not the Same:
  - Presidential Budgets
    - BUDGET2005, BUDGET2006 → BUDGET
    - Processed individually but searched together
  - Congressional Reports printed in the congressional record
    - Will have two collection codes: CR, CRPT
    - Will show up in both collections!
- Possibility for "virtual collections" of the future
  - For example: All documents related to Education

Packages

- Roughly Equivalent to a bound, paper document
- Examples:
  - One issue of the Federal Register
  - One issue of the Congressional Record
  - A single Congressional Bill
  - One issue of the Weekly Compilation of Pres. Docs.
  - One volume of the Code of Federal Regulations (!)
  - One volume of the United States Code
  - The 9-11 Report
  - A single congressional committee report

Packages Example

Granules

- *"The most usefully searchable unit"*
- *Examples:*
  - *A single FR article*
  - *A single CR unit of business (separated by ◆——◆)*
  - *A CFR section (e.g. §57.402 in Title 40)*
  - *An entire bill*
  - *An entire report*
  - *A Pres. speech*

Implications of Granules

- Search over granules
  - Search results are *much more accurate*
- Retrieve individual granules
  - Can view a single granule rather than sifting through an entire issue or volume
- Granules can be digitally signed
  - To ensure authenticity of the data
- No single page retrieval
  - Instead, you get the best granule
  - PDF viewer automatically jumps to the page you asked for

Architectural Philosophy

FDsys is a Data Driven Architecture
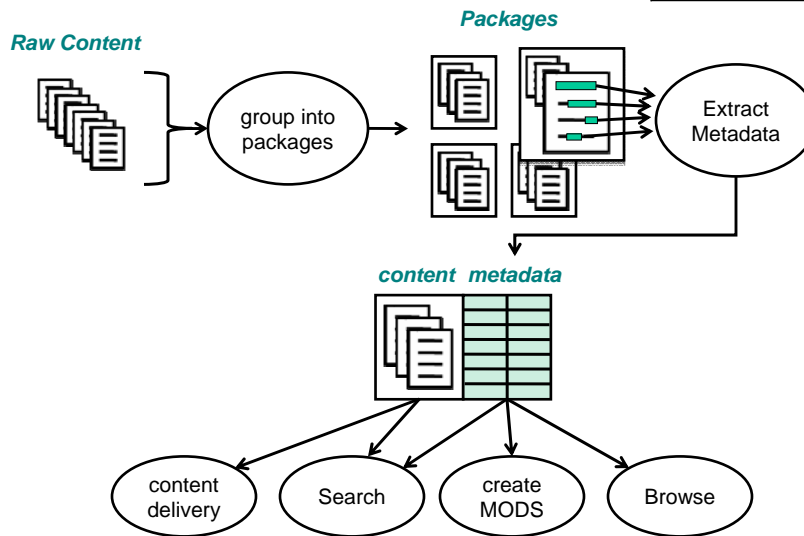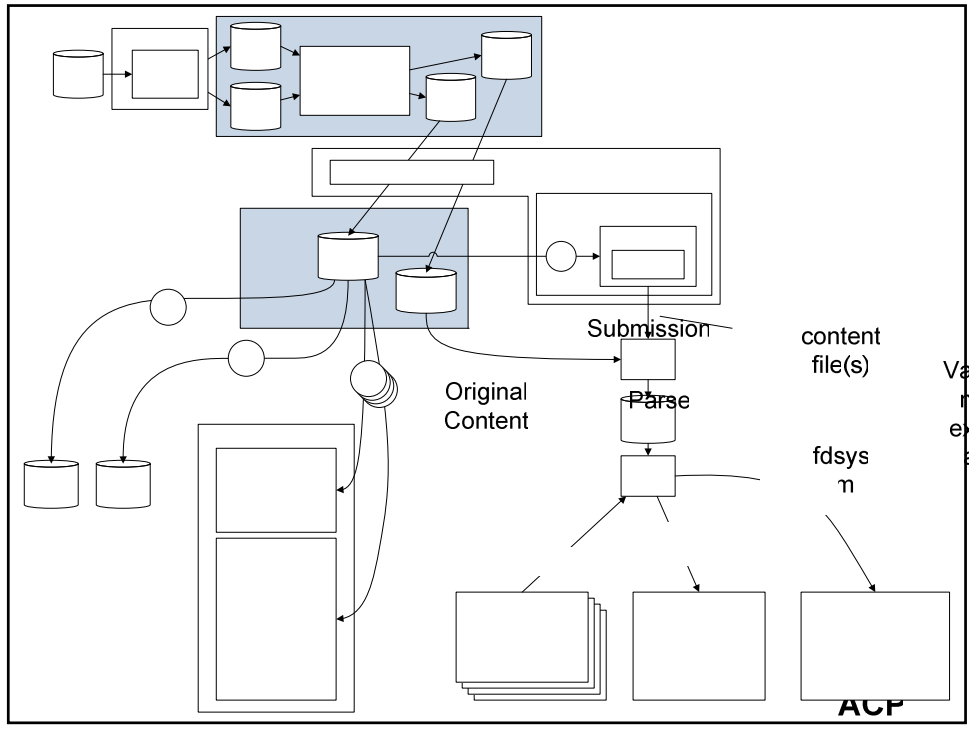
Submission

