

Cato Institute

Federal Legislative XML Guide

Version 1.0
January, 2013

This is a guide to XML, how it is currently being used by the United States Congress, and how the Cato Institute plans to improve legislative transparency with enhanced tagging and improved data relationships.

Table of Contents

XML Overview	2
House XML	5
Structural Elements	5
Procedural Elements	7
Organizational Elements	10
Display/Authoring Elements	16
Nominal/Descriptive Elements	19
Cato XML	22
Cato Namespace	22
The <entity-ref>, <entity>, and <property> Elements	23
Federal Organizational Units	23
Committees and Subcommittees	23
Federal Elective Officeholders	24
Existing Law and Parallel Law Citations	24
Proposed Laws and Citations	26
New Authority	27
Budget Authority Properties	28
Lookup Tables	29
Appendix	32

Font variations in this document:

Font	Meaning
12 Point Font, Bolded Arial	Main Section Headers
11 Point Font, Bolded Arial	Element Category Header
9 Point Font, Bold Arial	XML Terms and Concepts
<Courier New in Brackets>	XML Elements
Courier New not in Brackets	XML Attributes
"Quoted Courier New"	XML Attribute Values
Courier New in Gray	XML Example
<i>9 Point, Italicized Arial Text</i>	Printed/Parsed Text from XML

XML Overview

XML (short for eXtensible Markup Language) is a protocol used to define and transport data. Similar to HTML, the markup language underlying the World Wide Web, XML structures and contains data, making it more useful than unstructured information. For novices, right-clicking on a Web page and selecting "View Source" will reveal the HTML version of the page—the simpler the page the better—and provide an idea of what XML looks like.

Where HTML is a defined standard, XML does not have any predefined or standardized tags. All elements in XML are defined and structured by the author in a separate document – either a DTD (Document Type Definition) or XML schema file. When the elements are parsed by an external software program, the information can be put to whatever use that program facilitates. Because XML documents can be processed by computers, they are considered "machine readable."

A very basic XML document may look something like this:

```
<?xml version="1.0" encoding="UTF-8"?>
<presidents>
  <name name-id="W000178">
    <firstname>George</firstname>
    <lastname>Washington</lastname>
  </name>
  <name name-id="A000038">
    <firstname>John</firstname>
    <lastname>Adams</firstnme>
  </name>
</presidents>
```

In this document:

`<?xml version="1.0" encoding="UTF-8"?>` is an **XML Declaration**. Every XML document must start with a declaration specifying that it is XML, as well as the version number. Declaring the encoding can also be important because it denotes the way characters – numbers letters, and symbols – are encoded in the document.

`<presidents>` is the **root element**. All XML documents must have a root element as a container for all other tags. The root element will generally use a name that describes the overall content of the document.

Parent elements and **child elements** add information in a way that makes clear the relationships among the things they describe. The root element is the top-most parent element, so `<presidents>` is the parent element of its child element, `<name>`. `<name>` in turn is the parent of child elements `<firstname>` and `<lastname>`. An element is closed by a tag repeating its name preceded by a forward slash, e.g. `</firstname>` or `</lastname>`. All child elements must be closed before their parent element closes. This is called **nesting**.

`name-id` is an **attribute**. Attributes provide additional information about elements. So, an attribute of `<name>` is its distinct identifier, `name-id`. Elements can contain multiple attributes, and different elements can share names and values, but each attribute can have only one value per element. So while `<name name-id="A000038" fullname="John Adams">` would be a valid use of attributes, `<name name-id="A000038" fullname="John Adams" name-id="R000027">` would not be valid.

When XML is used to tag elements in a text, it makes the semantics of the text easier to discover using a computer. The XML will "surround" the relevant piece of text. For example:

```
When <name name-id="A000038">President <lastname>Adams</lastname></name> lost
his reelection bid, he retired to Massachusetts.
```

Because the `name-id` value "A000038" is a unique identifier assigned to him, this tagging makes clear to a computer that the president referred to here is President John Adams, the second president of the United States, and not his son John Quincy Adams, the sixth president of the United States.

Other important XML terms and concepts include the following:

Namespaces

Namespaces avoid naming conflicts among elements. Because XML is not a predefined standard, two different versions of XML could use the same element names differently. The element `<title>`, for example, is already a standardized HTML element, but it is also frequently used in other XML schemata. So if one were to use HTML and XML in the same document, adding a namespace to the XML element would avoid confusion among elements and attributes that might have the same name in the two different vocabularies. Namespaces can also be used to define XML tags standardized by a specific group or organization.

There are multiple ways namespaces can be declared. Here are a few:

Format 1:

```
<[parentElement] xmlns:[abbrev]="[name]">
```

For example:

```
<cato:presidents xmlns:cato="http://namespaces.cato.org/catoxml">
  <cato:name name-id="A000038">
    <firstname>John</firstname>
    <lastname>Adams</lastname>
  </cato:name>
</cato:presidents>
```

In this format, the namespace is defined under the `parentElement` and is only used to prefix some of the elements. The attribute `xmlns` ("xml name space") defines the namespace name `"http://namespaces.cato.org/catoxml"` as well as the prefix `cato:`. It is common, though not required, for the namespace name to be a URL that points to the namespace's definitions. In this example, the `cato:` prefix is not used under the `<firstname>` and `<lastname>` elements. This indicates that these elements were not created by the originator of the `cato:` namespace.

Format 2:

```
<[parentElement] xmlns:[abbrev]="[name]">
```

For example:

```
<presidents xmlns:cato="http://namespaces.cato.org/catoxml">
  <cato:name name-id="A000038">
    <cato:firstname>John</cato:firstname>
    <cato:lastname>Adams</cato:lastname>
  </cato:name>
</presidents>
```

This format is similar to the first except that the `cato:` prefix is not used under the `parentElement` and all child elements adopt the namespace properties. This indicates that the `parentElement` `<presidents>` was not created by the `cato:` namespace originator. Namespace declarations, however, can still be made in the `parentElement`.

Format 3:

```
<[parentElement] xmlns="[name]">
```

For example:

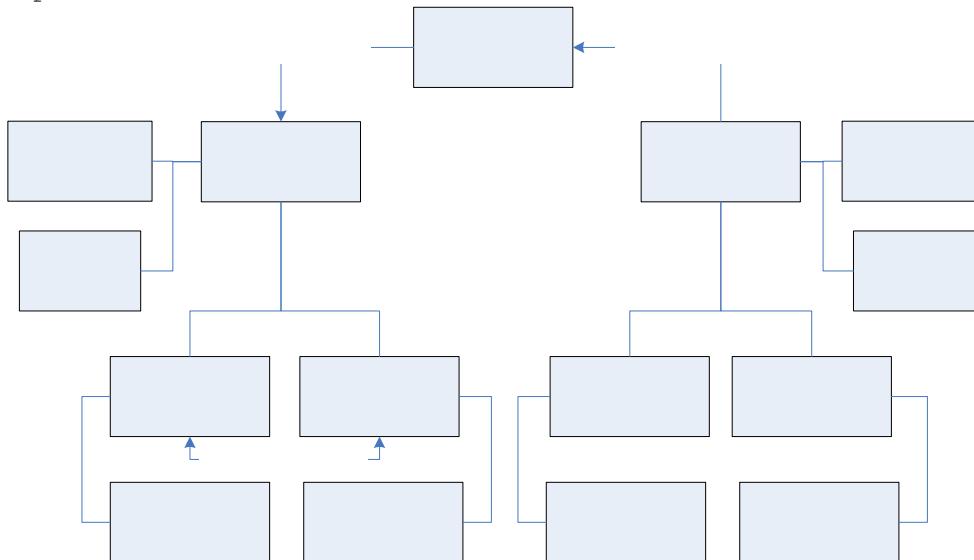
```
<presidents xmlns="http://namespaces.cato.org/catoxml">
  <name name-id="A000038">
    <firstname xmlns="">John</firstname>
    <lastname xmlns="">Adams</lastname>
  </name>
</presidents>
```

In this example, the "http://namespaces.cato.org/catoxml" namespace is defined under the parentElement <presidents> and all following child elements are assumed to inherit that namespace. If a child element does not share the namespace properties of its parent element, a new namespace declaration must be made to change or nullify those properties. In this example, <firstname> and <lastname> define an empty namespace xmlns="" to nullify the namespace properties of its parentElement. The prefix cato: is bound where ever the "http://namespaces.cato.org/catoxml" namespace is applicable. Cato XML most often uses Format 2, where the namespace is defined in the root element and prefixes denote the Cato elements, but any namespace declaration format will be recognized.

