CHERYL L. JOHNSON
CLERK

GLORIA J. LETT
DEPUTY CLERK

ROBERT F. REEVES
DEPUTY CLERK

Office of the Clerk
U.S. House of Representatives
Washington, DC 20515–6601

May 11, 2020

The Honorable Zoe Lofgren
Chairperson
Committee on House Administration
1309 Longworth House Office Building
Washington, DC 20515

The Honorable Rodney Davis
Ranking Member
Committee on House Administration
1216 Longworth House Office Building
Washington, DC 20515

Dear Chairperson Lofgren and Ranking Member Davis,

As required by Section 501 of House Resolution 756, agreed to by the House on March 10, 2020, I am attaching the initial report: Adopting Standardized Formats for Legislative Documents.

If you or others on the committee have questions about this report, please do not hesitate to contact me.

Sincerely,

Clerk of the House

Enclosure

Adopting Standardized Formats for Legislative Documents

As required by Section 501 of House Resolution 756, agreed to by the House on March 10, 2020, the Clerk of the House submits this initial report about adopting standardized formats for legislative documents.

Background

As the members of the Committee know, the House has the responsibility of preparing, managing, distributing and preserving (archiving) our official documents. The production and exchange of our Congressional legislative documents is a joint effort accomplished through the work of four primary entities: The House, the Senate, the Government Publishing Office, and the Library of Congress. Staff across these four organizations work together to ensure that both the official document of record, along with its paper and digital representations are accurate and accessible to Members of Congress, congressional staff, and the public in a timely manner.

While the work of Congress continues to be driven by paper, Congress addressed the need to modernize its electronic data exchange in the mid-1990s. Today's work related to the United States Legislative Markup (USLM) schema is a direct outcome of over twenty-five years of work accomplished by current and former staff, Members, vendors, contractors and others. Specifically, this work is rooted in section 314 of Public Law 104-197[1] (2 U.S.C. 181[2]), as well as in a 1996 directive from the chairs of this committee and the Senate Committee on Rules and Administration.

This directive instructed the Clerk of the House and the Secretary of the Senate, respectively, to work together to establish common data standards for the exchange of legislative information. Following the directive, Congress embarked upon an initiative to establish a document exchange standard for the transformation, creation, and exchange of legislative documents in XML format.

XML[3] is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. For meaningful online interaction, machine-readable formats for data and document representation are a must. XML is one such format and made especially for long pieces of text like that found in legislative documents.

To coordinate and manage Congress's data standards, the Legislative Branch XML Working Group was established in 1996. The Working Group consists of non-partisan, technical staff from the House, the Senate, the Government Publishing Office, the Library of Congress, and other legislative branch organizations. It is co-chaired by two (2) professional staff—one each from the House Clerk's office and Senate Secretary's office, respectively. The primary responsibility of the Working Group is to manage and maintain the XML schemas owned by Congress. These schemas are used to exchange legislative documents between the House and Senate, as well as other Legislative Branch organizations. The Working Group also ensures that the schemas are publicly available. Decisions in the Working Group are

---

[1] https://www.govinfo.gov/content/pkg/PLAW-104publ197/pdf/PLAW-104publ197.pdf

[2] http://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title2-section181&num=0&edition=prelim

[3] Defined by The World Wide Web Consortium (W3C) at (https://www.w3.org/XML/) Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879)

made by consensus—a testament to the working relationships that our non-partisan, professional staff have across the Legislative Branch.

The 1996 directive included an initiative to customize XML editors for the creation of bills and resolutions in the House and Senate. In 2000, the Clerk of the House presented the SGML/XML Feasibility Study Final Report to the House Committees on Rules and Administration, and in 2001, the House Office of Legislative Counsel began to use an XML editor to draft legislation. The proliferation of Congressional XML continued in the early 2000s with House votes in 2003, House and Senate bills in 2004, and Senate Votes in 2009[4].

The work of the XML Working Group was provided much needed support by the creation of the Bulk Data Task Force. Specifically, on June 1, 2012, House Report 112-511 that accompanied H.R. 5882— Legislative Branch Appropriations Act of 2013 directed the establishment of the House Bulk Data Task Force (BDTF) to expedite the process of providing bulk access to legislative information. The Task Force is a partnership of legislative branch agencies, including the House of Representatives (Office of the Clerk, Office of General Counsel, Office of Law Revision Counsel, Office of the Parliamentarian, Legislative Branch Subcommittee, Committee on House Administration, Committee on Rules, and House Leadership), Senate, Library of Congress, the Government Publishing Office, and the Government Accountability Office that collaborate to bring about openness and transparency. One of the group's primary goals is to provide to the public more data and documents online and, over time, in bulk and machine-readable formats using XML standards.