content
file(s)

Original
Content

Parse

fdsys
m

fdsys
xml

publish

**ACP
Cache**

fdsys
xml

content
file(s)

xslt

Parsing Introduction

xslt
xslt
xslt

- Runs regular expressions to extract metadata

  Regular Expression:

  `(Public Law|Pub. L.|PL|P. L.) (1[0-9][0-9])-([0-9]+)`

  Example: `Pub. L. 109-130`

  Produces: `<publicLaw congressNum="109">130</publicLaw>`

- Produces an instance of **fdsys.xml**

  – Parsed metadata is also available in the MODS

Searc

field1
field2
field3
field4

[*]

Federal Register Example Metadata

GPO FDsys

agencies → DEPARTMENT OF AGRICULTURE / Forest Service

title → Notice of Southwest Idaho Resource Advisory Committee Meeting

AGENCY: Forest Service, USDA.

action → ACTION: Notice of meeting.

summary → SUMMARY: Pursuant to the authorities in the Federal Advisory Committee Act (Pub. L. 92–463) and under the Secure Rural Schools and Community Self-Determination Act of 2000 (Pub. L. 106–393), the Boise and Payette National Forests' Southwest Idaho Resource Advisory Committee will conduct a business meeting, which is open to the public.

dates → DATES: Wednesday, August 16, 2006, beginning at 10:30 a.m.

ADDRESSES: Valley County Courthouse, 219 North Main Street, Cascade, Idaho.

SUPPLEMENTARY INFORMATION: Agenda will include review and approval of project proposals, and is an open public forum.

contact → FOR FURTHER INFORMATION CONTACT: Doug Gochnour, Designated Federal Officer, at 208–392–6681 or e-mail dgochnour@fs.fed.us.

Dated: July 26, 2006.
Lana S. Thurston,
Administrative Officer, Boise National Forest

FR Doc Number → [FR Doc. 06–6601 Filed 7–31–06; 8:45 am]

Billing Code → BILLING CODE 3410–11–M

GPO Federal Digital System – Architecture and Design   October 21, 2008

Extracted Metadata Example

GPO FDsys

```
<descMdGroup id="id-05-10658">
  <title>Editorial Modifications of the Commission's Rules</title>
  <printPageRange first="31372" last="31374"/>
  <migratedDocID>fr01jn05-12</migratedDocID>
  <collectionSpecific>
    <accessId>05-10658</accessId>
    <granuleClass>RULE</granuleClass>
    <agency order="1">FEDERAL COMMUNICATIONS COMMISSION</agency>
    <effectiveDate>2005-04-14</effectiveDate>
    <billingCode>6712-01-P</billingCode>
    <frDocNumber>05-10658</frDocNumber>
    <action>Final rule.</action>
    <summary>This document amends twelve sections ... updating the postal
address of the Arecibo Radio Astronomy Observatory.</summary>
    <dates>Effective April 14, 2005.</dates>
    <contact>Rodney Small, Office of Engineering and Technology, (202) 418-
2452.</contact>
    <cfr title="47">
      <part number="25"/>
      <part number="73"/>
      <part number="74"/>
    </cfr>
    <tocSubject1>Practice and procedure:</tocSubject1>
    <tocDoc>Commission's rules; editorial modifications, </tocDoc>
  </collectionSpecific>
</descMdGroup>
```