DOM and Nodes

The standard that defines all pieces of an XML document is called the **DOM**, or the **Document Object Model**. The DOM is the unseen interface that breaks down the pieces of any XML document into a series of **nodes**. An entire document is a document node, elements are element nodes, the root element is the root node, strings of text surrounded by elements are text nodes, and attributes are attribute nodes. The DOM breaks down all nodes into a node-tree which explains the hierarchical nature of the document. The following is a sample XML document followed by its DOM:

```
<?xml version="1.0" encoding="UTF-8"?>
<presidents xmlns:cato="http://namespaces.cato.org/catoxml">
  <cato:name name-id="W000178">
    <firstname>George</firstname>
    <lastname>Washington</lastname>
  </cato:name>
  <cato:name name-id="A000038">
    <firstname>John</firstname>
    <lastname>Adams</firstnme>
  </cato:name>
</presidents>
```



For more information on XML DOM, please read <http://www.w3schools.com/dom/default.asp>.

Metadata

Metadata is data about data, or additional information about the document's content that does not appear in the data itself – information like the language the document uses, the copyright status of the document, and so on. In XML, it is very simple to incorporate metadata into documents to make document management easier. Metadata is generally included at the top of the page before the body of text. Elements may differ among documents, but the metadata section typically looks something like this:

```
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
```

```

<dublinCore>
<dc:title>[title of the document]</dc:title>
<dc:format>text/xml</dc:format>
<dc:language>EN</dc:language>
<dc:rights>Pursuant to Title 17 Section 105 of the United States Code, this
file is not subject to copyright protection and is in the public
domain.</dc:rights>
</dublinCore>
</metadata>

```

dc: is the namespace abbreviation for Dublin Core, which is a project intended to provide standardized metadata vocabularies for a wide variety of XML projects. This makes for easier interpretation by web crawlers and other programs that interact with XML.

Metadata elements are elements that define metadata for a given text node. Attributes frequently add metadata to a span of text. House XML and Cato XML define multiple metadata elements, which are discussed below.

House XML

House XML is our name for the set of standardized XML elements used to define the most commonly used structures, sections, references, and terms in legislative documents that Congress publishes. The definitions were first created in 1998 and were most recently updated in 2004. A list of all standardized elements and attributes can be found in the bill.dtd file at xml.house.gov. As of now, a large majority of the elements defined by the House are unused. The elements that are used mostly define the structure of the document, but some identify meaningful content.

We have divided our catalogue of House XML elements by their major purposes, and we categorize them in five ways:

- Structural elements, which properly format the XML.
- Procedural elements, which define information about the procedural status of the document.
- Organizational elements, which define specific levels and types of text found in the document.
- Display elements, which define how text should be printed.
- Nominal/Descriptive elements, which identify terms, documents, and other meaningful things referred to in legislative documents.

The elements found in House XML are discussed in this order below.

Structural Elements

Many of the in-use House XML elements are purely structural, meaning that they do not define specific content, but rather help properly format the XML. Structural elements include:

Root Elements

<bill> Root element for all XML formatted bills. All other elements are hierarchically under this element. This element defines the overall type and structure of the bill. For example:

```

<bill bill-stage="Introduced-in-House" bill-type="olc" dms-
id="HA151391BBB5E4F9CAB0D7249E9F93A85" key="H" public-
private="public">

```

Attributes used vary, but they can include:

bill-stage Values in use include "Introduced-in-House", "Engrossed-in-House", "Placed-on-Calendar-Senate", "Received-in-Senate", "Reported-in-House", and so on.

bill-type Denotes the style of the bill. This generally affects header and indentation style in the PDF and print-out versions of the bill. Values include "olc" (Office of the Legislative Counsel, the default style that a vast majority of bills use),

"traditional" (used for some very short bills), and "appropriations" (used for some appropriations legislation).

dms-id This defines a document management system URI that can be used for references to the bill itself. The value of `dms-id` is usually prefixed with "H" (but occasionally it is prefixed with "ID") and followed by 32 random characters, although a few bills have a `dms-id` value of simply "A1".

key This is an infrequently used attribute that identifies the organization that created the present version of the bill. Valid values are "H" (House), "S" (Senate), "G" (GPO), and "L" (Library of Congress).

public-private Declares whether the bill is "public" (pertaining to the general population) or "private" (pertaining to a specific individual or group of individuals). A vast majority of legislation is public, but occasionally private legislation is proposed, often to provide compensation to, or change the immigration status of, specific individuals.

<resolution> Root element for all XML formatted resolutions. Like `<bill>` all other elements are hierarchically under this element and its attributes define the overall type and structure of the resolution. For example:

```
<resolution dms-id="HAC1645E692F249FCB7AB1E893F87F7E6" key="H" public-private="public" resolution-stage="Introduced-in-House" resolution-type="house-concurrent" star-print="no-star-print">
```

Attributes include:

dms-id This defines a document management system URI that can be used for references to the resolution itself. The URI is prefixed with "H" followed by 32 random characters.

key This is an infrequently used attribute that identifies the organization that created the present version of the bill. Valid values are "H" (House), "S" (Senate), "G" (GPO), and "L" (Library of Congress).

public-private Declares whether the resolution is "public" (pertaining to the general population) or "private" (pertaining to a specific individual or group of individuals).

resolution-stage States the resolution's stage in the legislative process. In-use values include "Introduced-in-House", "Reported-in-House", "Engrossed-in-House", "Placed-on-Calendar-Senate", "Enrolled-Bill", "Received-in-Senate", and "Referred-in-Senate".

resolution-type Declares what type of resolution the document is. In-use values include "house-concurrent", "house-resolution", "house-joint", "senate-concurrent", "senate-resolution", "senate-joint" and "constitutional-amendment".

star-print States whether the resolution is a reprint or not. "first-star-print" is used to mean it's the first reprint, while "no-star-print" means that it is not a reprint.

<amendment-doc> Root element for all XML formatted amendments. All other elements are hierarchically under this element and its attributes define the overall type and structure of the amendment. For example:

```
<amendment-doc amend-stage="proposed" amend-type="house-amendment" amend-degree="first">
```

Attributes in-use include:

amend-stage	Identifies the stage of the amendment. Values include "approved", "pending", "proposed", and "submitted".
amend-type	Identifies the type of amendment. Values in-use include "engrossed-amendment" and "house-amendment".
amend-degree	Identifies the degree of the amendment. Valid values are "first" for first degree amendments (amendments to the measure under consideration) and "second" for second degree amendments (amendments to the first degree amendment while it is pending).

Text Container Elements

Following the root element, there are typically a number of procedural elements (discussed below), followed by text container elements that surround the text of the measure.

`<legis-body>` Parent element to all legislative content in bills. Uses the `id` and `style` attributes. For example:

```
<legis-body id="HCFE52D9633C14CCE8DC1E2C5DCEADF4E" style="OLC">
```

`<resolution-body>` Parent element to all content in resolutions. Uses the `id` and `style` attributes. For example:

```
<resolution-body id="H1EF2D386C92045C18EA2E47A2C81EEDC" style="traditional">
```

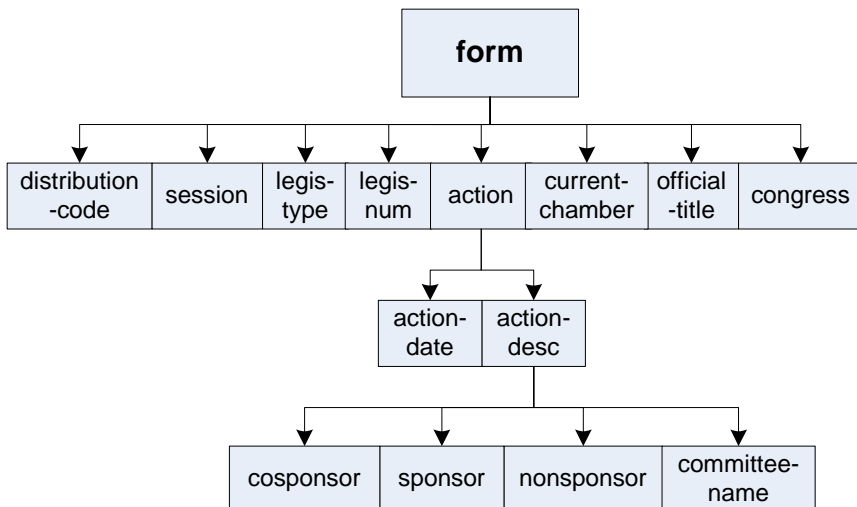
`<amendment-body>` Parent element to all content in amendments.

Procedural Elements

Many of the elements in House XML are used to define procedural content in legislative documents. These elements primarily specify standardized actions in the legislative process. These can appear at the beginning of a document under the `<form>` element, or at the very end of the document under the `<attestation>` element.

