



Inter-Parliamentary Union  
For democracy for everyone.



THE NATIONAL ASSEMBLY OF  
THE REPUBLIC OF KOREA

# World e-Parliament Conference 2014

8 - 10 May 2014 // National Assembly of the Republic of Korea // Seoul

# XML in the United States House of Representatives

Kirsten Gullickson  
Office of the Clerk

World e-Parliament Conference 2014

Thank you. It is exciting to be here sharing our lessons learned regarding XML adoption and implementation in the U.S. House of Representatives.

Today, I am going to share a brief history of what the U.S. House has done with XML, share some “lessons learned”, and wrap up with a description of a toolkit that was created as part of a data challenge or contest sponsored by the Library of Congress.

113TH CONGRESS  
2D SESSION

# H. R. 4032

To exempt from Lacey Act Amendments of 1981 certain water transfers by the North Texas Municipal Water District and the Greater Texoma Utility Authority, and for other purposes.

---

IN THE HOUSE OF REPRESENTATIVES

FEBRUARY 11, 2014

Mr. HALL (for himself, Mr. SESSIONS, and Mr. SAM JOHNSON of Texas) introduced the following bill; which was referred to the Committee on Natural Resources, and in addition to the Committee on the Judiciary, for a period to be subsequently determined by the Speaker, in each case for consideration of such provisions as fall within the jurisdiction of the committee concerned

---

## A BILL

To exempt from Lacey Act Amendments of 1981 certain water transfers by the North Texas Municipal Water District and the Greater Texoma Utility Authority, and for other purposes.

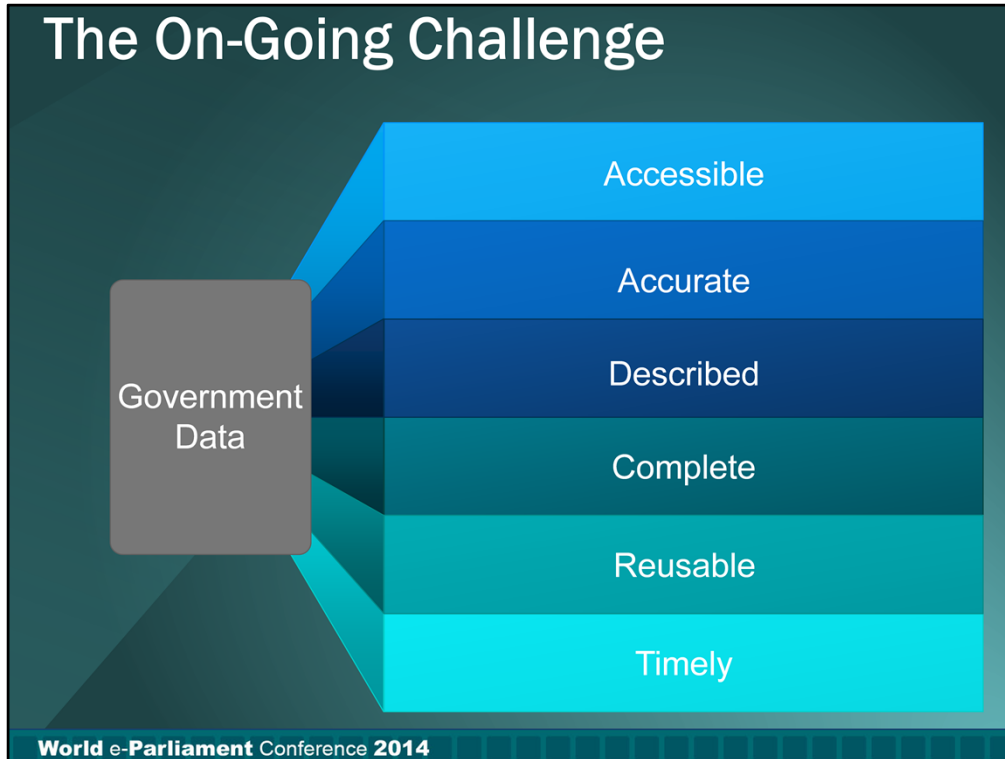
1 *Be it enacted by the Senate and House of Representa-*  
2 *tives of the United States of America in Congress assembled,*  
3 SECTION 1. SHORT TITLE.  
4 This Act may be cited as the "North Texas Invasive