Using the Table of Contents

- Subject oriented table of contents
  - Federal register:
    - Subject headings
  - Weekly Compilation of Presidential Documents:
    - Category headings
  - Congressional Record:
    - Daily digest

subject headings used to annotate granules
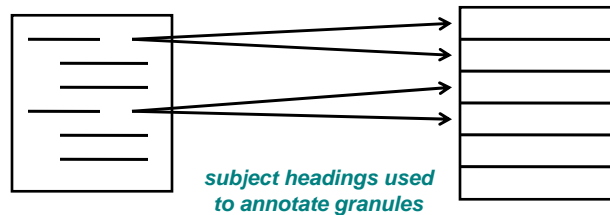
TOC Metadata Example

```
<descMdGroup id="id-05-10658">
   <title>Editorial Modifications of the Commission's Rules</title>
   <printPageRange first="31372" last="31374"/>
   <migratedDocID>fr01jn05-12</migratedDocID>
   <collectionSpecific>
     <accessId>05-10658</accessId>
     <granuleClass>RULE</granuleClass>
     <agency order="1">FEDERAL COMMUNICATIONS COMMISSION</agency>
     <effectiveDate>2005-04-14</effectiveDate>
     <billingCode>6712-01-P</billingCode>
     <frDocNumber>05-10658</frDocNumber>
     <action>Final rule.</action>
     <summary>This document amends twelve sections ... updating the postal
address of the Arecibo Radio Astronomy Observatory.</summary>
     <dates>Effective April 14, 2005.</dates>
     <contact>Rodney Small, Office of Engineering and Technology, (202) 418-
2452.</contact>
     <cfr title="47">
       <part number="25"/>
       <part number="73"/>
       <part number="74"/>
     </cfr>
     <tocSubject1>Practice and procedure:</tocSubject1>
     <tocDoc>Commission's rules; editorial modifications, </tocDoc>
   </collectionSpecific>
</descMdGroup>
```

TOC Searches

- Find all articles in the Federal Register which are listed as "meetings" in the Contents

- Find all entries from the body of the Congressional Record listed under "Measures Passed" in the Daily Digest

- Find all entries from the body of the the Congressional Record identified as "The Patriot Act" in the Daily Digest

- Find all presidential documents identified as "Communications to Congress" in the contents

Table of Contents Example

Table of Contents

Federal Register Vol. 69, Issue 66, Tuesday, April 6, 2004

Administration for Children and Families (within the Health and Human Services Department)

NOTICES
Grants and cooperative agreements; availability, etc.:
American Indian-Alaska Native Head Start-University Partnerships Program,
Pages 18085-18094 [FR DOC# 04-7260]   PDF | Text | More

Agriculture Department

See: Animal and Plant Health Inspection Service
See: Forest Service
See: Natural Resources Conservation Service
See: Rural Housing Service

Air Force Department (within the Defense Department)

NOTICES
Meetings:
HQ USAF Scientific Advisory Board,
Page 18061 [FR DOC# 04-7673]   PDF | Text | More

Congress Member and Committee Normalization

- Congress members and committees are normalized
- Congress Member Codes

```
<congMember authorityId="308" chamber="S" congress="109" party="D"
                                    role="SPONSOR" state="ND">
  <name type="parsed">Mr. Dorgan </name>
  <name type="authority-fnf">Byron L. Dorgan</name>
  <name type="authority-lnf">Dorgan, Byron L.</name>
  <name type="authority-other">Byron Leslie Dorgan</name>
</congMember>
```

- Congressional Committee Codes

```
<congCommittee authorityId="ssga00" chamber="S"
                              congress="109" type="S">
  <name type="authority-standard">
     Committee on Homeland Security and Governmental Affairs</name>
  <name type="authority-short">
     Homeland and Governmental Affairs</name>
</congCommittee>
```

Uses of Normalized Names

- Display the official names of congress members, presidents, and committees

- Search on a committee even if the name changes over time

- Search by congress member state codes or party affiliations

  - "Final all documents sponsored by a senator from Maryland"

Search Features

Search Results

1. 69 FR 18059 - Petition Requesting Mandatory Fire Safety Standards for Candles and Candle Access
   Federal Register. Notices.Notice. Tuesday, April 6, 2004.
   ...COMMISSION Petition Requesting Mandatory **Fire Safety** Standards for Candles and Candle...the Commission is
   candle...1/HP 04-1, Petition for **Fire** More Information

Navigators

Section
Notices (519)
Rules and Regulations (423)
Proposed Rules (335)
See More...