Opening Procedural Elements

```
<form>
  <distribution-code display="yes">IV</distribution-code>
  <congress display="yes">112th CONGRESS</congress>
  <session display="yes">2d Session</session>
  <legis-num>H. CON. RES. 105</legis-num>
  <current-chamber>IN THE HOUSE OF REPRESENTATIVES</current-chamber>
  <action display="yes">
    <action-date date="20120228">February 28, 2012</action-date>
    <action-desc>
      <sponsor name-id="H001050">Ms. Hanabusa</sponsor>(for herself and
      <cosponsor name-id="H001042">Ms. Hirono</cosponsor>) submitted the
      following concurrent resolution; which was referred to the <committee-
      name committee-id="HHA00">Committee on House Administration</committee-
      name>
    </action-desc>
  </action>
  <legis-type>CONCURRENT RESOLUTION</legis-type>
  <official-title display="yes">Authorizing the use of Emancipation Hall in the
  Capitol Visitor Center for an event to celebrate the birthday of King
  Kamehameha.</official-title>
</form>
```

- `<form>` Container element for all procedural elements at the beginning of documents.
- `<distribution-code>` Contains the Government Printing Office code used to identify the stage or type of legislation. In-use distribution codes include "I" (Public Bills introduced in the House), "IA" (Joint Resolutions introduced in the House), "IB" (Public Bills and Joint Resolutions reported in the House), "II" (Public Bills and Joint Resolutions reported in the Senate), "IV" (Simple and Concurrent House Resolutions), and "V" (Private Bills introduced in the House). The `display` attribute denotes whether the content of the element will be visible.
- `<congress>` Identifies the two-year meeting of Congress in which the document was introduced (e.g. "112th Congress")
- `<session>` Identifies the one-year session of the Congress in which the document was introduced. Text will either be "1st Session" or "2d Session." This element can include the attribute `display`, which signifies whether the text will be printed or not, the values of which can be either "yes" or "no".
- `<legis-num>` An element that tags the bill number. It will be formatted "H. R. [number]" for House bills, "H. RES. [number]" for House resolutions, "S. [number]" for Senate bills, "S. RES. [number]" for Senate concurrent resolutions, "H. CON. RES. [number]" for House concurrent resolutions, "S. CON. RES [number]" for Senate concurrent resolutions, "H. J. RES. [number]" for House joint resolutions, or "S. J. RES. [number]" for Senate joint resolutions.
- `<current-chamber>` Identifies the chamber in which the version of the bill is being considered. This element will either contain "IN THE HOUSE OF REPRESENTATIVES" or "IN THE SENATE OF THE UNITED STATES"
- `<action>` contains information about what has happened with the bill, including the date of introduction (`<action-date>`) and a description of the action (`<action-desc>`), including sponsors (`<sponsor>`), cosponsors (`<cosponsor>`), committee referrals (`<committee-name>`), and other relevant information. Elements under the `<action>` element include:
- `<action-desc>` Container element for all action elements.
- `<sponsor>`,
`<cosponsor>`,
and
`<nonsponsor>` Used to identify office holders having something to do with the document. For example:
`<cosponsor name-id="B000755">Mr. Brady of Texas</cosponsor>`

Each `<sponsor>`, `<cosponsor>`, and `<nonsponsor>` element contains the

attribute `name-id`. The value of this attribute is the elected official's identifier from the Congressional bioguide found at <http://bioguide.congress.gov/biosearch/biosearch.asp>. This is a standardized system for assigning every federal elective official – present and past – with a unique identifier.

`<committee-name>` Identifies the Senate and House committees to which the bill has been referred. For example:

```
<committee-name committee-id="HED00">Committee on  
Education and the Workforce</committee-name>
```

The `committee-id` attribute references the committee codes established by the House and Senate. The House committee identifiers are officially documented at http://cha.house.gov/sites/republicans.cha.house.gov/files/documents/committee_docs/CommitteeRepository-NamingConventions-v1-2-1.pdf and Senate committee identifiers can be found at http://www.senate.gov/legislative/LIS_MEMBER/cvc_member_data.xml. While inconsistencies exist between codes for the joint committees, the codes defined by the Senate are the more frequently used identifiers. Committee identifiers can be found in Appendix 1.

`<legis-type>` Element that defines the kind of legislation. Types include *CONCURRENT RESOLUTION*, *JOINT RESOLUTION*, *RESOLUTION*, *A BILL*, *AN ACT*, and *Amendment*.

`<official-title>` Identifies the official, long-form title of the given piece of legislation (such as "*Establishing the Congressional Commission on the European Union, and for other purposes.*").

`<short-title>` Identifies the shortened title for a given piece of legislation (such as "*Congressional Commission on the European Union Resolution*").

`<endorsement>` An infrequently used element used to contain the details of an endorsement. Endorsements are procedural details printed on the back of hardcopy legislation and are added by the GPO or clerks. This element uses the `<action-date>` and `<action-description>` elements to tag details of the endorsement. For example:

```
<endorsement display="yes">  
  <action-date date="20110531">May 31, 2011</action-date>  
  <action-desc>Committed to the Committee of the Whole House on the  
  State of the Union and ordered to be printed</action-desc>  
</endorsement>
```

When used, endorsements are included at the end of the document, after the closing `</legis-body>` tag and before the closing root element (either `</bill>` or `</resolution>`). In print, this information is found at the back of the bill or resolution and is added by either the GPO or the House Enrolling Clerks.

Closing Procedural Elements

`<attestation>` and `<attestation-group>` Infrequently used container elements containing the details of the person who certifies the occurrence of events in the legislative process. For example:

```
<attestation>  
  <attestation-group>  
    <attestation-date date="20120605" chamber="House">Passed the House  
of Representatives June 5, 2012.</attestation-date>  
    <attestor display="no">Karen L. Haas,</attestor>  
    <role>Clerk.</role>  
  </attestation-group>  
</attestation>
```

Additional attestation elements are:

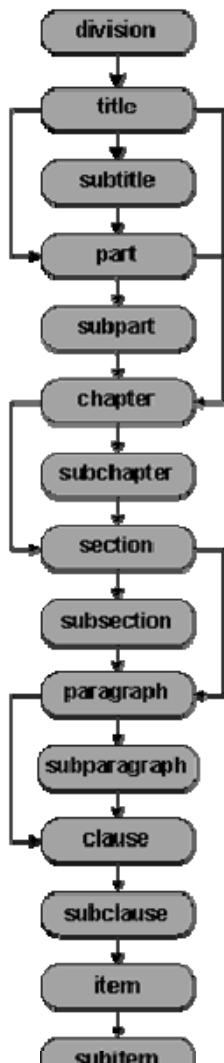
- <attestation-date> Details the date and chamber in which the event occurred. Attributes used include `date` (the value of which is formatted YYYYMMDD, so May 20, 2011, for example, would be "20110520") and `chamber` (which will be either "House" or "Senate").
- <attestor> The name of the individual who certified the occurrence.
- <role> The job title of the attestor. This is generally either "Clerk" or "Secretary".

Organizational

Federal legislation is highly structured. The basic unit of legislation is the section. Sections can be nested under seven higher levels (division, title, subtitle, part, subpart, chapter, and subchapter). Sections can contain seven levels of hierarchy nested under them (subsection, paragraph, subparagraph, clause, subclause, item, and subitem).

Text Hierarchy Elements

Many of the organizational elements define blocks and strings of text in this hierarchy. In print, these are signaled by indentations, enumerators, and text styling. There are some exceptions and levels can be skipped, but 99% of the time the textual hierarchy is as follows:



<division> This is the topmost level of text hierarchy. It is a container element for specific sections and defining legislative structure and uses the `id` and `style` attributes and is typically enumerated with an uppercase letter. For example:

```
<division id="H9D9752474A6F4BF0BB2A49FD7DC69E1E" style="OLC">
  <enum>A</enum>
  ...
</division>
```

<title> Defines titles and uses the `id`, `style`, `commented`, and `level-type` attributes for reference. The purpose of `commented` is not entirely clear, but where used the value is always "no", while `level-type` signifies how the subsequent enumerator will be designated in the printed version of the bill (the only in-use value is "subsequent" which is printed as "Sec."). For example:

```
<title commented="no" id="H7E0FDE1875CE4878BF1AB93FAFA8FE4B" level-type="subsequent">
  <enum>II</enum>
  <header display-inline="yes-display-inline">Leasing program for land within Coastal Plain</header>
  ...
</title>
```

This text would print as the following in the PDF or hardcopy version of the bill:

Sec. II Leasing Program for land within Coastal Plain

<subtitle> Similar rules to <title> element but defines subtitles and only uses the `id` attribute. Typically enumerated with uppercase letters. For example:

```
<title id="H428F3E42FF3A4B1593C50F0B787679A3">
```

```

        <enum>IV</enum><header>Drilling</header>
        <subtitle
id="HE06E5C589EA44F2BAFE8D662B717B0C2">
        <enum>A</enum><header>Tax
provisions</header>
        ...
        </subtitle>
</title>

```

<part> The next lower unit in the hierarchy. Uses the `id` attribute and is generally enumerated with uppercase roman numerals (but frequently also uses numbers and uppercase letters) For example:

```

<part id="H118F7D3D580B42E6AF71BEC9BDF766E8">
  <enum>VIII</enum>
  ...
</part>

```

<subpart> Similar to `<part>` but hierarchically under the `<part>` tag. Typically enumerated with uppercase roman numerals, numbers, and uppercase letters. For example:

```

<subpart id="HAB705E989FC24C29801DE3EA5BC1DCF2">
  <enum>B</enum>
  ...
</subpart>