**The paper version is the official document of record.**

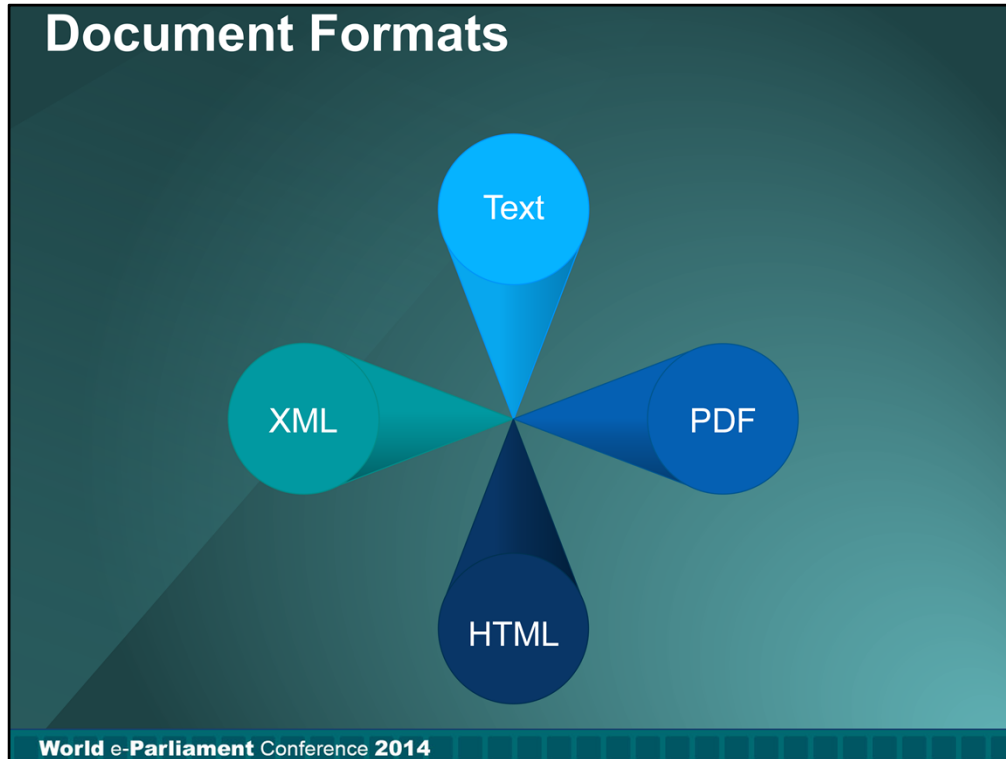
**World e-Parliament Conference 2014**

Although the United States Constitution does not mandate open sessions, the United States House of Representatives opened its doors to the public starting in April of 1789.

Since those first meetings, the House has been preparing, managing, distributing and archiving its official proceedings and documents. Like it was in those beginning days, the physical paper document is still the official document of record.



For all legislative bodies, the on-going challenge with regards to publishing its proceedings, documents and related data, is to ensure that is accessible, accurate, described, complete, reusable, and timely. This includes making proceedings, documents, and related data available to the elected and appointed members of the legislative body, its staff, and the public.



Since its formation, the U.S. Congress has been making its official documents available on paper, but in the late 20th century, Congress began publishing its proceedings and other legislative documents on the internet in electronic formats.

Today, there are still paper copies available, and depending on the type of legislative document, it is publicly available in text, PDF, HTML, and/or XML formats.

*Show an actual Congressional Record and a bill.*



## Lessons Learned: Bills, Resolutions and Amendments in XML

World e-Parliament Conference 2014

I would like to talk about some of the lessons learned as we implemented XML.

For those who are new to XML, I wanted to take a few seconds to describe it.

---

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**.

---

**World e-Parliament** Conference **2014**

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.

# XML Basics

```
<recorded-vote>  
  <legislator name-id="C001101" party="D">  
    Ms. Clark</legislator>  
    <vote>Yea</vote>  
  </recorded-vote>  
  
<recorded-vote>  
  <legislator name-id="B001250" party="R">  
    Mr. Bishop</legislator>  
    <vote>Yea</vote>  
  </recorded-vote>
```

World e-Parliament Conference 2014

An XML element is everything from (including) the element's start tag to (including) the element's end tag.

