New GovInfo Content and Enhancements

October 18, 2022

Speakers

- Jessica Tieman, Digital Preservation Librarian, GPO

- Amanda Dunn, Program Planner, GPO

Newly added content in FY22:

- 1,498 Congressional Hearings from the 79th-97th Congress
- 85 Committee Prints from the 84th-96th Congress
- 20 GPO Publications including Annual Reports, Semiannual Reports to Congress, GPO Management Responses to IG Reports, GPO Transmittal Letters for OIG Semiannual Reports to Congress, Annual Budget Justifications, and LSCM Year in Review
- 84 Publications from the Railroad Retirement Board
- 200 Publications from the Census Bureau
- Various Publications from the Defense Department and Federal Aviation Commission

Newly added content in FY22:

- 7 New US Courts Added
- 4,550 Web-harvested Monographs
- USLM XML for the Constitution, Jefferson's Manual, and the Rules of the House of Representatives, 117th Congress
- 4,373 publications within the Serial Set collection featuring:
  - House & Senate Reports
  - House & Senate Documents
  - House Executive Documents
  - House & Senate Miscellaneous Documents
  - Senate Journals and more across the 15th-82nd Congress

Newly added Partnership Content in FY22:

- Boise State University has contributed 16 publications from the Interior Department, Agriculture Department, Pacific Northwest River Basins Commission, and other Government authors on topics such as water pollution, water and soil conservation, environmental disaster effects, fish and wildlife rehabilitation, and much more relating to local environmental assessments of the Snake River and Idaho regional area

Content Coming Soon:

- Congressional Directory in PDF
  - 216 volumes from the 41st Congress to the 104th Congress.
- Depository Shipping Lists in PDF
  - 10,407 Federal Depository Library Program daily shipping lists from January 1980 through December 1989 made available individually
- Digitized Statutes at Large in PDF and USLM XML
  - 110 volumes beginning in 1789 through 1950 and conversion to USLM XML formats for all digitized volumes and volumes from 1951-2002 already within GovInfo
- Thousands more Web-harvested Monographs!

Trustworthy Digital Repository Certification

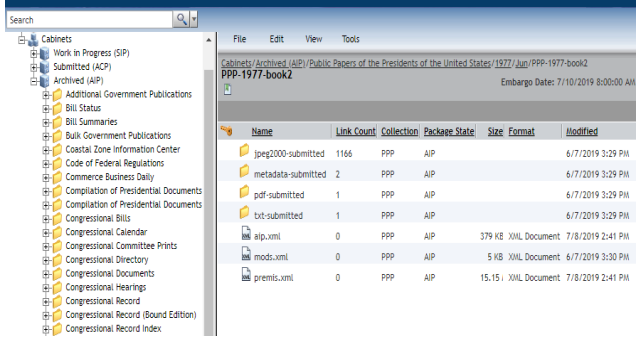
- GovInfo remains the only Digital Repository in the world to maintain ISO 16363 certification.
- In Summer 2022, GPO made the decision to apply for CoreTrustSeal certification. GPO intends to maintain certification through both CoreTrustSeal and ISO 16363.
- A dual-model of certification allows GPO to remain certified at all times, in the unforeseen event that availability of ISO 16363 certifying bodies are not available. Involvement in CTS also provides GPO a means to stay engaged in a community of over 110 repositories around the world, including 10 federally operated data repositories in the US.

GovInfo Enhancements

GovInfo Components

GovInfo System Primary Components

Content Management System



manages digital content to ensure authenticity and integrity

Preservation Repository


```
<?xml version="1.0" encoding="UTF-8"?>
- <mets:mets OBJID="V0b002ee1a153fd32" xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:mets="http://www.loc.gov/METS/">
  - <mets:metsHdr LASTMODDATE="2019-06-07T15:30:57-04:00" CREATEDATE="2019-06-
    07T15:17:49-04:00">
      - <mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
          <mets:name>United States Government Publishing Office</mets:name>
        </mets:agent>
        <mets:altRecordID TYPE="ACP">P0b002ee1a14ca7f1</mets:altRecordID>
    </mets:metsHdr>
        <!--MODS object-->
  - <mets:dmdSec ID="DMD_OTHER">
        <mets:mdRef ID="D09002ee1a154169d"
            xlink:href="file:/u01/app/emc/data/storage_aip3/00002ee1/80/37/2b/d4.xml"
            xmlns:xlink="http://www.w3.org/1999/xlink" LOCTYPE="URL" MIMETYPE="text/xml"
            MDTYPE="MODS"/>
    </mets:dmdSec>
        <!--PREMIS OBJECT -->
  - <mets:amdSec ID="AMD_OTHER">
      - <mets:techMD ID="D09002ee1a1541927-TEC">
```


follows preservation standards enabling access to digital content

Public Website and API

combines extensive metadata, robust search, and modern design

FY22 Development Highlights

- December 2021 –
  - VeraPDF software to identify PDFs with embedded files
  - House and Senate Congressional Hearing "Event IDs" added to MODS metadata
  - Access to related Committee Prints from Bills details pages and API service
  - Related Documents API service for Federal Register and CFR
  - Serial Set enhancements from feedback
  - Search engine upgrade and internal content management system (CMS) upgrade
- March 2022 –
  - Improved access to individual Senate amendments in the Congressional Record
  - Related Documents for Committee Prints from Public and Private Laws
  - Citation Search enhancements to display multiple matching results
  - API parameters for Serial Set to filter by volume or individual documents
  - Numerous enhancements to the Miscellaneous Government Publications collection