```

<chapter> Used to define chapters which typically contain a series of sections, subsections, paragraphs, and subparagraphs. Contains attributes for `id`, commented ("yes" or "no" value, the meanings of which are unclear), and `level-type` ("undesignated" or "subsequent", the meanings of which are also unclear). Typically enumerated with numbers. For example:

```

<chapter id="H22AD1F7C3D4F4AFB904538C5479D5238">
  <enum>37</enum>
  ...
</chapter>

```

<subchapter> Similar rules as `<chapter>` but used for subchapters and only uses the `id` attribute. Typically enumerated with uppercase roman numerals and uppercase letters. For example:

```

<subchapter id="H91A9D2EE586B41AB8E38A53834D8C52E">
  <enum>V</enum>
  ...
</subchapter>

```

<section> The basic unit of legislative text. Attributes used include `display-inline` ("yes-display-inline" or "no-display-inline"), `id`, and `section-type` ("undesignated-section", "section-one", "subsequent-section", or "resolved"). The attribute `section-type` signifies how the section will be designated in the printed version of the bill ("section-one" appears as *Section*, the default "subsequent-section" appears as *Sec.*, "resolved" appears as *Resolved*, and "undesignated-section" appears as nothing). Typically enumerated with a number followed by a period. For example:

```

<section id="H38CC28DAE4414969ABDC1D4F8EEC31AC" section-type="section-
one"><enum>1.</enum>
<header>Short title</header>
<text display-inline="no-display-inline">This Act may be cited as
the<quote><short-title>Small Company Capital Formation Act of
2011</short-title></quote>.</text>
</section>

```

This text would print as the following in the PDF or hardcopy version of the bill:

SECTION 1. SHORT TITLE.

This Act may be cited as the "Small Company Capital Formation Act of 2011".

- <subsection>** Contains similar content beneath **<section>** in the hierarchy and only uses an **id** reference. Typically enumerated with a lowercase letter in parentheses. For example:
- ```
<subsection
id="HE3CFEDF5D3B8489BB073D7A047C388A0"><enum>(a)</enum><header>In
general</header><text display-inline="yes-display-inline">The Greater
Washington Soap Box Derby Association (in this resolution referred to
as the <quote>sponsor</quote>) shall be permitted to sponsor a public
event, soap box derby races (in this resolution referred to as the
<quote>event</quote>), on the Capitol Grounds.</text></subsection>
```
- <paragraph>** Specifies the text level as a paragraph, and uses the **id** attribute for reference. Typically enumerated with a number in parentheses. For example:
- ```
<paragraph id="HA518779C57EB4C1E958FC43F9C9E11EC">
  <enum>(3)</enum>
  <text>a reformed and simplified Tax Code should include properly
structured tax incentives to maintain and contribute to such plans and
to strengthen retirement security for all Americans.</text>
</paragraph>
```
- <subparagraph>** Similar to **<paragraph>** but only situated within paragraphs. Typically enumerated with an uppercase letter in parentheses. For example:
- ```
<subparagraph id="H79FE3C73C9454DDDB45F22D5B875848F">
 <enum>(A)</enum>
 <text>requests that the Government of China honor its obligations
under the United Nations Convention relating to the 1951 Status of
Refugees and the 1967 Protocol; and</text>
</subparagraph>
```
- <clause>** Used to tag clauses. Contains an **id** attribute for unique identification. Typically enumerated with lowercase roman numerals. For example:
- ```
<clause id="H2DCEC9A094B044DE96406CAD77689D26">
  <enum>(ii)</enum>
  <text>an urgent, pressing, and compelling need requiring immediate
action;</text>
</clause>
```
- <subclause>** Same rules as **<clause>** but used for subclauses. Typically enumerated with uppercase roman numerals. For example:
- ```
<subclause id="HC96B5C2C02604FEBA3A91C1D1606C22E">
 <enum>(II)</enum>
 <text>subparagraph (A) shall not apply in determining such
individual's primary insurance amount.</text>
</subclause>
```
- <item>** Structure beneath subclauses. Uses the **id** attribute and is typically enumerated with two lowercase letters. For example:
- ```
<item id="idEF068C24C5F445A9A8D5B962266F1F4F">
<enum>(aa)</enum>
<text>race;</text>
</item>
```
- <subitem>** Same rules as **<item>** but used for subitems. Typically enumerated with two uppercase letters. For example:

```

<subitem id="H5DD4A19CF28B4220A099D881401C621A">
  <enum>(AA)</enum>
  <text>for violations arising from a single inspection or examination
shall not exceed $2,500; and</text>
</subitem>

```

Table of Contents Elements

The Table of Contents section in a legislative document is defined using a series of section-specific elements. For example:

```

<toc regeneration="no-regeneration">
<toc-entry bold="on" level="section">Chapter</toc-entry>
<multi-column-toc-entry bold="on" level="chapter">
<toc-enum>1.</toc-enum>
<level-header level="chapter">Definitions</level-header>
<target>1</target>
</multi-column-toc-entry>

```

Elements specific to the table of contents include:

<code><toc></code>	Used to define elements in the Table of Contents. Attributes include:
container-level	Defines the element the table of contents is used for. In-use values include "legis-body-container", "quoted-block-container", "subtitle-container", "chapter-container", "subchapter-container", "part-container", and "subpart-container".
lowest-bolded-level	Defines the level in which bolded text should be used. Currently, the only in-use value is "division-lowest-bolded".
lowest-level	Indicates the lowest level of text used in a table of contents. Currently, the only in-use values are "section" and "title".
quoted-block	Determines whether or not quoted blocks are used in the table of contents. The value can be either "yes-quoted-block" or "no-quoted-block".
regeneration	Denotes whether or not the table of contents allows for regeneration. The value can be either "yes-regeneration" or "no-regeneration".
<code><toc-entry></code>	Element under <code><toc></code> that's used to define specific entries in a table of contents. Attributes include <code>idref</code> and <code>level</code> .
<code><multi-column-toc-entry></code>	Used to define the contents of multi-column table of content entries. Attributes can include <code>bold</code> and <code>level</code> .
<code><toc-enum></code>	Identifies Table of Contents enumerators.
<code><level-header></code>	Used to identify headers under <code><multi-column-toc-entry></code> . Attributes include <code>level</code> and <code>bold</code> , which denotes whether or not the printed text is bolded. The value can be either "yes" or "no".
<code><target></code>	Element under <code><multi-column-toc-entry></code> that defines items in the last column of Table of Contents.
<code><toc-quoted-entry></code>	Contains Table of Content entries (<code><toc-entry></code>) and uses the <code>style</code> attribute. For example: <pre> <toc-quoted-entry style="OLC"> <toc-entry idref="HA1F6A1E14B7D444096F25DCFA381477E" level="title">Title V—Fair elections financing of House election campaigns</toc-entry> </pre>

Table Elements

Some legislative documents contain tables, which are used to organize data into an easier to read form. For example:

```
<table align-to-level="section" blank-lines-before="1" colsep="0"
frame="topbot" line-rules="no-gen" rowsep="0" rule-weights="4.0.4.0.0.0"
subformat="S6211" table-type="3-Generic:-1-text,-1-num,-1-text">
  <ttitle>Schedule B</ttitle>
  <tdesc>Long-term Schedule B contingent capacity and associated firm energy for
offers of contracts to Boulder Canyon project contractors</tdesc>
  <tgroup cols="3" grid-typeface="1.1" rowsep="0" thead-tbody-ldg-
size="10.10.10" ttitle-size="10">
    <colspec coldef="txt" colname="col1" colsep="1" colwidth="129pts" min-data-
value="125" rowsep="0"/>
    <colspec coldef="fig" colname="col2" colsep="1" colwidth="56pts" min-data-
value="10"/>
    <colspec coldef="txt-no-ldr" colname="col3" colsep="0" colwidth="129pts"
min-data-value="125" rowsep="0"/>
    <tbody>
      <row>
        <entry align="left" colname="col1" leader-modify="force-ldr" rowsep="0"
stub-definition="txt-ldr" stub-hierarchy="1">After such enactment date</entry>
        <entry align="right" colname="col2" leader-modify="force-ldr"
rowsep="0">zero</entry>
        <entry align="left" colname="col3" leader-modify="clr-ldr"
rowsep="0">zero</entry>
      </row>
    </tbody>
  </tgroup>
</table>
```

Like the text hierarchy elements, table elements are organized into a specific order. They include:

<code><table></code>	Defines the properties of the table. Attributes include <code>align-to-level</code> , <code>blank-lines-before</code> , <code>colsep</code> , <code>frame</code> , <code>line-rules</code> , <code>rowsep</code> , <code>rule-weights</code> , <code>table-template-name</code> , <code>table-type</code>
<code><tdesc></code>	Identifies descriptions of tables.
<code><ttitle></code>	Identifies the title of a table.
<code><tgroup></code>	Identifies a group in a table and uses the <code>cols</code> (number of columns in the table, must be greater than zero) and <code>rowsep</code> (0 or 1 value denoting the presence or absence of separator lines. 0 = no separator lines, 1 = visible separator lines) attributes to define the separator properties.
<code><tbody></code>	"Table body," container element for other table details (<code><row></code> and <code><entry></code> elements).
<code><colspec></code>	"Column specification," used to establish settings and sizes in tables.
<code><thead></code>	Container element that defines the table headers. For example:

```
<thead>
  <row>
    <entry align="left" colname="column1" morerows="0"
namest="column1" rowsep="0">
      <bold>If taxable income is:</bold>
    </entry>
    <entry align="right" colname="column2" morerows="0"
namest="column2" rowsep="0">
      <bold>The tax is:</bold>
```

```

        </entry>
      </row>
    </thead>

```

`<row>` Used to contain table entries.

