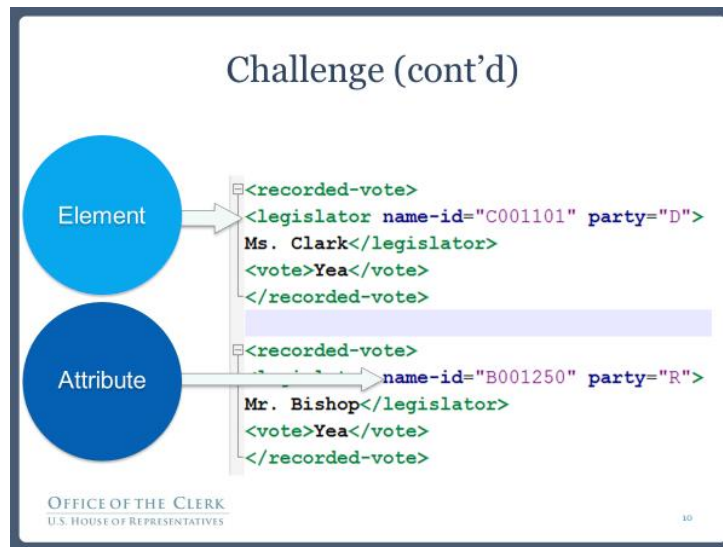

**STATEMENT BEFORE THE SELECT COMMITTEE ON THE
MODERNIZATION OF CONGRESS**

ROBERT F. REEVES, DEPUTY CLERK
MAY 10, 2019

Chair Kilmer, Vice Chair Graves, Members of the Committee: Thank you for the opportunity to testify about the accomplishments and ongoing efforts of the Office of the Clerk and the Bulk Data Task Force (BDTF) to make legislative data easily accessible, accurate, timely, free, reusable and available in standard machine-readable formats.

Today the House has the ongoing challenge of preparing, managing, distributing and archiving its official documents both on paper and in digital formats. There is no longer a question of will we have publicly available documents but rather how do we modernize our systems and workflows to utilize today's newer technologies to help citizens and staff access free and timely legislative and legal information. A challenge for the House is to examine our own policies and procedures so that we can support online access to bills, amendments, votes and proceedings in structured and machine-readable formats.

There are several machine-readable formats such as JSON and XML. These formats are different from traditional word processing documents and portable document format (PDF) files, which are generally difficult for machines to interpret. The House's preferred standard for legislative documents is XML. XML is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. An XML document has a structure that lends itself to automated analyses; it is not meant to just display information on a screen or on paper for "publishing" purposes, but rather for automated "processing" of information. Below is a sample of XML coding:



In 2011, the rules of the US House of Representatives for the 112th Congress called for greater transparency and accessibility to legislative documents. The intention of the change was to place electronic distribution on par with traditional printing.