An attribute is a name-value pair attached to the element's start tag.

*Use the pointer, describe the XML in more detail -- re: the element and attribute*





## Validate the XML with a DTD or Schema

**World e-Parliament** Conference 2014

XML documents need to follow some rules, similar to grammar rules, in order for the XML to be valid. A DTD or Document Type Definition document or a XML schema document defines the rules.

## Lessons Learned: Bills, Resolutions and Amendments in XML

World e-Parliament Conference 2014

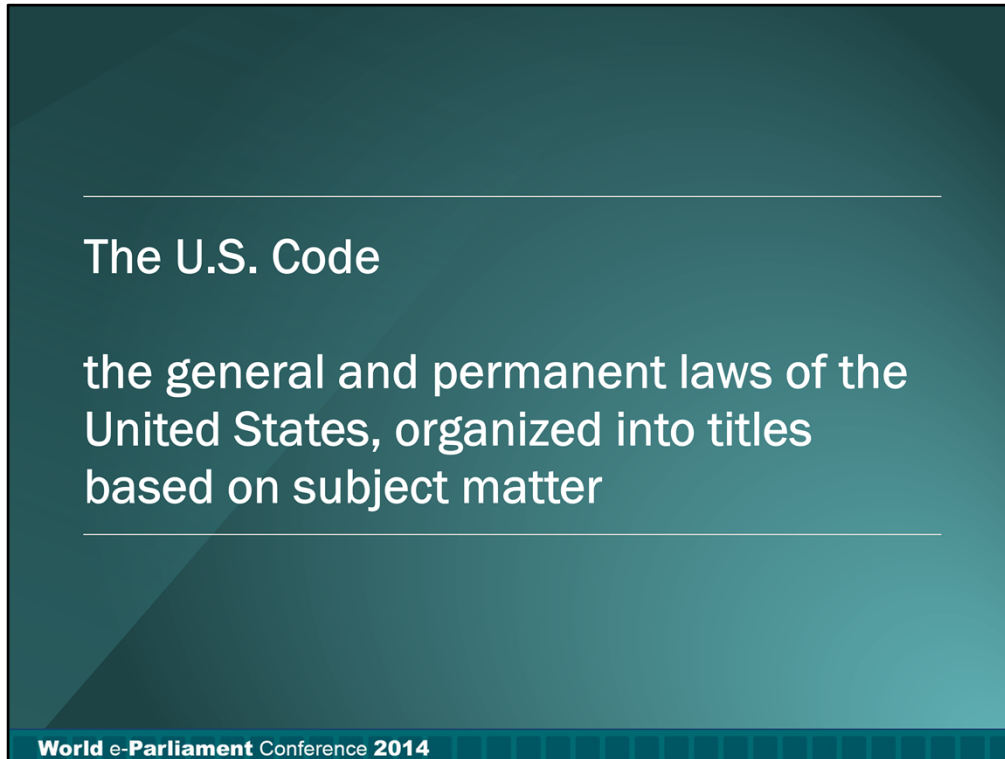
In the late 1990's when the House and Senate agreed to move to XML, a Legislative Branch XML Working Group was established to coordinate and manage Congress' move to XML.

The Working Group is a group of non-partisan support staff from the U.S. House, U.S. Senate, and two Legislative Branch support agencies -- the Government Printing Office and the Library of Congress. We make decisions by consensus. This group agreed on the XML elements and attributes that would be used to mark-up House and Senate bills, resolutions and amendments.

Congress has no central IT governance office, so creating a Working Group to work across the two chambers was really important for moving both chambers from paper to the web.

In the beginning, both the U.S. House and U.S. Senate customized an off-the-shelf XML editor called XMetaL. We selected this software in 1999 because it was affordable, used common development languages, and provides fully featured tags-off editing view. XMetaL is usable desktop application, but it was costly for the initial customizations and its on-going maintenance. It has been in use for the last twelve years and it works.

We do have three separate DTDs for the three document types – bill, resolutions and amendments. We created the DTDs in 1997 when XML tools were in their infancy or were not created yet. We sunk a significant amount of time and money in the DTDs, the XML Editor XMetaL, and other tools. Anyone starting a new effort in XML should design a schema. We do know now that we could have created one DTD for all three document types. We didn't know then what we know now. By all measures the effort to publish bills, resolutions and amendments in XML has been a success.



