

GPO

Data Modernization



September 10, 2020

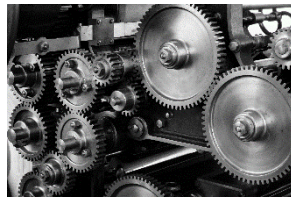
Programs, Strategy, and Technology



XPub

XPub

XML-based composition ecosystem to replace legacy Microcomp system and locator-coded text format



Accomplishments

United States Code

Composed 2018 main edition with XPub cutting production time by 50 percent.

Tools and Workflows for OFR

Created web-based tools and workflows for Public Laws and the Statutes at Large to produce USLM XML and print-ready PDFs via the Oxygen XML editor and XPP composition engine.

In-Process

Bills, Resolutions, and Amendments, Public Laws, and the Statutes at Large

Deploy XPub to House and Senate systems and begin composing with XPub for the 117th Congress.

Reports Prototype

Convert styled Microsoft Word files into USLM XML, compose the USLM XML with XPub, and provide flexibility to explore modern formatting options with the House and Senate.

Roadmap

House and Senate Calendars

Use XPub to compose House and Senate calendars.

United States Code USLM Update

Update USLM files to add information required to efficiently compose future main editions through XPub.

USLM for Remaining Bill Versions

Model data and create transforms for remaining bill versions via XPub.

Publications, Formats, & Functionality

Deploy iterative releases, create templates for authoring in Word, convert files to USLM XML and typesetting XML, and create a web interface for composition.

XPub Enterprise Application

- Content intake, data conversion, and job monitoring tools.
- Integrated workflow management and OCR scanning tools.
- Digital proof reading, correcting, and editing functionality.
- Support for structural and semantic XML markup at scale.
- Job assembly tools to produce print and digital products.



USLM

USLM

United States Legislative Mark-up, the modern XML schema being embraced by legislative data partners



Accomplishments

Enrolled Bills, Public and Private Laws, and the Statutes at Large

Building upon the successes of the House Office of the Law Revision Counsel’s release of the U.S. Code in USLM XML and the House Rules Committee’s release of the House Rules and Manual in USLM XML—

Performed data modeling activities, created transforms to convert enrolled bills, public and private laws, and Statutes at Large files from the 113th Congress forward from locator-coded text files into USLM XML, and made the beta files available on govinfo.

Federal Register and Code of Federal Regulations Prototype

Performed initial data modeling activities, created prototype transforms to convert a subset of the Federal Register and Code of Federal Regulations from SGML files into USLM XML, and made the prototype files available for public review and comment on GPO’s USLM GitHub repository.

In-Process

COMPS in USLM

Perform data modeling activities and create transforms to convert over 2,000 Statute Compilations into USLM XML, and make the files available on govinfo.

Roadmap

USLM for Remaining Bill Versions

Model data and create transforms for remaining bill versions via XPub.

Bulk Data Task Force / Legislative XML Working Group’s Roadmap for Publishing More Legislative Documents in USLM XML Format

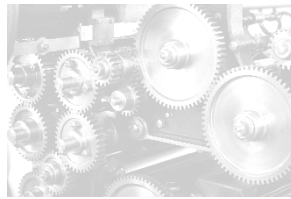
- Committee Reports
- Congressional Record, including the Daily Digest
- Hearings
- Calendars
- Precedents of the U.S. House of Representatives
- House Journal
- Additional Legislative Publications



govinfo

govinfo

*ISO 16363 Trustworthy
Digital Repository system
ensuring free online access
today and into the future*



Accomplishments

XML Metadata

METS, MODS, and PREMIS XML metadata for all govinfo content.

XML Bulk Data Repository

One-click downloads plus XML and JSON endpoints for 12 collections.

Bulk Download Search Results

Available in CSV and JSON formats.

API

api.govinfo.gov also includes an endpoint for Bill Status bulk data.

Developer Hub

Visit govinfo.gov/developers for info about RSS feeds, sitemaps, the Link Service, and more!

GitHub Repositories

Review documentation, submit issues, and request features via github.com/usgpo.

MARC Records on GitHub

Download MARC record sets extracted from GPO's Catalog of U.S. Government Publications in UTF-8, MARC-8, and MARC XML formats at github.com/usgpo/cataloging-records.

In-Process

Technology Refresh

Hardware and storage refresh included moving to a completely virtualized server infrastructure and upgrading applications and operating systems.

Files from XPub

- Bills – Responsive HTML without hard returns, beta USLM XML for enrolled bills, bill DTD XML, and PDF.
- Public and Private Laws and the Statutes at Large – Responsive HTML without hard returns, beta USLM XML, and PDF.

Curated Content Search

Search feature for curated govinfo content such as release notes and articles.

Roadmap

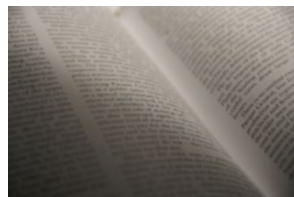
API Search Service

Visit github.com/usgpo/api and enter an issue for new features or functionality.

Content

Content

Current and historical Federal Government information from all three branches



Accomplishments

Notable Historical Content

Bound Congressional Record to 1873; Federal Register to 1936; Statutes at Large to 1951; United States Government Manual to 1935; Public Papers of the Presidents to 1929; GPO Style Manuals; various House and Senate hearings; and executive agency, board, and commission publications including the U.S. Army in the World War, 1917-1919, NIST publications, and over 11,000 Railroad Retirement Board decisions and opinions.

In-Process

10 Additional Years of Bill Status XML Bulk Data Files

In partnership with the Library of Congress, House, and Senate, make Bill Status XML back to the 108th Congress available as bulk data on govinfo in early FY21.

Digitized Serial Set

In partnership with the Library of Congress, provide access to the digitized Serial Set back to 1817.

Roadmap

Upcoming Digitized Content

Additional House and Senate committee hearings, House and Senate committee prints, and the Congressional Directory.

H. Rept. 115-696

In H. Rept. 115-696, the report that accompanied the FY19 House legislative branch appropriation bill, the House Appropriations “Committee direct[ed] the GPO to assess the costs associated with converting Statutes at Large from 1789-2002 into USLM XML format.”

In FY21, GPO is moving forward with a project to digitize the remaining Statutes at Large volumes back to Volume 1 (1789-1799). As part of the project, GPO plans to prototype the conversion of a subset of digitized Statutes at Large into USLM XML in order to assess the technical feasibility and determine future budgetary requirements.