`<entry>` Used to define the properties of a table. Attributes include:

```

    align  States the kind of alignment. In-use values include "left", "right", and
           "center".

    colname Defines the name of the current column.

    colsep  Short for "column separation." Either a "yes" or "no" value.

    morerows Appears to be only used in single row entries, so "0" is the only used value.

    nameend  Short for "ending name." Defines the name of the last column in the table.

    namest   Short for "starting name." Defines the name of the first column in the table (this
           is generally identical to colname).

    rowsep   Short for "row separation." Either a "yes" or "no" value.

    valign   Short for "vertical align." Value can be "top", "middle", or "bottom".

```

Other Organizational Elements

`<continuation-text>` Text that belongs to a hierarchical level above an intervening sublevel. In the example below, the `<paragraph>` element includes `<subparagraph>` elements followed by a return to paragraph-level text within the `<continuation-text>` element:

```

<paragraph id="H717619BF645445278CDF00A75172E376">
  <enum>(2)</enum>
  <header>Hostilities</header>
  <text>The term<term>hostilities</term>includes a situation where
  overt acts of warfare are taking place, such as—</text>
  <subparagraph id="H2954FDFB873946218D2F52F3E7663F36">
    <enum>(A)</enum>
    <text>the exchange of fire between regular or irregular military
  forces, or</text>
  </subparagraph>
  <subparagraph id="H4E9C41AEE28846D29FFFC57AC2BAA4F9">
    <enum>(B)</enum>
    <text>the aerial or naval bombardment of military or civilian
  targets by regular or irregular military forces,</text>
  </subparagraph>
  <continuation-text continuation-text-level="paragraph">regardless of
  whether the Armed Forces are participating in the exchange of fire or
  the bombardment;</continuation-text>
</paragraph>

```

`<enum>` As shown in the text hierarchy, this identifies enumerators for the hierarchical levels of text (such as subsections and paragraphs).

`<header>` Used to tag headers throughout legislative documents. For example:

```

<header>Submission to Congress</header>

```

`<subheader>` Used to tag subheaders throughout legislative documents. Subheaders always come after headers. For example:


```
<header display-inline="yes-display-inline">other federal drug control programs</header><subheader>(including transfers of funds)</subheader>
```

<text> Element that identifies any kind of text that cannot be categorized as a specific level. Sometimes uses the `display-inline` element. For example:

```
<text display-inline="yes-display-inline">That Congress recognizes the centennial of the State of Arizona.</text>
```

<header-in-text> Used to tag headers within text of legislation. Attributes include `level` and `style`. For example:

```
<text display-inline="yes-display-inline">in the second undesignated paragraph, by striking <quote>In like manner</quote> and inserting <quote>(b) <header-in-text level="subsection" style="OLC">Additional disclaimer or dedication</header-in-text>.In the manner set forth in subsection (a),</quote>.</text>
```

<list> Used to define lists. Attributes can include `list-type`, which is always defined as "none", and `level`, which can be "paragraph", "subparagraph", "section", or "subsection". For example:

```
<list list-type="none">
  <list-item>Fiscal year 2011: $1,453,000,000,000.</list-item>
  <list-item>Fiscal year 2012: $874,000,000,000.</list-item>
  <list-item>Fiscal year 2013: $545,000,000,000.</list-item>
  <list-item>Fiscal year 2014: $325,000,000,000.</list-item>
  <list-item>Fiscal year 2015: $280,000,000,000.</list-item>
  <list-item>Fiscal year 2016: $295,000,000,000.</list-item>
  <list-item>Fiscal year 2017: $179,000,000,000.</list-item>
</list>
```

<list-item> Element used to tag individual items under the `<list>` element.

Display/Authoring Elements

Some of the elements do not define specific content, but specify how text will look in print. Others are vestiges of XML authoring software that appear to have no public significance. Such elements include:

<appropriations-major> The `<appropriations-major>` element is only used for large headers in appropriations bills. The "-major" suffix indicates that the header font should be larger than that of `<appropriations-intermediate>` and `<appropriations-small>`. For example:

```
<appropriations-major commented="no"
id="HC01D6666541049D684A272D7947D3940">
  <header display-inline="yes-display-inline">Other Department of
  Defense Programs</header>
</appropriations-major>
```

<appropriations-intermediate> The `<appropriations-intermediate>` element is used to contain text detailing the recipients of appropriated funds. The "-intermediate" suffix indicates that the header font should be smaller than that of `<appropriations-major>` and larger than that of `<appropriations-small>`. For example:

```
<appropriations-intermediate commented="no"
id="H9E78465F5B084C8DA0770BDEBF7270CF">
  <header display-inline="yes-display-inline">United States
  Institute of Peace</header>
  <text display-inline="no-display-inline">For necessary expenses of
  the United States Institute of Peace, as authorized by the United
  States Institute of Peace Act, $30,589,000, to remain available
  until September 30, 2013, which shall not be used for construction
  activities.</text>
```

	<code></appropriations-intermediate></code>
<code><appropriations-small></code>	<p>The small appropriations element is used to contain paragraphs and provisos detailing the recipients of appropriated funds. The “-small” suffix indicates that the header font should be smaller than both <code><appropriations-major></code> and <code><appropriations-intermediate></code>. For example:</p> <pre> <appropriations-small id="H6BFDEA5771974E6EA5E9E6108F256A83"> <text display-inline="no-display-inline">For fees and expenses of jurors as authorized by 28 U.S.C. 1871 and 1876; compensation of jury commissioners as authorized by 28 U.S.C. 1863; and compensation of commissioners appointed in condemnation cases pursuant to rule 71.1(h) of the Federal Rules of Civil Procedure (28 U.S.C. Appendix Rule 71.1(h)), \$51,908,000, to remain available until expended:<proviso> <italic>Provided</italic> </proviso>, That the compensation of land commissioners shall not exceed the daily equivalent of the highest rate payable under 5 U.S.C. 5332.</text> </appropriations-small> </pre>
<code><bold></code>	<p>Denotes a bolded word, term, or section of text. For example:</p> <pre> <bold>Defense Agencies: Outside the United States</bold> </pre>
<code><fraction></code>	<p>Used to denote that a fractional number should be displayed inline. It would direct a parser to print 3/4 as ¾. For example:</p> <pre> <fraction>1/53</fraction> </pre>
<code><editorial/></code>	<p>The purpose of this tag is unclear. This is an empty tag in that it is included within blocks of text, does not have a closing tag, and contains no printed text. Examples include:</p> <pre> <text> <editorial/> <editorial/>20.2 percent of all households with children in the United States experienced food insecurity at some point during the year; and</text> <header>Prohibition on use of sodium cyanide for predator control<editorial/> </header> </pre>
<code><enum-in-header></code>	<p>Identifies enumerators under the <code><header></code> element so they can be formatted properly in print. Header enumerators can include numbers, letters, parentheticals, and abbreviations. For example:</p> <pre> <header>Policy on means-<enum-in-header>T</enum-in-header>ested welfare programs</header><text display-inline="yes-display-inline"> </pre> <p>This prints as the following:</p> <pre> POLICY ON MEANS-TESTED WELFARE PROGRAMS </pre>
<code><italic></code>	<p>Denotes italicized text. This is most frequently used under the <code><proviso></code> element to signify that "Provided" and "Provided further" are italicized when printed, but it is occasionally used to signify italicization elsewhere in a document.</p>
<code><linebreak/></code>	<p>Element that signifies a line break. <code><linebreak/></code> is an empty tag in that it has no closing tag and contains no parsable information.</p>
<code><pagebreak/></code>	<p>Empty tag used to force a page break in the printed text.</p>