Last summer, the U.S. House, via the Office of Law Revision Counsel, released the United State Code in XML. The U.S. Code contains the general and permanent laws of the United States, organized into titles based on subject matter. The printed version of the U.S. Code is over 200,000 pages.

To put the U.S. Code in XML, a number of lessons learned and current best practices were implemented.

The U.S. House hired a contractor to lead the project. The contractor has experience implementing XML projects with other legislative bodies. With the expertise of the Congressional staff and the contractor, we had a good team of people working on the task.

To begin, the Legislative Branch XML Working Group agreed that a schema needed to be designed. It would have been silly to direct the contractor to create a DTD given today's toolsets and industry best practices surrounding XML schemas.

Four key "lessons learned" can be seen in the XML schema for the U.S. Code.

# Lesson Learned: Model the data as it appears

1 *Be it enacted by the Senate and House of Representa-*  
2 *tives of the United States of America in Congress assembled,*  
3 **SECTION 1. SHORT TITLE.**

4 This Act may be cited as the “National Park Service  
5 100th Anniversary Commemorative Coin Act”.

6 **SEC. 2. FINDINGS.**

7 The Congress

```
1 <section id="H354CA4636A1B41A693F90E736E818B0A"  
  section-type="section-one"><enum>1.</enum><header>Short  
  Title</header><text display-inline="yes-display-inline">  
  This Act may be cited as the <quote>National Park  
  Service 100th Anniversary Commemorative Coin Act</quote>  
  . </text></section>  
2 <section id="H5181530F2C244C558618C5BD142456CD"><enum>2.  
  </enum><header>Findings</header><text display-inline=  
  "yes-display-inline">The Congress finds the following:  
  </text> </section>
```

World e-Parliament Conference 2014

The first lesson we learned was “model the data as it appears”.

So, in the XML for the U.S. Code, we are doing just that.

To the greatest extent possible, text that is published is maintained in the main body of the document in the order it appears when presented in the publication.

## Not this...

1 *Be it enacted by the Senate and House of Representa-*  
2 *tives of the United States of America in Congress assembled,*  
3  
4 SECTION 1. This Act may be cited as the "National  
5 Park Service 100th Anniversary Commemorative Coin  
6 Act".

7  
8 FINDINGS  
9 SEC. 2. The Congress finds the following:

```
1 <section id="H354CA4636A1B41A693F90E736E818B0A"  
2 section-type="section-one"><enum>1.</enum><header>Short  
3 Title</header><text display-inline="yes-display-inline">  
4 This Act may be cited as the <quote>National Park  
5 Service 100th Anniversary Commemorative Coin Act</quote>  
6 . </text></section>  
7 <section id="H5181530F2C244C558618C5BD142456CD"><enum>2.  
8 </enum><header>Findings</header><text display-inline=  
9 "yes-display-inline">The Congress finds the following:  
10 </text> </section>
```

## Lesson Learned: Avoid generated text

3 SECTION 1. FINDINGS AND DEFINITIONS.  
4 (a) FINDING.—Congress finds it in the public interest  
5 to authorize the sale of certain federally owned land in

```
<section id="H171EBBB26FB14E62A44B0B19946A7C98"  
section-type="section-one"><enum>1.</enum><header>  
Findings and definitions.</header>  
<subsection id="HC3CC357C072B42A0B016AF9F81724346"  
><enum>(a)</enum><header>Finding.—</header><text  
display-inline="yes-display-inline">Congress finds it  
in the public interest to authorize the sale of  
certain federally owned land in
```

World e-Parliament Conference 2014

In general, generated text is avoided in the schema used for the U.S. Code.

Over the past decade, as Congress has gained experience with bill drafting in XML, drafters have found that the use of generated text can be problematic, particularly when working with existing law.

Generated text can be useful when drafting proposed legislation or new law, -- it keeps the drafter from making typographical errors --- but when trying to represent current law, generated text is not always flexible when working with law that may contain typographical errors. In our older laws, we do have places where punctuation is missing.