The House, via the adoption of H. Res 756, continues to support the use of the XML, specifically the USLM schema for all its legislative documents and data. Staff of the XML Working Group and the BDTF are grateful for this continued support.

The XML generated since 2001 can be characterized as **"First-Generation"**— meaning the markup primarily expresses the visual characteristics of a document. The First Generation is colloquially referred to as BILL DTD XML. It is still in use today to create, edit, and publish bills, resolutions, and amendments.

In 2013, the Office of the Law Revision Counsel with assistance from the staff on the XML Working Group (and support from this committee) launched Congressional XML into its **"Second-Generation"** with the release the United States Code using the United States Legislative Markup (USLM) schema[5]. The USLM schema, which is interoperable with the LegalDocML international standard[6], goes beyond what the First-Generation XML can do. Not only does the USLM describe the presentation of a document, it also expresses the structure of a document (how it is organized) and the semantics of a document (what it means). Among other things, this structure will enable richer searching and an increased ability to link documents and data to one another. Simply put, USLM is an XML information model designed to represent United States legislative and regulatory documents in a schema that is compatible with international standards. (International standards that the House and the Library of Congress were involved in setting.)

---

[4] http://xml.house.gov/resources/TechTimeline.htm
[5] http://uscode.house.gov/download/resources/USLM-User-Guide.pdf
[6] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legaldocml

The "More Documents in USLM" Roadmap

Following the development of the USLM schema, staff within the Legislative Computer System division of the Clerk of the House laid out a strategy and priority for moving legislative publications into USLM XML. The Bulk Data Task Force (and the XML Working Group) adopted this strategy in the form of the roadmap included below. This roadmap was first presented publicly to the Committee on House Administration during then-Clerk Karen Haas' testimony on October 11, 2017[7]. Subsequently, Deputy Clerk Robert Reeves included the roadmap in his testimony before the Committee on the House Modernization on May 20, 2019[8].

Roadmap for Publishing Legislative Documents in USLM XML Format

Completed	U.S. Code
Completed	Enrolled bills, Public Laws, and Statutes at Large
IN PROGRESS	HOLC/SOLC statute compilations
B	Remaining bill/resolution versions (introduced, reported, engrossed, all amendments, et. al.)
C	Committee Reports
D	House portions of the *Congressional Record*, including the *Daily Digest*
E	Hearing Record
F	House Calendar
G	*Precedents of the U.S. House of Representatives*
H	*House Journal*
NOTE: *This table is House-centric in many respects and can be updated to include the Senate versions of like documents.*	

Many legislative documents contain portions of other legislative documents and this creates a natural and a logical progression to the roadmap and an order to which documents should be converted to USLM. The progression is to start with the law (codified and non-codified provisions), then proceed to bills and resolutions, followed by documents that contain bill and resolution language. The development of this strategy was a direct result of lessons learned from work with the First-Generation BILL-DTD XML as well as years of work across our legislative branch organizations to gather and understand this particular set of documents.

Past projects

The U.S. Code (Stage one of the OLRC House Modernization Project)

The United States Code was released in USLM to the public on July 29, 2013. In addition to releasing the Code in USLM, a stylesheet for online presentation of the USLM was created, and a comprehensive user guide was written.

---

[7] https://docs.house.gov/meetings/HA/HA00/20171011/106494/HHRG-115-HA00-Wstate-HaasK-20171011.pdf
[8] https://docs.house.gov/meetings/MH/MH00/20190510/109468/HHRG-116-MH00-Wstate-ReevesR-20190510.pdf

This project was led by the Office of the Law Revision Counsel with assistance from the Office of the Clerk, GPO, the Secretary of the Senate and other staff from the XML Working Group and the Bulk Data Task Force. Prior to its public release, the XML Working Group agreed upon the schema in November 2012. With this agreement, the USLM schema was added to the set of schemas managed by the Legislative Branch XML Working Group.

The U.S. Code, the USLM schema, and related files (stylesheets and User Guide) are available at https://uscode.house.gov/download/download.shtml.

Following the initial adoption of the USLM schema, the Bulk Data Task Force, as the request of the XML Working Group, directed GPO in making the USLM XML schema available as an authoritative source on GitHub. It can be found at https://github.com/usgpo/uslm.

Enrolled Bills, Public Laws, and Statutes at Large

In December of 2018, a subset of enrolled bills (113th Congress forward), public laws (113th Congress forward), and the Statutes at Large (108th Congress forward) were released to the public in USLM XML.

This project was led by GPO with assistance from the Office of the Clerk, the Office of the Federal Register, the Secretary of the Senate and other staff from the XML Working Group and the Bulk Data Task Force. Prior to its public release, the XML Working Group agreed upon changes to the USLM schema during the second quarter of 2018. In addition, the XML Working Group released the proposed schema changes on GPO's GitHub pages for public comment and feedback.