<code><quote></code>	<p>Used to identify a quoted line of text. For example:</p> <pre><whereas><text>Whereas according to article I, section 8, of the United States Constitution, <quote>The Congress shall have Power To declare War, grant Letters of Marque and Reprisal, and make Rules concerning Captures on Land and Water</quote>; and</text></whereas></pre>
<code><subscript></code> , <code><superscript></code>	<p>Used to define subscripted and superscripted characters. For example:</p> <pre><subscript>2.5</subscript></pre>
<code><added-phrase></code>	<p>Used for minor additions of text not including amendments. This seems to be used when legislative language is recycled. For example:</p> <pre><text display-inline="yes-display-inline">The Secretary of the Interior may use discretionary funds to pay private attorney fees and costs for employees and former employees of the Department of the Interior reasonably incurred in connection with <added-phrase reported-display-style="italic">Cobell</added-phrase> v. <added-phrase reported-display-style="italic">Salazar</added-phrase> to the extent that such fees and costs are not paid by the Department of Justice or by private insurance.</text></pre> <p>The element can use the <code>reported-display-inline</code> attribute, which is typically valued as <code>"italic"</code> to stylize the text.</p>
<code><deleted-phrase></code>	Used for minor deletions of text not including amendments.
<code><quoted-block></code>	<p>Defined in the House XML guide located at http://xml.house.gov/quoted-block.html as "Textual matter used to amend current law. Quoted block is so named because it is printed or displayed with quotation marks in front of each structural and non-structural level. (These quotation marks are generated for print or display; they are not part of the data.) The name "quoted block" is not meant to imply that the words within this element exist somewhere else and are being quoted. Rather, they signify text to be used to amend current law." Attributes can include <code>id</code>, <code>style</code>, and <code>display-in-line</code>, which can be valued at either <code>"yes-display-in-line"</code> or <code>"no-display-in-line"</code>. For example:</p> <pre><quoted-block display-inline="no-display-inline" id="H4B6127128F7F4C20A858CB85174CFA25" style="OLC"><subparagraph id="H081A8BEBAA854AA49F9E2FC9E5051CC9"><enum>(D)</enum><text display-inline="yes-display-inline">\$100,000 in the case of taxable years beginning in 2013, and</text> </subparagraph><after-quoted-block>, and</after-quoted-block></quoted-block></pre>
<code><quoted-block-continuation-text></code>	<p>Element used within a <code><quoted-block></code> (and before <code><after-quoted-block></code>) reference to denote a text level typically hierarchically higher than the level the text is contained in. This changes the indentation stylizing in the printed version of the bill. This element uses the attribute <code>quoted-block-continuation-text-level</code> to denote the text level, For example:</p> <pre><quoted-block display-inline="no-display-inline" id="HCF01E1B9C7E24913AA9C05B9A4A333BB" style="OLC"><quoted-block-continuation-text quoted-block-continuation-text-level="paragraph">In applying subparagraph (B), a group health plan (or a health insurance issuer with respect to health insurance coverage) may vary premiums and cost-sharing by up to 50 percent of the value of the benefits under the plan (or coverage) based on participation (or lack of participation) in a standards-based wellness program.</quoted-block-continuation-text><after-quoted-block>.</after-quoted-block></quoted-block></pre>

`<after-quoted-block>` Used to define text that is in the same line as `<quoted-block>` but is not actually part of the amendment. In the example above, "and" is tagged with this identifier.

Nominal/Descriptive Elements

House XML currently has some elements that identify meaningful references, terms, and concepts.

Cross-References

The `<external-xref>` element is used to cross-reference legislative, regulatory, and treaty materials external to the document. For example:

```
<external-xref legal-doc="usc" parsable-cite="usc/42/11411">42 U.S.C.  
11411</external-xref>
```

`legal-doc` This attribute labels the type of document being tagged. Accepted values include:

- "act" An act is the primary means of proposing a law in the House.
- "bill" A bill is the primary means of proposing a law in the Senate.
- "executive-order" An executive order is a directive from the president, usually addressed to cabinet agencies.
- "public-law" A public law is a bill or resolution that has been passed and signed into law.
- "regulation" The Code of Federal Regulations (CFR) is a codification of regulations issued by federal agencies.
- "senate-rule" The Standing Rules of the Senate are a series of forty-four rules dictating Senate procedures.
- "statute-at-large" The Statutes at Large is a serial compilation of laws Congress has passed.
- "treaty-tias" Treaties and Other International Acts Series is a compilation of treaties and international legal agreements to which the United States is a party.
- "treaty-ust" A U.S. treaty is a specific agreement between the United States and foreign states or organizations.
- "usc" The U.S. Code is the official compilation of all permanent federal laws.
- "usc-act" This is used to cite an act appearing in the U.S. Code by popular name.
- "usc-appendix" This reference is used to cite a specific appendix in the U.S. Code.
- "usc-chapter" This reference is used to cite a specific chapter of the U.S. Code.
- "usc-subtitle" This reference is used to cite a specific subtitle in the U.S. Code.

`parsable-cite` This attribute defines the type of document with an abbreviated identifier followed by two numbers that make up a citation appropriate to the type of document, formatted `parsable-cite="[legal-doc]/[first #]/[second #]"`. Numbering systems for each `legal-doc`, followed by an example of their use, are as follows:

`public-law` `pl/[Congress#]/[law #]`

```
<external-xref legal-doc="public-law" parsable-  
cite="pl/112/10">Public Law 112-10</external-xref>
```

`regulation` `cfr/[title#]/[chapter#].[subsection#]`

```
<external-xref legal-doc="regulation" parsable-
cite="cfr/10/436.31">section 436.31</external-xref>
```

usc usc/[title#]/[section#]

```
<external-xref legal-doc="usc" parsable-
cite="usc/22/3903">22 U.S.C. 3903</external-xref>
```

usc-act usc-act/[name of act]/[section#]

```
<external-xref legal-doc="usc-act" parsable-cite="usc-
act/Inspector General Act of 1978 /3">section
3(b)</external-xref>
```

usc-chapter usc-chapter/[title#]/[chapter#]

```
<external-xref legal-doc="usc-chapter" parsable-cite="usc-
chapter/31/51">chapter 51</external-xref>
```

The rest of the documented legal-doc types – act, bill, executive-order, senate-rule, statute-at-large, treaty-tias, treaty-ust, usc-appendix – are not used.

The <internal-xref> element is used to reference another part of the current document. For example:

```
<internal-xref idref="HB246BD9FEEB241C9965716920071EB73" legis-
path="202.(a)(1)(A)">subparagraph (A)</internal-xref>
```

idref This attribute is a URI that matches the unique ID of another part of the document. In this case, the reference <subparagraph id="HB246BD9FEEB241C9965716920071EB73"> refers to a subparagraph (A) elsewhere in the document.

legis-path This attribute lists the section, subsection, paragraph, and/or subparagraph to which the internal-xref refers.

Acts

In addition to <external-xref>, House XML sometimes uses a separate system for identifying existing acts of Congress. <act-name> is the element that tags prior law using popular and famous names. For example:

```
<act-name parsable-cite="ESA">Endangered Species Act of 1973</act-name>
```

The parsable-cite attribute identifies an abbreviation for the act. See the table in Appendix 2 for full list of parsable-cite values.

Other Descriptive Elements

<constitution -article> Container element used for constitutional amendments. The only attribute is id, whose value always starts with "H" and is followed by either "1234" or a series of 32 numbers or letters.

<rules- clause> Occasionally House Rules are amended in resolutions, which use a series of hierarchical "rules" elements. These are always structured within a <quoted-block> element and are ordered: <rules-clause>, <rules-clause-header>, <rules-paragraph>, <rules-subparagraph>, <rules-subdivision>, and <rules-item>.

<calendar> An infrequently used element that tags references to legislative calendars (e.g. "Union Calendar No. 439"). For example:

```
<calendar display="yes">Union Calendar No. 439</calendar>
```

The display attribute determines if the information will appear in the printed version of the

document.

- `<preamble>` Container element for "whereas" statements in resolutions.
- `<whereas>` Used to identify the individual "whereas" clauses in a resolution. Typically, this is followed by a `<text>` element. For example:
- ```
<whereas>
 <text>Whereas, on November 17, 2008, Hungary joined the Visa Waiver
 Program;</text>
</whereas>
```
- `<proviso>` Used to tag "*Provided*" and "*Provided further*" provisos. This is typically followed by the `<italic>` tag. For example:
- ```
<proviso>
  <italic>Provided further</italic>
</proviso>
```
- `<term>` Many terms are tagged with the `<term>` element. This is a generic identifier, and it can be used to tag anything from "CBO" to "terrorism" to "aboriginal, indigenous, native people." The purpose of this is evidently to create a collection of terms within a document that give a general idea of what the bill, resolution, or amendment is about. Typically, information tagged by the `<term>` element is followed by a description of the term. For example:
- ```
<text>The term <term>patent infringement claim</term> means any
allegation made to an ANDA filer, whether or not included in a
complaint filed with a court of law, that its ANDA or drug to be
manufactured under such ANDA may infringe any patent.</text>
```

## Frequently Used Attributes

Because attributes can share names and values across different elements, many House XML elements share attributes and some have identically defined sets of values. Some of the most frequently used attributes include:

- `style` This attribute is typically used in structural elements to define the style of the particular document. Values are predefined and include: "defense", "OLC", "tax", "traditional", "USC", "appropriations", and "constitutional-amendment". These dictate the formatting of a particular document or section, such as display issues, text and enumerator generation, and indentations.
- ```
<title id="H55E1C3037E9948A1A5E8B1D3DC6451C6" style="appropriations">
```
- This attribute is used by `<division>`, `<title>`, `<subtitle>`, `<part>`, `<header-in-text>`, `<legis-body>`, `<quoted-block>`, `<resolution-body>`, and `<toc-quoted-entry>` elements.
- `id` and `idref` These attributes are used to uniquely identify specific elements by defining a URI so it can be referenced within the document management system. For example:
- ```
<section id="HAEAC836C266149659DB66C0FC9D821AD">
```
- This attribute is used by `<appropriations-small>`, `<appropriations-intermediate>`, `<appropriations-major>`, `<chapter>`, `<clause>`, `<constitution-article>`, `<division>`, `<internal-xref>`, `<item>`, `<legis-body>`, `<paragraph>`, `<part>`, `<quoted-block>`, `<resolution-body>`, `<rules-clause>`, `<rules-item>`, `<rules-paragraph>`, `<rules-subdivision>`, `<rules-subparagraph>`, `<section>`, `<subchapter>`, `<subclause>`, `<subitem>`, `<subparagraph>`, `<subpart>`, `<subsection>`, `<subtitle>`, `<title>`, and `<toc-entry>` elements.
- `level` This attribute is used in structural elements to define the level of text. Values are predefined

and include: "division", "subdivision", "title", "subtitle", "part", "subpart", "chapter", "subchapter", "section", "subsection", "paragraph", "subparagraph", "clause", "subclause", "item", and "subitem". This is so the text within a specific element can be formatted according to its given level. For example:

```
<toc-entry idref="HE10B8EBD4D8F4D76AB0D4ED2F87D2B7B"
level="section">Sec. 2961. Short title.</toc-entry>
```

This attribute is used by <header-in-text>, <level-header>, <list>, <multi-column-toc-entry>, and <toc-entry> elements.

## Cato XML

Cato XML is a developing set of standards for public documents that expands on the established set of House XML elements. It uses the current set of House XML identifiers (both used and unused) and introduces new elements to identify things frequently referred to in legislative documents. The new and expanded elements cover:

- Organizational units of government (generally federal agencies, bureaus, programs, and projects);
- Existing law;
- Senate and House committees and subcommittees;
- Federal elective officeholders;
- Authorizing language, regulations, and budget authorities (both authorizations of appropriations and appropriations).

They may be extended to cover:

- Regulations and regulatory documents;
- Locations;
- Non-federal governments;
- Dates, times, and durations;
- and so on.

In the future, the standards may be used to mark up other government documents, including:

- Transcripts;
- Committee reports;
- Regulations and regulatory documents;
- Inter- and intra-branch communications;
- and so on.

Including more elements and attributes from a standardized XML vocabulary will make more available the specific content of these documents, and it will create much needed links among actors and entities in government processes. The end-result we hope for is better computer-aided oversight of government (both internal and external) and a more transparent government.

## Cato Namespace

To differentiate Cato XML from House XML, all Cato XML documents include a namespace declaration in the root element and include the `cato:` prefix in Cato XML elements. The Cato namespace helps external developers and XML analysts know which elements have what provenance. For example:

```
<bill xmlns:cato="http://namespaces.cato.org/catoxml" bill-stage="Introduced-in-House"
bill-type="olc" dms-id="HEC9B27C65C814B5EA33B456D6807413B" public-private="public">
...
<text>Lands may be leased pursuant to this subtitle to any person qualified to obtain
a lease for deposits of oil and gas under the <cato:entity entity-type="law-
citation"><cato:entity-ref entity-type="act">Mineral Leasing Act </cato:entity-ref>
(<external-xref legal-doc="usc" parsable-cite="usc/30/181">30 U.S.C. 181 et
seq.</external-xref>)</cato:entity.</text>
...
```

```
</bill>
```

As previously mentioned, Cato XML can use any method for declaring namespaces. The above example declares the namespace in the root element `<bill>` and uses the `cato:` prefix to denote Cato XML elements.

## The `<entity-ref>`, `<entity>`, and `<property>` Elements

One of the most integral Cato XML elements is `<entity-ref>`. This element is used to identify references to things not in the document, including federal organizational units (executive branch agencies, bureaus, and so on, as well as organizational units in the legislative and judicial branches), federal elective officeholders, legal citations, and other things frequently referenced in government documents. The `<entity-ref>` element is formatted as follows:

```
<cato:entity-ref entity-type="[entity-type]" entity-id="XXX-XX">entity
text</cato:entity-ref>
```

For example:

```
<text display-inline="yes-display-inline">The Departments of <cato:entity-ref
entity-type="federal-body" entity-id="1500">Justice</cato:entity-ref>,
<cato:entity-ref entity-type="federal-body" entity-
id="2000">Treasury</cato:entity-ref>, and <cato:entity-ref entity-
type="federal-body" entity-id="7000">Homeland Security</cato:entity-ref> are
the agencies with the greatest capacity to investigate, prosecute and dismantle
the highest level of drug trafficking organizations. Low-level drug offender
investigations and prosecutions divert Federal personnel and resources from
prosecuting high-level traffickers.</text>
```

Because `<entity-ref>` is used for a variety of things, the `entity-type` attribute identifies the type of entity being referred to. In the above examples, this is "federal-body", meaning a federal government body. The `entity-id` is a unique identifier that ties the reference to a specific federal government body in the federal-body lookup table. While "Treasury" is not the official name of the agency, the `entity-id` of "1500" allows any XML parsing software to correlate the term to its official name—in this case, the Department of the Treasury. All Cato XML references with unique identifiers are defined in **lookup tables**, which are discussed in greater detail below.

The purposes of the general `<entity-ref>` element – versus creating unique and compartmentalized elements for each entity type – are to:

- Simplify the tagging process by putting all "things" under one frequently used element
- Uniquely identify entities through cross-referenced attributes
- Allow for convenient data updating in the referenced lookup tables
- Make it easier for developers to include all of the new identifiers in their XML software
- Keep references generic so they can be used more easily by external authors of non-Cato XML documents

The `<entity>` element is used to identify things in the document itself, such as parallel law citations and new exercises of federal authority.

### Federal Organizational Units

One of the uses for this element is identifying federal organizational units such as agencies and bureaus. Agencies, bureaus, and programs are referenced countless times throughout legislative documents but are not delineated with XML in any way. In the example above, the Department of Justice, the Department of the Treasury, and the Department of Homeland Security are identified using the `<entity-ref>` element. Because the `entity-type` is valued as "federal-body", the software parsing or using the XML refers to the federal organizational unit lookup table. Thanks to the unique value assigned to its `entity-id`, the software can discover the specific entry in the table. In the above example, the numbers (1500, 2000, and 7000) are references to the agencies themselves (Department of Justice, Department of the Treasury, and Department of Homeland Security, respectively).

### Committees and Subcommittees

Another use of the `<entity-ref>` element is identifying committees and subcommittees throughout legislative text. As previously mentioned, House XML already identifies committees in the procedural section of a document with



<committee-name>, but references within legislative text have been left unidentified. Cato XML identifies these references using the <entity-ref> element. For example:

```
<text>For the purposes of this Act, the term <term>congressional defense committees</term> means the <cato:entity-ref entity-type="committee" entity-id="HAS00">Armed Services Committee of the House of Representatives</cato:entity-ref>, the <cato:entity-ref entity-type="committee" entity-id="SSAS00">Armed Services Committee of the Senate</cato:entity-ref>, the <cato:entity-ref entity-type="committee" entity-id="SSAP02">Subcommittee on Defense of the Committee on Appropriations of the Senate</cato:entity-ref>, and the <cato:entity-ref entity-type="committee" entity-id="HAP02">Subcommittee on Defense of the Committee on Appropriations of the House of Representatives</cato:entity-ref>.</text>
```

As seen, committee references give the `entity-type` attribute the value "committee" and the `entity-id` attribute as its official committee code (described earlier on page 9 for <committee-name>). All full committee IDs begin with their official committee identifier followed by 00. Subcommittees IDs begin with their parent committee's identifier followed by their unique two-digit number documented by the House and Senate. For committee and subcommittee codes, see Appendix 1.

### Federal Elective Officeholders

The <entity-ref> element is also used to identify elected officials mentioned in text. While officials are identified in the procedural section using the <sponsor> and <cosponsor> elements, references within legislative text are not identified. Cato XML identifies these references using the <entity-ref> element. For example:

```
<text>Whereas <cato:entity-ref entity-type="person" entity-id="B000243">Senator Baucus</cato:entity-ref>, in a Senate Finance Committee hearing he chaired on March 9, 2011, with Ambassador Kirk as the sole witness, said that<quote>... the time has passed to ratify the Colombian free trade agreement. It's long past. I mentioned in my opening remarks, we're losing market share hand over fist. Hand over fist, we're losing market share This is a no-brainer, Mr. Ambassador, no-brainer, and I just hope we get this passed quickly</quote>;</text>
```

As seen, elected officials references value the `entity-type` attribute as "person" and the `entity-id` attribute as the official's Congressional Bioguide identifier (described earlier on page 8). Non-elected officials do not currently have unique identifiers and cannot be identified at this time.