## Not this...

3 SECTION 1. FINDINGS AND DEFINITIONS.

4 (a) FINDING.—Congress finds it in the public interest

5 to authorize the sale of certain federally owned land in

```
<section id="H171EBBB26FB14E62A44B0B19946A7C98"  
section-type="section-one"><enum>1.</enum><header>  
Findings and definitions</header>  
<subsection id="HC3CC357C072B42A0B016AF9F81724346"  
><enum>(a)</enum><header>Finding</header><text  
display-inline="yes-display-inline">Congress finds it  
in the public interest to authorize the sale of  
certain federally owned land in
```

## Lesson Learned:

No attribute text in online or printed presentation

### SECTION 1. FINDINGS AND DEFINITIONS.

(a) FINDING.—Congress finds it in the public interest to authorize the sale of certain federally owned land in

```
<section id="H171EBBB26FB14E62A44B0B19946A7C98"
section-type="section-one"><enum>1.</enum><header>
Findings and definitions.</header>
<subsection id="HC3CC357C072B42A0B016AF9F81724346"
><enum>(a)</enum><header>Finding.—</header><text
display-inline="yes-display-inline">Congress finds it
in the public interest to authorize the sale of
certain federally owned land in
```

World e-Parliament Conference 2014

XML attributes are reserved for metadata and normalized representations of the element text. No attribute text should ever appear, as is, in the online or printed presentation of the United States Code.



## Not this...

### SECTION 1. FINDINGS AND DEFINITIONS.

(a) FINDING.—Congress finds it in the public interest to authorize the sale of certain federally owned land in

```
<section id="H171EBBB26FB14E62A44B0B19946A7C98"
section-type="section-one" enum="1"><header>Findings
and definitions.</header>
<subsection id="HC3CC357C072B42A0B016AF9F81724346"
enum="(a)"><header>Finding.—</header><text
display-inline="yes-display-inline">Congress finds it
in the public interest to authorize the sale of
certain federally owned land in
```

# Lesson Learned:

## Reuse elements and data models from other standards

Akoma Ntoso

XHTML

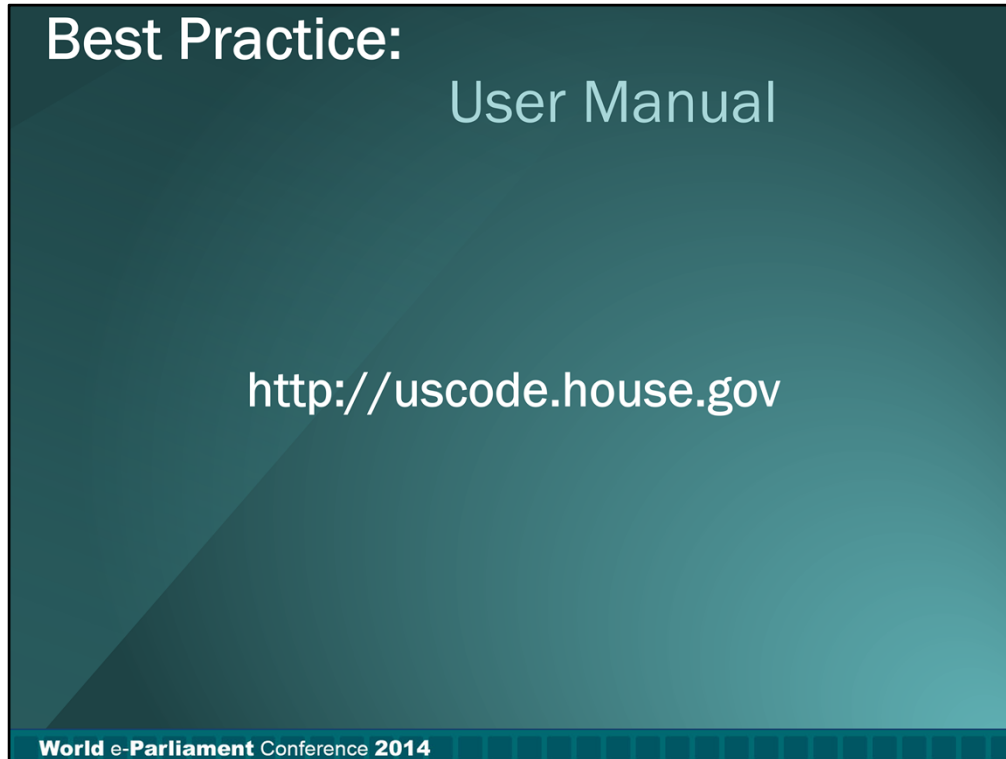
```
<paragraph style="-uslm-lc:I11" class="indent0" id="id1c9b745f-76d8-11e3-b262-e79c7393386e" identifier="/us/uscc/t28/s332/e/10">  
<num value="10">(10)</num><content>  
Preparing an annual report to the circuit and to the Administrative Office of the United States Courts for the preceding calendar year, including recommendations for more expeditious disposition of the business of the circuit.  
<p style="-uslm-lc:I11" class="indent0">  
All duties delegated to the circuit executive shall be subject to the general supervision of the chief judge of the circuit.</p>  
</content>
```