Agency
Transportation Department (375)
Federal Aviation Administration (320)
Homeland Security Department (93)
See More...

Cfr Citation
⊟ 14 CFR (334)
    Part 39 (276)
    Part 25 (31)
    Part 23 (10)
    Part 121 (8)
⊞ 40 CFR (79)
⊞ 33 CFR (65)
See More...

Collection Browsing

PUBLIC AND PRIVATE LAWS

About the Public and Private Laws.

⊟ 105th Congress (1997 - 1998)   ← *congress*
  ⊟ Private Law (Pvt. L.)   ← *law type*
    ⊞ Less than 2
    ⊞ From 3 to 4   ← *number range*
    ⊟ From 5 to 7

Private Law (Pvt. L.)	5	For the relief of Heraclio Tolley.	PDF| Text| More
Private Law (Pvt. L.)	6	For the relief of Larry Errol Pieterse.	PDF| Text| More
Private Law (Pvt. L.)	7	For the relief of Mai Hoa ``Jasmin" Salehi.	PDF| Text| More

List of Laws →

    ⊞ 8 and above
  ⊞ Public Law (Pub. L.)
⊞ 107th Congress (2001 - 2002)

Advanced Search Form

ADVANCED SEARCH

Publication Date: If publication date is not available, search is performed on date of submission into FDsys.

| Date is ▼ | January ▼ | 1 ▼ | 2006 ▼ |

Collections:

Available Collections

All Collections
Public and Private Laws
Compilation of Presidential Documents
Federal Register

Add >>
<< Remove

Selected Collections

Congressional Bill

Search in:

| Sponsors and Co-Sponsors ▼ | for | Cardin |
| Bill Version ▼ | for | Considered and Passed House ▼ |

Add more search criteria (5 max)

Search   Clear   Search Tips | Retrieve by Citation

Data Mapping Examples

Internal Data Storage	Web Application
• 110	• 110th Congress (2007-2008)
• V	• Part V
• FR	• Federal Register
• RULE	• Rules and Regulations
• congnum	• Congress Number
• 2006-02-01	• February 1st, 2006
• [2006-02-01;]	• After February 1st, 2006
• accode=PPL congnum=110 billnumber=1234 publishdate=2008-01-01	• Public and Private Laws. 110th Congress. H.R. 1234. January 1, 2008.

69 FR 18059 PETITION REQUESTING MANDATORY FIRE SAFETY STANDARDS FOR CANDLES AND CANDLE ACCESSORIES (PETITION NO. CP 04-1/HP 04-1)
Back to search results

Download Files

Formats:	Text (NaN KB)
Descriptive Metadata:	MODS	PREMIS
All Format & Metadata Files:	ZIP file

Metadata

Publication Title:	Federal Register Volume 69, Issue 66 (Tuesday, April 6, 2004)
Category:	Regulatory Information
Collection:	Federal Register
SuDoc Class Number:	AE 2.7:
	GS 4.107:
	AE 2.106:
Publisher:	Office of the Federal Register, National Archives and Records Administration
Section:	Notices
Action:	Notice.
Dates:	The Office of the Secretary must receive comments on the petition by June 7, 2004.
Contact:	Rockelle Hammond, Office of the Secretary, Consumer Product Safety Commission, Washington, DC 20207; telephone (301) 504-6833, e-mail rhammond@cpsc.gov.
Summary:	The United States Consumer Product Safety Commission (Commission or CPSC) has received a petition (CP 04-1/HP 04-1) requesting that the Commission issue mandatory fire safety standards for candles and candle accessories. The Commission solicits written comments concerning the petition.
Agency Name:	CONSUMER PRODUCT SAFETY COMMISSION
Page Number Range:	18059 - 18059
Federal Register Citation:	69 FR 18059

- Simple queries: Barbara Mikulski and the Bay
  – NOTE: Defaults to AND
- Exact Phases: "Department of Defense"
- Boolean operators:

  (Mikulski or Cardin) and bay not seasoning
  – or, and, ( ), not
- Proximity Operators: house near/10 action
  – near/#, before/#, adj
- Date and numeric ranges
  – range(110,130)
  – range(2008-01-01, 2009-01-01)