FY22 Development Highlights (continued)

- June 2022 –
  - Development for Digitized Congressional Committee Prints
  - Link Service for individual Senate amendments in the Congressional Record
  - Support for addenda to Congressional Hearings
  - API functionality to return more than 10,000 results over multiple pages
  - API related endpoints for Public and Private Laws collection
  - Major upgrade to public web content management system (CMS)

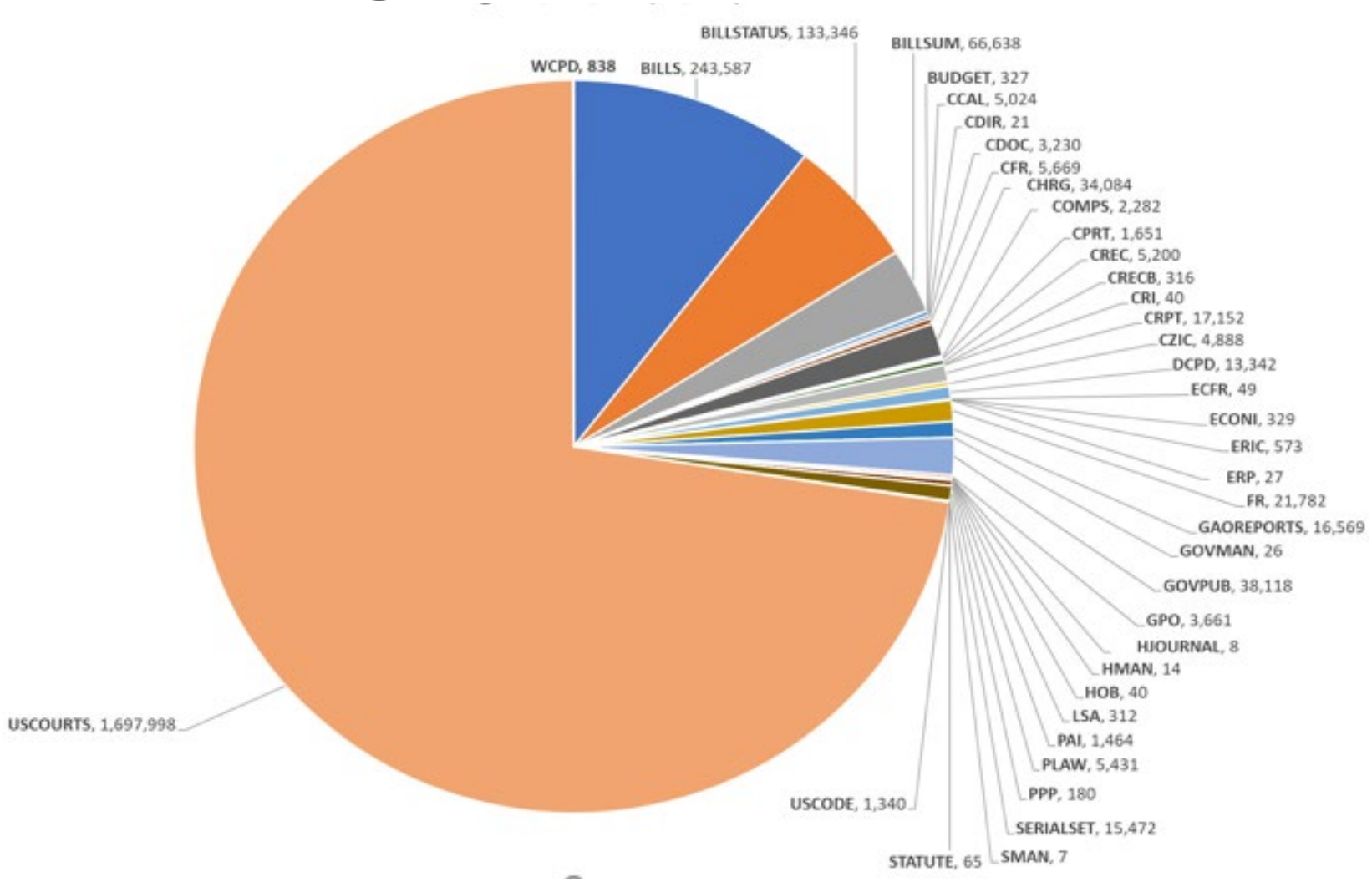
- September 2022 –
  - Access to Related Congressional Record documents from Bills details and API
  - Related Documents for Public Laws from Congressional Record details and API
  - Access to individual Senate introductory statements in the Congressional Record and API
  - API update to provide up to 1,000 results per page
  - Development for processing of and access to cataloged born-digital content
  - Tools to OCR, optimize, and detect corrupt PDF files in initial collections
  - Collection-specific citations for Chicago Style in the Bibliographic Citation Generator