World e-Parliament Conference 2014

Where practical, the equivalent Akoma Ntoso elements are used.

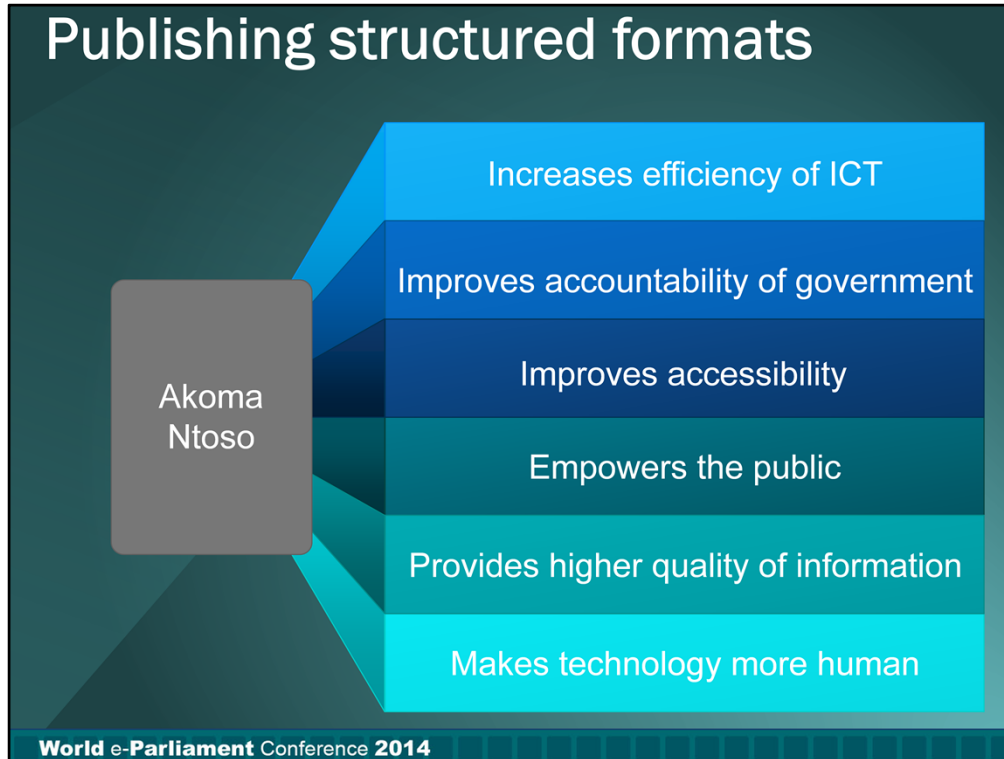
The schema leverages the XML version of HTML 4.0 or XHTML.

Many of the attributes and elements names purposely coincide with their XHTML equivalents.



Following best practices, a comprehensive user guide was created and published. It is a well written guide to the schema and the U.S. Code in XML.

The United States Code and its user guide can be found at this website.



Why an international standard. Why not?

What the staff in the U.S. House learned about XML implementation over the last ten years is really the same that other people were experiencing. Some of our lessons learned are best practices today.

A common standard can help all parliaments. Early XML projects across the public and private sector proved that XML and other structured formats improved operations and saved money. We also learned that legislative documents across the various jurisdictions have more in common than what one might initially realize.

In 2008, when Monica Palmirani (OASIS LegalXML Technical Committee) first described the vision for a common standard, I was skeptical. Laws and legislative documents are complex. Today, I am in full support.

To support the efforts of the OASIS LegalXML Technical Committee and the developments of the standard Akoma Ntoso, the House worked with the Library of Congress to host an open innovation challenge or contest.