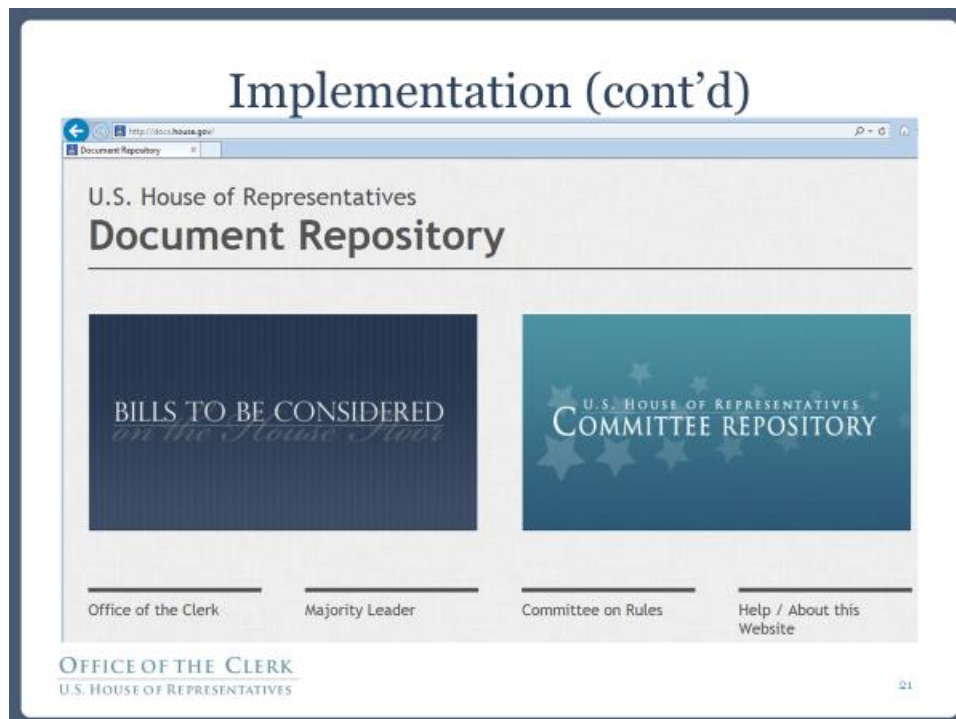
BULK DATA TASK FORCE

H. Rept. 112-511 that accompanied H.R. 5882 Legislative Branch Appropriations Act 2013, was passed by the House on June 8, 2012 and directed the establishment of the House Bulk Data Task Force to examine increased dissemination of congressional information via bulk data download by non-governmental groups supporting openness and transparency in the legislative process. The Bulk Data Task Force is a partnership of representatives from the various Legislative Branch Agencies such as: House of Representatives, Senate, Library of Congress, Government Publishing Office, that work in a cooperative manner to coordinate and accomplish openness and transparency goals. The BDTF meets approximately three times a year internally and two to three times a year with the Civil Society groups in public meetings.

In 2013, the House through the Bulk Data Task Force (BDTF) decided to move away from the current generation one XML structure and encouraged the use of the international standard commonly called Akoma Ntoso to the greatest extent practicable. Some structures of the U.S. Code could not be handled by Akoma Ntoso at the time of the project, so the House, in consultation with the Office of Law Revision Counsel, the Legislative Branch XML Working Group and the House's contractor designed its own generation two XML schema called United States Legislative Markup (USLM) which has now become our standard going forward. Having a single data standard can provide many benefits:

- Reuse and transformation of source documents into other formats ensuring consistency and quality
- Improved data exchange internally or across borders
- Decrease amount of human error in drafting
- Long term preservation of parliamentary documentation
- More accurate search capabilities
- Improved time for typesetting and printing of the document
- Easier tracking of document versions
- Unprecedented level of possibilities for data analysis. Interpretation and visual representation of content
- Wider multichannel dissemination of information through different media, as well as facilitated access for persons with disabilities.

The BDTF with the approval of Leadership began to move more data and documents online in bulk using the USLM/XML standard. In 2013 the first project, Bills in Bulk Data was completed. At the same time the Office of the Clerk was completing a project to develop the docs.house.gov website to provide access to legislative proposals (bills, resolutions and amendments) at a crucial point in the legislative process that was not previously available to Members, staff and the public.



Since then many other projects have been completed and can be seen on our publicly available Technology Timeline <https://xml.house.gov/resources/TechTimeline.htm>

- 2013 House bills in XML. House bills were made available as XML bulk Data on the FDsys Bulk Data Repository. Senate bills were added to the collection in January 2015.
- 2013 U.S. Code made available in XML. In July the Office of Law Revision Counsel made US Code available in XML. This was the first release of “generation two” schema named “United States Markup Language” or USLM.
- 2013 Redesign of the THOMAS system by the Library of Congress. Combining the functions of the internal Legislative Information System (LIS) with THOMAS the public system. One system for everyone (Members, staff, and the public) at <https://www.congress.gov>



- 2013 Data Challenges - the Library of Congress announced the winner of Data Challenge I.

First Legislative Data Challenge Winner Announced

December 19, 2013 by [Tina Gheen](#)

Today we have the great pleasure of announcing Jim Mangiafico as the winner of our [first legislative data challenge](#), [Markup of US Legislation in Akoma Ntoso](#) and the \$5,000 prize. The challenge, which was open from July 16 to October 31, invited participants to create XML versions of US bill text using the [Akoma Ntoso](#) standard. We also asked for feedback about the Akoma Ntoso standard itself to help ensure US legislation can be incorporated properly into this international framework.