GovInfo Metrics

- 10.6M PDF files and 24.2M image files in the GovInfo Preservation Repository
- PDF, Text, XML, and JP2 are the predominate file formats
- 152K packages added to GovInfo in FY22
- 5 minute average processing time from ingest to availability on the public website and API
- 72M average monthly file retrievals from GovInfo in FY22 compared to 70M in FY21
- 869M file retrievals from GovInfo in FY22, an increase of 32M compared to FY21
- 5.7B public file retrievals since January 2009 and 9.7B public file retrievals since 1994

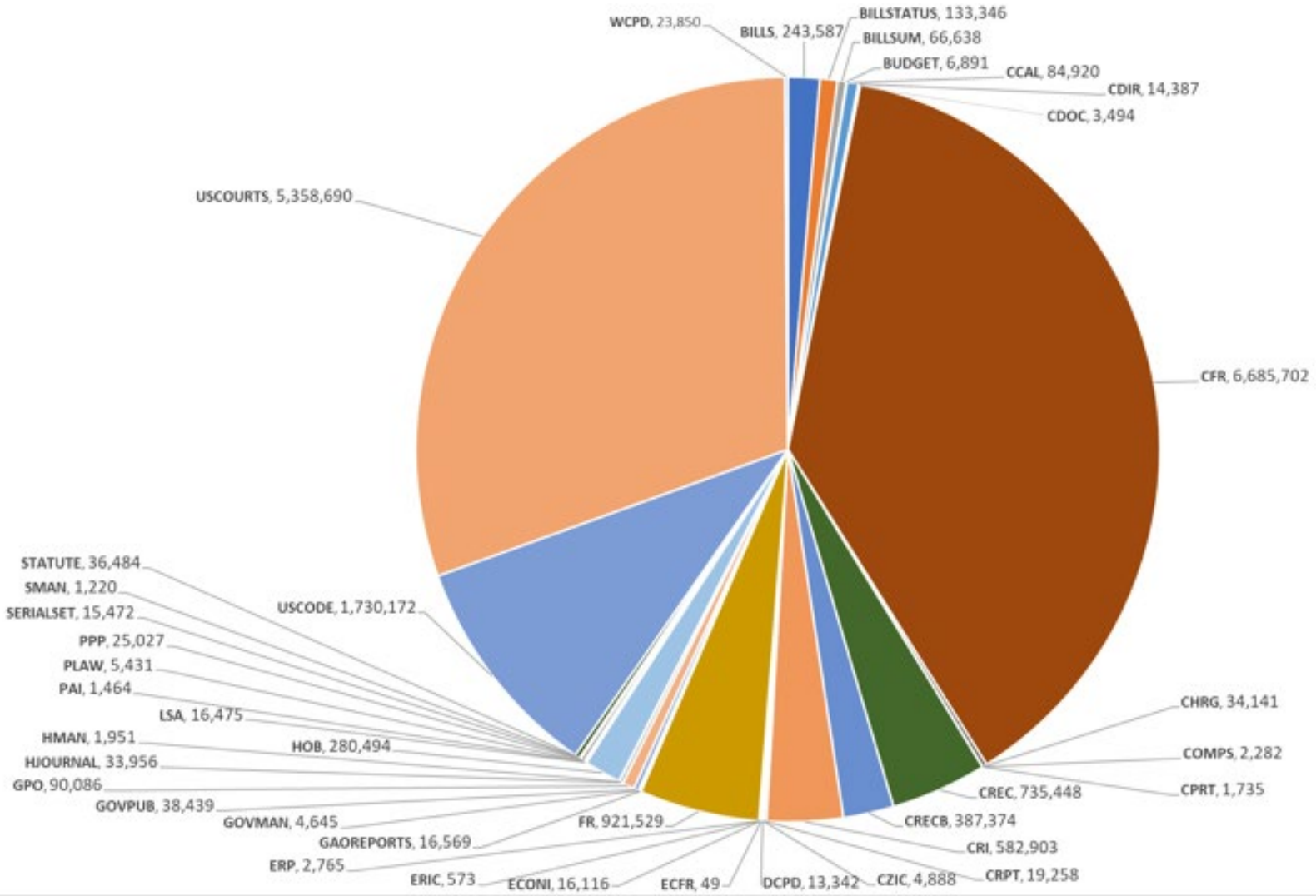
GovInfo Packages

October 2022
2.3M



GovInfo Granules

October 2022
17.6M



Live Demo

Release Notes

https://www.govinfo.gov/features/release-notes

Questions