Query Language: Examples

GPO FDsys

- hearing
- "congressional hearing"
- congressional adj hearing
- congressional hearing
- ways and means
- "ways and means"
- ways "and" means
- congressional or congress
- congress and (report or meeting or notice)
- congnum:range(103,110)

- congressional not report
- congressional –report
- +cardin congressional committee
- congresional not (committee report)
- congressional not (committee or meeting)
- representative near/10 cardin
- representative before/10 cardin

The Field Operator

GPO FDsys

- Fields can be targeted with the field: operator
- For R1C2: A selected list of fields are available
- Each collection has its own list of searchable fields

Note: More fields available with field: than are available on the advanced search form

- Examples:
  - title:(environmental controls)
  - congressnum:110
  - title:"executive communications"
  - congressnum:110 and billnumber:range(1000,1500)

Search over MODS

- Expected for a near-term future release
- Can perform complex queries directly over the MODS
- Since the MODS contains all extracted metadata, this means just about any search is possible
- Examples:
  - mods:identifier:(@type:citation and "70 FR 14232")
  - collection:BILLS and mods:congMember:(@congress:house and @state:MD and @role:SPONSOR) and mods:USCode:(@title:14 and section:@number:673)

Relevancy Ranking

- Priority 1: Citations
  - Public Law 108-5
  - 70 FR 31389
  - H.R. 1413
- Priority 2: Words in the title
- Priority 3: Names (agencies, organizations, congress members, committees)
- Priority 4: Other summary metadata (abstracts)
- Priority 5: Referenced documents (i.e. U.S.C. references, Stat. L. references)
- Priority 6: Everything else

Citations are Normalized and Expanded for Search

GPO FDsys

- All citation methods are normalized based on metadata
- Citations are then expanded to include all common variations
  - HR 1042
  - H.R. 1043
  - House Bill 1043
- All variations are indexed, a search on any variation will retrieve the desired record

Other search features

GPO FDsys

- Spell Check
  - Basic English plus common proper names
  - To be expanded in the next release
- Synonyms
  - None in this release
  - Capability installed and available for next release
- Wildcards
  - Standard with software
    - legis*
    - presid*al
    - theat??

GPO FDsys

Accessing Documents

Access IDs

GPO FDsys

- Federal Register   *Example:*
  - Package: Issue Date   FR-2008-01-03
  - Granule: FR Document Number   06-1325
- Congressional Bills
  - Package: Congress number, bill type, bill number, bill version
    code   BILLS-109hr1323rfs
- Weekly Compilation of Presidential Documents
  - Package: Issue Date   WCPD-2008-01-03
  - Granule: Issue Date + page number + count
                                                     WCPD-2008-01-03-Pg423-2
- Public and Private Laws
  - Package: congress, type, number   PLAW-108publ5

Things you can download

- Granules
  - The PDF of the entire granule (not just page at a time)
  - An HTML rendition of the granule
  - Granule MODS
- Packages
  - PDF files of the entire package, the front matter, the end matter, and any nested granule
  - PDF files of any nested granule
  - ZIP file of all publicly available package contents
  - Package level MODS
  - Package PREMIS data
- Possibly other renditions (e.g. XML)

Package-Level URLs

- Package Content Detail
  - http://www.gpo.gov/fdsys/pkg/FR-2006-01-01/content-detail.html
- Package Metadata Standards
  - http://www.gpo.gov/fdsys/pkg/FR-2006-01-01/mods.xml
  - http://www.gpo.gov/fdsys/pkg/FR-2006-01-01/premis.xml
- Package Table of Contents
  - http://www.gpo.gov/fdsys/pkg/FR-2006-01-01/toc.html
- Today's Table of Contents
  - http://www.gpo.gov/fdsys/html/FR/todays_toc.html

Granule-Level URLs

- HTML and PDF Files
  - http://www.gpo.gov/fdsys/pkg/FR-2006-01-01/html/E6-1423.html
  - http://www.gpo.gov/fdsys/pkg/FR-2006-01-01/pdf/E6-1423.pdf
- Granule Content Detail
  - http://www.gpo.gov/fdsys/granule/FR-2006-01-01/ E6-1423/content-detail.html
- Granule Metadata Standards
  - http://www.gpo.gov/fdsys/granule/FR-2006-01-01/ E6-1423/mods.xml