The decision was a difficult one for [our three judges](#), but in the end Mr. Mangiafico's submission, [Four US Legislative Documents in Akoma Ntoso](#), earned the highest overall score in the challenge. The judges were particularly impressed by the complexity of his application of the metadata section of Akoma Ntoso to the bill text and the submission's cascading style sheet.

We want to thank everyone who participated in the challenge. We all learned a great deal from the entries we received, and we hope everyone who entered will continue to contribute to the development of the Akoma Ntoso standard.

- 2014 House bill summaries were made available as XML bulk data on FDsys Bulk Data Repository. Senate bill summaries were added to this collection in January 2015
- 2014 the Library of Congress announced the Data Challenge II winner (an international contest to map US and UK legislative documents to Akoma Ntoso) winner(s).

February 25, 2014 the Library of Congress Announces Legislative Data Challenge Winners

The Library of Congress today announced the winners of two legislative data challenges conducted using the Challenge.gov platform. The challenges, conducted at the request of the U.S. House of Representatives, will advance the exchange of legislative information worldwide.

Jim Mangiafico is the winner of the "Markup of U.S. Legislation in Akoma Ntoso" challenge. Mangiafico and Garrett Schure won first and second place, respectively, in the "Legislative XML Data Mapping" challenge.

"Markup of U.S. Legislation in Akoma Ntoso" invited competitors to apply the Akoma Ntoso schema to the text of four U.S. federal bills. Akoma Ntoso is a framework used in many other countries around the world to annotate and format electronic versions of parliamentary, legislative and judiciary documents. Mangiafico was awarded first place and a \$5,000 prize for his submission, which explained the practical decisions and difficulties required to apply the Akoma Ntoso schema to U.S. bill text. Mangiafico, who earned a doctorate degree from Vanderbilt University and a J.D. degree from the University of Michigan Law School, is an attorney in the Washington, D.C., metropolitan area.

"Legislative XML Data Mapping" invited competitors to map the Akoma Ntoso schema to established U.S. and United Kingdom legislative markup languages to improve Akoma Ntoso's support of U.S. and U.K. legislative data. Mangiafico's submission earned first place and a \$10,000 prize. Schure's submission garnered second place and a \$5,000 prize. Both submissions accomplished the mapping using open source software, which resulted in foundational options that can be learned from and expanded on by others.

- 2014-2015 the Office of the Clerk released the new Electronic Financial Disclosure system. The Ethics in Government Act of 1978 ("EIGA"), as amended by the STOCK Act, requires the Clerk of the House to deploy an electronic filing system for all forms filed pursuant to EIGA, by January 2014. All forms required by EIGA, including Form A and Periodic Transaction Reports (PTRs) are available for electronic filing
- In January 2015 the House transitioned to an online Lobby Disclosure Act Filing system and paper forms were discontinued
- 2015 the House began publishing basic Member data in XML format

- 2015 Stage 2 of the Office of Law Revision Counsel and House Office of Legislative Counsel’s House Modernization Project (Positive Law Codification System) was made operational
- 2016 House and Senate bill status information was made available in XML format on GPO’s Bulk Data Repository. Sample files were released for comment on GitHub prior to the release through FDsys, govinfo and Bulk Data Repository
- 2016 GPO’s Next Generation FDsys website <https://www.govinfo.gov> released



- 2016 Speaker Ryan announces the start of the “More Documents in United States Legislative Markup (USLM)” project which was approved by the House Legislative Branch Appropriations Subcommittee. This multi-year project seeks to convert enrolled bills, public laws, and the Statutes at Large from GPO locator-coded text format into an XML format using the USLM schema and to make the resulting XML available on GPO’s Federal Digital System (FDsys) / govinfo. The first phase of this project is to develop three stand-alone, java-based utilities that build upon the House’s existing U.S. Code utility to transform from GPO locator-coded text to USLM XML. It is assumed that separate utilities will be required for enrolled bills, public laws and Statutes at Large. This project is important because now that we have finished the creation of bills, bill summaries and bill status in bulk data XML format using the USLM schema, it allows us to go back and start converting all our other legislative documents into the same new standard. XML files are now available on Govinfo.gov
- 2016 the Office of the Clerk makes available a new mobile-friendly version of the House Telephone Directory (<https://directory.house.gov>) designed to help reduce the number of print copies of the House Telephone Directory and provide online access to its information