Challenge.gov allows federal agencies and organizations to run open innovation or crowdsourcing contests. This government sponsored and managed initiative is similar to private sector contests.

As described on the website, “Challenge.gov is a collection of challenge and prize competitions, all of which are run by more than 50 agencies across the United States federal government. These include technical, scientific, and creative competitions where the U.S. government seeks innovative solutions from the public, bringing the best ideas and talent together to solve mission-centric problems.”

Our challenge had the goal of advancing the development of international exchange standards for legislative data and identifying potential gaps in the Akoma Ntoso schema. Challenge participants were asked to map existing elements from US and UK bill XML to the most recent Akoma Ntoso schema for eight specific pieces of US and UK legislation. These selected examples of legislation employed either the US Legislative XML DTD or the UK’s Crown Legislation Markup Language schema.

113TH CONGRESS  
2d Session

# H. R. 4032

To exempt from Lacey Act Amendments of 1981 certain water North Texas Municipal Water District and the Greater Authority, and for other purposes.

IN THE HOUSE OF REPRESENTATIVES

FEBRUARY 11, 2014

Mr. HALL (for himself, Mr. SESSIONS, and Mr. SAM JOHNSON of Texas) introduced which was referred to the Committee on Natural Resources, and in addition the Judiciary, for a period to be subsequently determined by the Speaker consideration of such provisions as fall within the jurisdiction of the committee

## A BILL

To exempt from Lacey Act Amendments of 1981 certain water North Texas Municipal Water District and the Greater Authority, and for other purposes.

*Be it enacted by the Senate and House of Representative States of America in Congress assembled,*

SECTION 1. SHORT TITLE.

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<akomantoso xmlns="http://docs.oasis-open.org/legaldocml/ns/akn/3.0/CSD06"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://docs.oasis-open.org/legaldocml/ns/akn/3.0/CSD06
http://akomantoso.googlecode.com/svn/release/trunk/schema/akomantoso30.xsd"
113 HR 4032 IH: North Texas Invasive Species Barrier Act of 2014>
<docIntroduction refersTo="#H000067">Mr. Hall</docIntroduction>
<docProponent refersTo="#S000250">Mr. Sessions</docProponent>
<docProponent refersTo="#J000174">Mr. Sam Johnson of Texas</docProponent>
<docCommittee refersTo="#H1100">Committee on Natural Resources</docCommittee>
<docCommittee refersTo="#HJU00">Judiciary</docCommittee>
<content currentId="sec1">
<num>Section. 1.</num>
<heading currentId="sec1-heading">Short title</heading>
<content currentId="sec1-content">
<p>
This Act may be cited as the
<extractText startQuote="" endQuote="">North Texas Invasive Species Barrier
Act of 2014</extractText>
</p>
</content>
</section>
<section currentId="sec2">
<num>Sec. 2.</num>
<heading currentId="sec2-heading">Compliance with Lacey Act Amendment of
1981</heading>
```

World e-Parliament Conference 2014

The judges were very excited to award prizes to two winners.

The first prize winner built a browser plug-in for Google Chrome.

The tools Mr. Mangiafico developed (including a Chrome extension) were so refined one judge remarked, "It is a starting point for something and not just a proof of concept." His mapping eked out only a very narrow win because of his mastery of temporal metadata and his ability to maintain the end of line (<eol/>) designations in UK document mapping.

In addition, both submitters exhibited a good understanding of the complexity of US and UK legislation by maintaining needed elements in both sets of documents.

A solution like the Google Chrome plug-in provides:

- a solution that allows us to provide U.S. Legislation in AK without have to re-engineer our existing toolsets and workflows.
- On-demand availability.

---

## Data Challenge

<http://legislative-data-mapping.challengepost.com/>

---

---

## Basic Legislative Workflow

<http://xml.house.gov/wep2014>

---



# Finding Legislative Information United States House

DOCS.HOUSE.GOV

*Text of legislation being considered now*

HOUSE.GOV

*House information*

CONGRESS.GOV

*House and Senate data and documents*

GPO.GOV

*all government documents*

Thank you  
Merci  
Gracias

**Kirsten Gullickson**

Office of the Clerk  
U.S. House of Representatives

[Kirsten.Gullickson@mail.house.gov](mailto:Kirsten.Gullickson@mail.house.gov)

**World e-Parliament** Conference 2014