---

MODS Example

```
− <titleInfo>
    − <title>
        To amend titles IV, XVIII, and XIX of the Social Security Act to improve the provision of care under the programs under such titles, a
      </title>
      <title>Health and Welfare Relief Act of 2005</title>
      <title>S. 2164 (IS)</title>
  </titleInfo>
− <originInfo>
      <publisher>U.S. Government Printing Office</publisher>
      <dateIssued encoding="w3cdtf">2005-12-21</dateIssued>
      <issuance>monographic</issuance>
  </originInfo>
− <physicalDescription>
      <note type="source content type">deposited</note>
      <digitalOrigin>born digital</digitalOrigin>
  </physicalDescription>
  <classification authority="sudocs">Y 1.6:</classification>
  <classification authority="sudocs">Y 1.4/1:</classification>
  <identifier type="uri">http://www.gpo.gov/fdsys/pkg/BILLS-109s2164is</identifier>
  <identifier type="stock number">021-610-00252-9</identifier>
  <identifier type="Parent ILS system id">000501532</identifier>
  <identifier type="Child ILS system id">000325576</identifier>
  <identifier type="Parent ILS title">Congressional bills</identifier>
  <identifier type="Child ILS title">Senate bills</identifier>
  <identifier type="former package identifier">f:s2164is.txt</identifier>
− <location>
    − <url displayLabel="Content Detail" access="object in context">
        http://www.gpo.gov/fdsys/pkg/BILLS-109s2164is/content-detail.html
      </url>
```

Repository Design

Johnny Gee

Repository Key Requirements

- **Basic Requirements**
  - Submission
  - Ingestion
  - Access Processing
  - Metadata Management
  - Preservation
  - Publishing
- **Integration Interfaces**
  - Content Parser
  - Virus Check
  - Adobe LiveCycle
  - AIP Reconstruct Utility
  - ILS/ESB
  - Publishing to ACP Cache

- **Use Cases:**
  - Login
  - Upload
  - Submit Package
  - Upload ACP Rendition
  - Update Content
  - Update Metadata
  - Delete ACP Package/Rendition
  - Delete AIP Package
  - Validate Package
  - Search
  - Browse
  - Reporting
  - Etc.

Enterprise Content Management Platform

Create

Authoring Applications

Electronic Forms

HTML/XML

Rich Media Sources

Manage

Workflow

Search and Index

Renditions and Transformations

Lifecycles

Native support for XML

Security and Audit trails

EMC Documentum ECM Platform

Deliver / Publish

Email

Print

Wireless

Integration with Enterprise Applications

Website/Portals
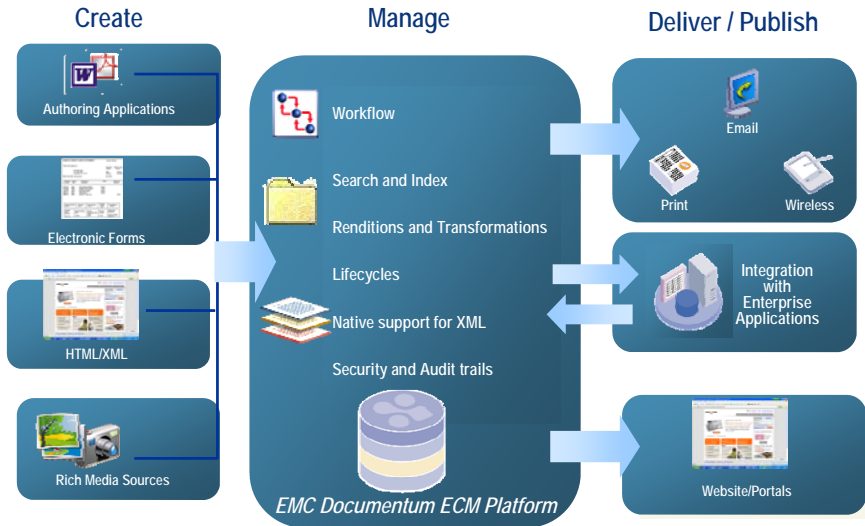
FDsys Repository Package

- Work In Progress (SIP)
  - Federal Register
    - FR-2008-09-20
  - Congressional Bills
    - BILLS-111hr1234enr