These datasets can be found on govinfo.gov—

- Enrolled measures https://www.govinfo.gov/bulkdata/BILLS/uslm
- Public and private laws https://www.govinfo.gov/bulkdata/PLAW
- Statutes at large https://www.govinfo.gov/bulkdata/STATUTE

House Rules and Manual

While not on the current roadmap, the *Jefferson's Manual and the Rules of the House of Representatives* was released in USLM to the public on March 23, 2016.

The project was led by the House Committee on the Rules with assistance from the Office of the Clerk, the House Parliamentarian and GPO. Prior to its release, the XML Working Group agreed upon small changes to the schema. The XML Working Group released the files on GPO's GitHub account for public comment and feedback (https://github.com/usgpo/house-manual). Since the initial release, the Rules Committee, the Parliamentarian's office, the Clerk's office and GPO have worked together to publish subsequent editions of this publication in USLM.

This publication can be found on govinfo.gov—

- https://www.govinfo.gov/app/collection/hman
- https://www.govinfo.gov/bulkdata/HMAN

Current Project(s)

Statute Compilations

Current project work involves the Statute Compilations[9] maintained by the House Legislative Counsel. Each Statute Compilation incorporates the amendments made to the underlying statute since it was originally enacted. This project is led by GPO with assistance from the Office of the Clerk, the Offices of the House and Senate Legislative Counsels, the Secretary of the Senate and other staff from the XML Working Group and the Bulk Data Task Force.

A set of 319 Compilations is now available on govinfo in PDF format and additional Statute Compilations will be added over time. This effort is the first phase of a project to provide a uniform set of laws in United States Legislative Markup (USLM) to enable downstream processes and increase efficiencies. The second phase of this project currently underway is to convert legacy Statute Compilations files into Beta USLM XML and provide access to those files on govinfo. It is anticipated that the project will be completed by end of the year.

Future projects

As you can see in this report, work with regard to transforming legislative documents into USLM XML is done in partnership with GPO, the Senate, the LOC and other House organizations, in particular HOLC and OLRC.

The current cross-organizational set of projects build upon the tools and successes of the previous projects. As we work towards the upcoming phases of the roadmap, the XML Working Group in consultation with the Bulk Data Task Force, will continue to analyze and recommend next steps that build on the work that has already been accomplished. Additionally, future recommendations will include analysis related to how past projects and systems may be impacted. As with any data standardization efforts, interoperability is a key goal and project work must address any potential impacts on those documents and systems currently in production.

Future Costs

Since 1996, costs for document standardization (and the related tools) have been shared across the Legislative Branch. We do not anticipate any changes in how this work is accomplished. In future years, we do anticipate that the Clerk's office and other data partners across the Legislative Branch will request additional funds to complete projects related to USLM XML, particularly with regard to upcoming phases of the roadmap, editing legislative documents in USLM, and publishing the documents using GPO's in-development XPub.

---

[9] The Office of the Legislative Counsel of the U.S. House of Representatives maintains a corpus of Statute Compilations of public laws that either do not appear in the U.S. Code or that have been classified to a title of the U.S. Code that has not been enacted into positive law

Related Project Work

At this time, in addition to the work to transform our documents into USLM described in this report, there is a cross-organizational set of systems that utilize the USLM schema. Each project builds upon the tools and successes of previous projects. A key benefit of implementing USLM XML for Congressional documents is increased efficiency, transparency, and interoperability. Each of these current projects receive active assistance from staff on the XML Working Group and the Bulk Data Task Force.

These projects are—

- OLRC Modernization project Stage 2: Hammurabi (codification bills), led by the Office of Law Revision Counsel
- OLRC Modernization Stage 3: Coda (U.S. Code editor), led by the Office of Law Revision Counsel
- The Comparative Print Project, led by the Office of the Clerk (and the House Office of Legislative Counsel)
- XPub (formerly Composition System Replacement), led by the GPO
- Tools to display the USLM XML on Legislative Branch-wide websites, led by GPO

In addition to the projects listed above, the current systems and technologies listed below use the generation-one bill-dtd XML and are actively maintained—

- Tools to display the generation-one XML online, led by GPO and LOC
- The XML editors used by the House and Senate to draft legislation, commonly called XMetaL and LEXA (maintained by the House Clerk and Senate Secretary's office with assistance from the House and Senate Legislative Counsels)
- The Ramseyer Editing System and Statute Maintenance system (maintained by the House Legislative Counsel with assistance the House Clerk and GPO)

Future Quarterly Reports

In subsequent quarterly reports as required by H. Res. 756, we will continue to report on current and future project work related to USLM. Current XML Working Group staff and staff on the Bulk Data Task Force are committed to working as a team across their organizations to accomplish involved in using XML schemas to create, edit, and publish legislative documents.