- 2016 the Office of the Clerk made an alpha version of its new website’s main page <https://clerkpreview.house.gov> and streaming video replacement page <https://live.house.gov> available to a controlled group of internal users for feedback.
- 2017 new House Rules called for greater transparency and accessibility to legislative documents. For the 115th Congress, the Rules also called for more tools to help the Members, staff and public see changes to existing law made by proposed legislation in certain stages of the legislative process. Specifically, three provisions were added to the House Rules. These provisions are being called the Posey Comparative Print rules. They read as follows:

(s) COMPARATIVE PRINTS FOR BILLS OR JOINT RESOLUTIONS CONSIDERED ON FLOOR. — Effective December 31, 2017, in rule XXI, add at the end the following new clause:

“12. (a) (1) Before a bill or joint resolution proposing to repeal or amend a statute or part thereof may be considered, there shall be made available on a publicly available website of the House an easily searchable electronic comparative print that shows how the bill or joint resolution proposes to change current law, showing **(to the greatest extent practicable)** by appropriate typographical devices the omissions and insertions proposed.

“(2) Before an amendment in the nature of a substitute may be considered if the amendment proposes to repeal or amend a statute or part thereof, there shall be made available on a publicly available website of the House an easily searchable electronic comparative print that shows **(to the greatest extent practicable)** how the amendment proposes to change current law, showing by appropriate typographical devices the omissions and insertions proposed.

“(b) If a committee reports a bill or joint resolution, before the bill or joint resolution may be considered with text different from the text reported, there shall be made available on a publicly available website of the House a document **that shows, by appropriate typographical devices**, the differences between the text of the bill or joint resolution as proposed to be considered and the text of the bill or joint resolution as reported.”.

Requirements were developed, and contracts were awarded to two vendors with experience in working with the House and the House Office of Legislative Counsel (HOLC). Deliverables were made by the one-year deadline.

- 2018 Work continues the Posey Comparative print solution with the goal of providing more House staff access to the program
- 2018 The Library of Congress released new searchable House Executive Communications

- 2018 the BDTF released the Legislative Branch Innovation Hub which highlights Legislative Branch activities that use technology to cultivate collaboration, foster data standardization, and increase transparency. The Hub includes information to help law-making bodies innovate, communicate, and legislate effectively, and it is being developed iteratively by non-partisan support staff in United States Legislative Branch organizations coordinated through the Bulk Data Task Force in collaboration with civil society using GitHub Pages. It can be found at: <https://usgpo.github.io/innovation/>.
- 2018 “More Documents in USLM” a multi-year project that converts enrolled bills, public laws, and the Statutes at Large from GPO locator-coded text format into an XML format using the USLM schema and makes the resulting XML available on GPO’s FDsys/govinfo was completed. The scope for the first phase of the project includes enrolled bills from the 113th Congress forward, public laws from the 113th Congress forward, and the Statutes at Large from the 108th Congress forward. Beta USLM XML is available on govinfo.
- 2018 GPO became the first organization in the United States and the second organization in the world to receive ISO 16363 certification for its digital repository system, govinfo
- 2018 Library of Congress combined House and Senate Committee Calendar
- 2018 CRS releases CSR Reports at <https://crsreports.congress.gov/>
- 2019 Office of the Clerk completes the new Truth in Testimony Form

The Bulk Data Task Force continues to monitor projects across the Legislative Branch and foster collaboration among the House, Senate, Library of Congress and GPO. One key project is transforming legislative documents into the USLM XML schema. The BDTF developed the project roadmap below. There is a logical progression and order to which documents should be put in USLM. The logical progression requires that the law goes first, including both positive and non-positive law, then, the bills, resolutions and amendments, followed by documents that contain bills and resolution language. The roadmap is House-centric but could easily be expanded to include the Senate versions of like documents.

In support of the BDTF, GPO, the Clerk of the House, the Secretary of the Senate and the Office of the Federal Register, GPO worked on a project to convert a subset of key legislative documents including enrolled bills, public laws, and the Statutes at Large into USLM. In early 2019 the first project was completed with the release of Beta USLM XML on GPO’s govinfo website. Additionally, in collaboration with the House Office of Legislative Counsel, the Senate Legislative Counsel, the Clerk of the House and the Secretary of the Senate, GPO is now making select Statute Compilations available in PDF format as a pilot on GPO’s website, govinfo. The second phase of this project will be to convert legacy Statute Compilations files into USLM and provide access to those on GPO’s website, govinfo.