- Submitted (ACP)
  - Federal Register
    - FR-2008-06-01
      - PDF
        - 164.pdf
      - TXT
        - 164.txt
      - HTML
        - 164.htm
        - images
          - image1.gif
      - Granule
        - 164a.txt
        - 164b.txt
        - 164c.txt
        - 164d.txt
    - FR-2008-06-02
  - Congressional Bills
    - BILLS-110hr3690enr

- Archived (AIP)
  - Federal Register
    - FR-2008-06-01
    - FR-2008-06-02
  - Congressional Bills
    - BILLS-110hr3690enr

Roles in FDsys

GPO FDsys

- Roles

Role Name	Description
ext_search_and_access	External Search and Access
submission_search_and_access	Submission Search and Access
standard_search_and_access	Standard Search and Access
content_submission	Content Submission
content_mgmt	Content Management
reporting_manifests	Manifest Reporting
reporting_metrics	Metrics Reporting
workflow_admin	Workflow Administrator

## FDsys Security

GPO FDsys

- Used on Work In Progress (SIP) and Submitted (ACP) cabinets. Document level security will be handled by alias sets and template ACLs.

	fdsys_sip_ cabinet_acl	fdsys_acp_ cabinet_acl
dm_owner	Delete+LOP	Delete+LOP
dm_world	None	None
sip_delete	Read	None
acp_read	None	Read
acp_delete	None	Read
aip_write	None	Read
fdsys_admins	Delete+LOP	Delete+LOP

- Used only in the Archived (AIP) cabinet and the contained folders and archive documents. A server method will perform deletions in order to support group approval for package deletions.

	fdsys_aip_ cabinet_acl	fdsys_aip_ folder_acl	fdsys_aip_ archive_doc_acl
dm_owner	None	None	None
dm_world	None	None	None
sip_delete	None	None	None
acp_read	None	None	None
acp_delete	None	None	None
aip_write	Read	Write	Write
fdsys_admins	Write+LP	Write+LP	Write+LP

Repository Process Overview
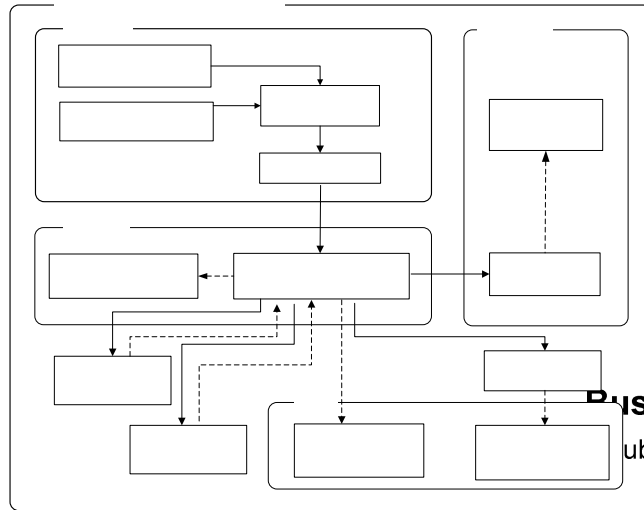
Business Process Overview

ubmission

Interactive Submission

Hot Folder Submission

Upload
Process

Submission W
& Rules F

Submission to FDsys

GPO **FDsys**

- FDsys will support uploading content files into FDsys via two methods:
  – Upload Package (interactive)
  – Hot Folder Job (folder-based)
- Hot Folder Job will perform the upload automatically, instead of manually by a user.
- Once uploaded, a user can select multiple packages to submit at one time. The submit action will launch the Ingestion Workflow.

Processing W
& Rules F

Access

Auth
Acce
(

e LiveCycle
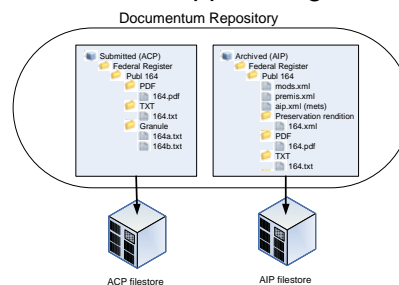tegration

Ingestion Workflow

- Create and Validate AIP Package
  - Generate METS, MODS, and PREMIS xml files
- Create CryptoTimeStamp (Fixity)
  - Generate two separate message digests using unique FDsys ID and ingestion date-time
- Create Preservation Rendition (XML)
  - Generate xml rendition from pdf content files using Adobe LiveCycle
- Invoke Access Processing Workflow