Bulk Data Task Force

- “More Documents in USLM” – Project Roadmap

	Publication to be put in USLM XML format
Completed	U.S. Code
Completed	Enrolled bills, public laws, and the Statutes at Large
Current project	HOLC/SOLC compilations (non-positive law)
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

OFFICE OF THE CLERK
U.S. HOUSE OF REPRESENTATIVES

12

Other projects include Stage 3 of the Office of Law Revision Counsel’s Modernizations project, continual improvements to Congress.gov as well as others.

While the Clerk's Office monitors and coordinates a number of transparency projects via the BDTF we also take the lead on several transparency related projects. I would like to discuss those projects now.

Comparative Print Project

After the successful completion of Phase 1 of the Comparative Print project at the end of 2017, the Clerk's Office has continued to work with the Office of the Legislative Counsel and the Law Revision Counsel on the planning and development of subsequent phases. In August 2018, the Clerk's Office awarded the contract for Phases 2 and 3 of the Comparative Print project. The long-term goal of the project is to deploy a robust, scalable and secure platform so all House staff (and potentially others) can create on-demand static comparative prints that illustrate the changes between:

- Two versions of a bill, resolution or amendment (document to document comparison)
- Current law and current law as proposed to be changed by amendments contained in a resolution or amendment to current law (positive and non-positive law)
- A bill resolution and the bill or resolution as proposed to be modified by amendments (amendment impact).

Phase 2 is anticipated to last from August 2018 to August 2019 with the following deliverables:

- Develop and build the backend framework to include security, user management, document repository, and document transformation—in progress, second quarter 2019
- Build natural language processor (NLP) to interpret amendatory language —initial design complete (fourth quarter 2018), in progress, first/second quarter 2019
- Implement base set of amendatory language and automated execution to create a stand-alone document that meets the provisions of clause 12(a) of House Rule XXI— second quarter 2019
- Implement an editor for those cases in which user intervention is needed— second/third quarter 2019
- Continuously improve the comparison document in HTML and PDF that meets the provisions of clause 12(b) of House Rule XXI—Second/third quarter 2019
- Improve billcompare.house.gov user interface—second/third quarter 2019
- Develop outputs in HTML, PDF, and XML—second/third quarter 2019.

Phase 2 of the project is on track and we are holding weekly status meetings with the vendor.

We expect Phase 3 of the project to begin later this year and anticipate it will last from August 2019 to August 2020. Below are the deliverables for Phase 3:

- Continue to build and refine NLP engine and implement an extended set of amendatory language—first/second quarter 2019

- Refine Phase 2 functionality: automation, editor, transformations, and user interface— third quarter 2020
- Migrate the current Amendment Impact Program (AIP) to the Posey Technology— fourth quarter 2019
- Deploy to larger user base—second quarter 2020
- Develop outputs in HTML, PDF, and XML—third quarter 2020.

If we do release this tool to the entire House community determining how we will provide support for it will be a challenge.

Consensus Calendar

Section 102, Subsection (r), of H.Res. 6 provides for a Consensus Calendar in the 116th Congress House Rules and mandates that the Speaker must designate, and the House must consider, at least one measure on the Consensus Calendar during any week in which the House convenes (except before March 1 at the beginning of a Congress and after September 30 at the end of a Congress). This subsection also provides that to be eligible for placement on the Consensus Calendar a measure must accumulate 290 cosponsors and must not have been reported by its primary Committee of jurisdiction. When this cosponsorship threshold is reached the sponsor of the measure may, while the House is in session, submit to the Clerk a written motion to place the measure on the Consensus Calendar. If these conditions have been met, the Clerk will note the motion's submission in the *Congressional Record* of that day and enter the motion on a comprehensive list of Consensus Calendar Motions, which will be available on the Clerk's website. After a measure has maintained at least 290 cosponsors for a cumulative period of 25 legislative days after the presentation of the motion, the measure will be placed on the Consensus Calendar, where it will remain until it is considered in the House or reported by the primary Committee. The Clerk's Office has developed a process that will ensure it meets the Consensus Calendar requirements.

Website Redesign

The Clerk's Office has released the Beta version of our new redesigned website at <https://clerkpreview.house.gov>. The new homepage highlights House Floor activity and provides easy access to vital legislative information in a user-friendly, responsive, and modern design. We have added several features, including a public vote search page, updated Member and Committee profiles, and a new video streaming site (<https://live.house.gov>) that includes the ability to rewind the live feed and search closed captioned text. The Clerk's Office also released an Alpha version of a new internal video clipping tool (<http://floorclips.house.gov>). The last major piece of the Clerk website redesign is a new site search tool, which is expected to be ready in June 2019.

Lobbying Disclosure - Justice Against Corruption on K Street Act of 2018

The Clerk's Office is implementing the Justice Against Corruption on K Street Act of 2018 passed on December 20, 2018 and signed into law by the President. The "JACK" Act became effective on January 3, 2019, and amends the LDA by requiring all registrations (form LD-1) and quarterly reports (form LD-2) filed on or after January 3, 2019, to include "for any listed lobbyist who was convicted in a Federal or State court of an offense involving bribery, extortion, embezzlement, an illegal kickback, tax evasion, fraud, a conflict of interest, making a false statement, perjury, or money laundering, the date of the conviction and a description of the offense."

The Clerk and the Secretary of the Senate notified registrants in January that our offices were developing the technical changes necessary on the two relevant forms to meet this reporting requirement. We made those changes and posted additional guidance, on March 29, 2019, and notified filers that the updates are available in the system. If the statutory timelines required them to file a registration or quarterly report before the amended forms were available, we advised them to do so, and, if necessary, amend the filing to include the relevant information when it became possible to do so. Below are the project deliverables and estimated timeframes:

- February 8, 2019: Define additional fields, schema, and modified XML (complete)
- March 8, 2019: Completed updates to LD-1/LD-2 and data import modules
- March 31, 2019: Completed updated forms moved to production
- May 2019: Completing a new search/download site system is an item that remains on our priority list but has been pushed back because of the addition of unforeseen projects.

Biographical Directory of the U.S. Congress

We have continued to redevelop the *Biographical Directory of the U.S. Congress* (BioGuide) website (<http://bioguide.congress.gov>). Upgrades include a robust search engine, responsive user interfaces, and a content management system that will enable House and Senate Historians to enter more extensive profile data and collaborate seamlessly. The Clerk's Office expects to launch the Alpha version of the website during the first half of 2019.

Legislative Information Management System Modernization and Redevelopment

Legislative Information Management System (LIMS) modernization and redevelopment project. (LIMS) is a critical part of the flow of legislative information. It brings in bill information, Floor activity, Member and Committee information, and Executive Communications from the House and Senate and distributes that information to the Government Publishing Office (GPO), the Library of Congress, Members, Committees, House Officers, and the public. However, the LIMS system was built more than 30 years ago and operates with an outdated programming language on a legacy platform from the mainframe era. The platform has undergone multiple migrations and currently runs in the Linux environment. However, the legacy codebase has not been upgraded and poses significant challenges moving forward. The

cost and technological risk of continuing this legacy platform are very high as finding skilled developers who can work on legacy platforms is difficult and maintenance and changes to the current system demand significant resources.

The migration of LIMS to a new modern technology will allow the application to be more flexible, extensible, robust and secure, and ultimately enable it to better meet the House's needs. This phase of the modernization project will comprise LIMS data modeling, backend development, Floor Action Reporting System (FARS), Committee Action Reporting System (CARS), Executive Action Reporting System (EARS), Reports Due to Congress, Senate Action Reporting System (SARS), and Quality Assurance Testing.

In closing, over the last six plus years the Legislative Branch has made a lot of progress in the areas of data standardization and transparency. That's not to say that there isn't still a lot of work to do but I believe we are moving in the right direction. When the Bulk Data Task Force started more direction had to be given regarding projects to work on, today the dynamic is completely different as all the staff and organizations involved including the civil society groups work together and on their own to help to make legislative information and data more readily available to Members, staff and the American public.

As we move forward I see a move toward accessing data via application program interfaces (APIs) instead of using bulk data files and the Clerk has positioned itself to move in that direction by developing our new website through the use of APIs. There is also a need to look at the whole legislative drafting process from desktop tools through the legislative drafting system to see where we can make improvements to the process.

Thank you again for the opportunity to testify. I would be happy to answer any questions you might have.