ACP and AIP Packages

- ACP packages are stored in ACP filestore
- AIP packages are stored in AIP filestore, which is physically different than the ACP filestore
- ACP is designed to support publishing to ACP Cache and daily updates to packages
- AIP is designed to be self-contained, self-describing, and independent of DCTM to support long-term preservation

Documentum Repository

Submitted (ACP)
Federal Register
Publ 164
PDF
164.pdf
TXT
164.txt
Granule
164a.txt
164b.txt

Archived (AIP)
Federal Register
Publ 164
mods.xml
premis.xml
aip.xml (mets)
Preservation rendition
164.xml
PDF
164.pdf
TXT
164.txt

ACP filestore

AIP filestore

METS, MODS, & PREMIS

- METS, MODS, & PREMIS are generated and updated when AIP package is created, and when any metadata and/or content files are updated
- MODS file will be versioned when there is a change in the metadata
  - This will facilitate tracking of metadata updates
- PREMIS file will have a accumulative provenance information.
- METS file will have complete information about metadata and content files for a package, independent of DCTM
- AIP Reconstruct Utility can be used to reassemble AIP package from AIP filestore, independent and outside of Documentum
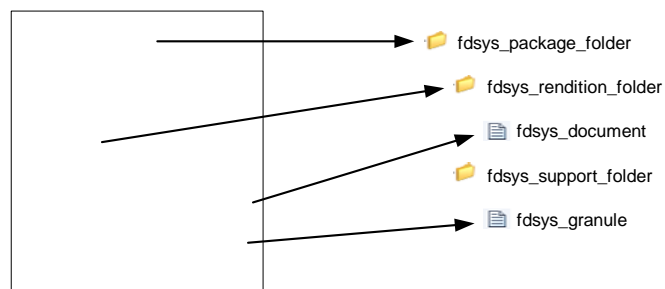
Access Processing Workflow

- Call Appropriate Parser
  - Load parsed data into Documentum for ease of metadata updating
  - Parser determines quality of metadata based on heuristic rules
- Validate Quality of Data
  - If quality errors exist or metadata is missing, notify metadata correction group to review and edit metadata
  - Users can search for package with low quality metadata and update metadata values appropriately
- Create PDF granules
  - Generate pdf granules from submitted PDF files using Adobe LiveCycle

Access Processing Workflow (cont'd)
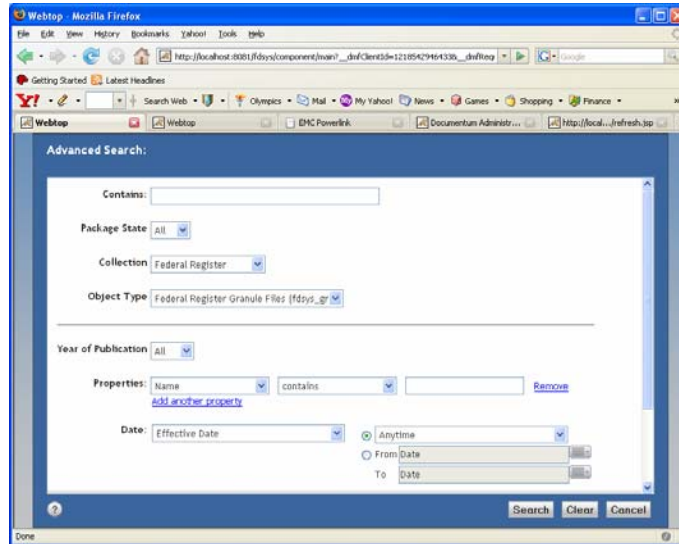
GPO FDsys

- Create HTML rendition
  - Extract images from PDF files using Adobe LiveCycle
  - Perform OCR on images to extract indexable text
  - Include text as part of ALT tag for the image in HTML
  - Convert URLs and E-mail addresses to clickable links
- Digitally Sign PDF (if required)
  - Apply GPO Digital Signature and Certificate
- Invoke ILS Workflow
- Publish Package for Access

FDsys.xml Metadata Mapping to CMS

GPO FDsys

fdsys_package_folder

fdsys_rendition_folder

fdsys_document

fdsys_support_folder

fdsys_granule

Metadata Management
for Authorized Users

For submission_search _and_access role

Metadata Management
for Authorized Users (cont'd)

For standard_search _and_access role

Advanced Search

Architecture & Design

- Q & A