### Existing Law and Parallel Law Citations

The <entity-ref> element is also used to identify references to existing laws. In most bills and resolutions, most references to existing laws first use a commonly cited popular name, followed by its Public Law number, a Statutes at Large citation, or a U.S. Code citation. This is called a parallel citation, as the popular name and the U.S. code citation refer to the same thing. For example:

```
<text>The withdrawal, limitation, or suspension of United States security assistance in accordance with section 502B of the Foreign Assistance Act of 1961 (22 U.S.C. 2304).</text>
```

In Cato XML, these are identified by marking up each component individually with the <entity-ref> identifier and grouping together the references as a single entity with a container element <entity>. Such text is marked up as follows:

```
<text>The withdrawal, limitation, or suspension of United States security assistance in accordance with <cato:entity type="law-citation"><cato:entity-ref entity-type="act" value="Foreign Assistance Act of 1961/sec:502B">section 502B of the Foreign Assistance Act of 1961</cato:entity-ref> (<cato:entity-ref entity-type="uscode" value="usc/22/2304">22 U.S.C. 2304</cato:entity-ref></cato:entity>).</text>
```

The <entity> element – as it is used in this instance – creates the parallel citation. The attribute `law-citation` signals that the child <entity-ref> elements each point to the same thing. All popular act names, Public Law

numbers, Statutes-at-Large citations, and U.S. Code citations are then labeled with the `entity-type` that corresponds to the type of citation to existing law they are.

The `<entity-ref>` element uses two attributes for law citations, `entity-type` and `entity-value`. The `entity-type` attribute signifies what type of reference to existing law it is. It can have one of the following values:

"act" The popular name for an act. When `entity-type="act"`, the `entity-value` attribute is assigned with a normalized title of the act. This is generally the most common name assigned to the act by the House of Representatives. A list of these names can be found at <http://uscode.house.gov/popularnames/popularnames.htm>.

Popular names are typically preceded by a reference to a section number (but they can also be preceded by a reference to any text level, multiple text levels, or nothing at all). In such instances, the reference shall be included in the `<property>` text node as well as the `value` attribute. For example:

```
<cato:entity entity-type="law-citation"><cato:entity-ref entity-type="act" entity-value="Personal Responsibility and Work Opportunity Reconciliation Act of 1996/s:402/ss:b/p:2">Section 402(b)(2) of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996</cato:entity-ref> (<cato:entity-ref entity-type="uscode" entity-value="usc/8/1612/b/2">8 U.S.C. 1612(b)(2)</cato:entity-ref>)</cato:entity>
```

In this example, "section 402" also includes references to subsection b and paragraph 2 (these are expressed in the parentheticals after the section reference). To denote what text levels are expressed in the text node, prefixes are added to the segments of the `entity-value`. These are always ordered from highest to lowest, and may vary from act to act. Text level prefixes include:

Division: d:	Subpart: spt:	Subsection: ss:	Subclause: scl:
Title: t:	Chapter: ch:	Paragraph: p:	Item: i:
Subtitle: st:	Subchapter: sch:	Subparagraph: sp:	Subitem: si:
Part: pt:	Section: s:	Clause: cl:	

"uscode" Machine-readable U.S. Code citation. When `entity-type="uscode"`, the `value` attribute uses the equivalent of the `parsable-cite` attribute from House XML's `<external-xref>`. For example:

```
<cato:entity-ref entity-type="uscode" entity-value="usc/20/9202">20 U.S.C. 9202</cato:entity-ref>
```

Specific U.S. Code chapters and appendix references are cited similarly to their `<external-xref>` counterparts. Examples include:

```
<cato:entity-ref entity-type="uscode" entity-value="usc-chapter/38/17">chapter 17 of title 38</cato:entity-ref>
```

```
<cato:entity-ref entity-type="uscode" entity-value="usc-appendix/50/533">50 U.S.C. App. 533</cato:entity-ref>
```

If a U.S. Code citation includes additional text levels (such as subsections, paragraphs, subparagraphs, and clauses), additional slashes (beyond those used in House XML) denote the extended citation. For example:

```
<cato:entity-ref entity-type="uscode" entity-value="usc/8/1612/b/2">8 U.S.C. 1612(b)(2)</cato:entity-ref>
```

U.S. Code citations with either "note" or "et. seq." are also included as an extended citation value. Examples include:

```
<cato:entity-ref entity-type="uscode" entity-value="usc/8/1103/note">8
U.S.C. 1103 note</cato:entity-ref>
```

```
<cato:entity-ref entity-type="uscode" entity-value="usc/50/541/etseq">50
U.S.C. 541 et seq.</cato:entity-ref>
```

"public-law" Machine-readable public law citation. Similar to "uscode", when name="public-law", the value attribute will use the same citation as Public Laws under <external-xref>. For example:

```
<cato:entity-ref entity-type="public-law" entity-
value="pl/111/148">Public Law 111-148</cato:entity-ref>
```

Sections, subsections, and other hierarchical levels of text are denoted the same as they are for U.S. Code citations.

"statute-at-large" Machine-readable citation to Statutes at Large.<sup>1</sup> When entity-type="statute-at-large", the value attribute will use a similar, normalized citation method as "public-law" and "usc". In this case, the citation for value will be formatted entity-value="stat/[volume#]/[page#]". For example:

```
<cato:entity-ref entity-type="statute-at-large" entity-
value="stat/64/568">64 Stat. 568</cato:entity-ref>
```

The act, uscode, public-law, and statute-at-large entity types do not use the entity-id attribute for two reasons: there are no universally agree upon identifiers for these references, and different references may have different values because they refer to existing law differently or in different sources.

If either of the House XML elements <act-name> or <external-xref> has been applied to a law citation, the <entity-ref> element is not applied to the law citation tagged with House XML, though it and parallel law citations are wrapped as such in the <entity> tag. To illustrate, text that looks like the following in House XML:

```
<text>The term <term>Smart Grid</term>has the meaning provided by section 1301
of the Energy Independence and Security Act of 2007 (<external-xref legal-
doc="usc" parsable-cite="usc/42/17381">42 U.S.C. 17381</external-xref>).</text>
```

Would look like the following in Cato XML:

```
<text>The term <term>Smart Grid</term>has the meaning provided by <entity
xmlns="http://namespaces.cato.org/catoxml" type="law-citation"><entity-ref
entity-type="act" entity-value="Energy Independence and Security Act of
2007/sec:1301">section 1301 of the Energy Independence and Security Act of
2007</entity-ref> (<external-xref xmlns="" legal-doc="usc" parsable-
cite="usc/42/17381">42 U.S.C. 17381</external-xref></entity>).</text>
```

Note how in Cato XML the xmlns="" attribute has been added to <external-xref> to nullify the namespace properties of the <entity> element. Likewise, wherever <act-name> is already used by House XML, <entity-ref> identifies the parallel citation.

### Proposed Laws and Citations

Frequently, legislative documents will mention act names and legal citations (most frequently amendments and additions) that have yet to be enacted into law. For example:

```
<text display-inline="yes-display-inline">The Administrator shall establish an
abbreviated application process for applicants that are managers of small
business investment companies that are licensed under section 301 and that are
applying to participate in the program. Such abbreviated process shall
```

---

<sup>1</sup> The name of the resource is Statutes at Large, plural, but House XML uses statute-at-large, singular. Cato XML follows House XML to simplify.

incorporate a presumption that such managers satisfactorily meet the selection criteria under paragraphs (3) and (5) of section 399D(b).</text>

In this block of text, the references to "*paragraphs (3) and (5) of section 399D(b)*" are references to paragraphs and sections in a proposed amendment. To delineate these references from existing law, Cato XML uses the attribute `proposed` (with the possible value of "true") to indicate that the reference does not yet exist as a popular name, public law, Statute at Large, or U.S. Code citation. This example would be marked up as follows:

```
<text display-inline="yes-display-inline">The Administrator shall establish an abbreviated application process for applicants that are managers of small business investment companies that are licensed under section 301 and that are applying to participate in the program. Such abbreviated process shall incorporate a presumption that such managers satisfactorily meet the selection criteria under <cato:entity-ref entity-type="act" proposed="true" value="Small Business Investment Act of 1958/s:399D/ss:b/p:3">paragraphs (3)</cato:entity-ref> and <cato:entity-ref entity-type="act" value="Small Business Investment Act of 1958/s:399D/ss:b/p:5" proposed="true">(5) of section 399D(b)</cato:entity-ref>.</text>
```

## New Authority

Along with amending existing exercises of authority (a very difficult challenge to render as data), a purpose of legislation is to exercise new congressional authority. Cato XML begins to represent new exercises of authority in at least a "thin" way.

The `<entity>` element is used to define new authority entities in Cato XML documents. The `entity-type` attributes are `auth-authorization`, `auth-regulation`, `auth-auth-approp`, and `auth-approp`, reflecting the four types of authority:

"auth-authorization" **Authorizations** consist of language that creates organs of government or defines the powers and functions of existing organs of government. Authorization language typically consists of defining what an agency, bureau, or government office "shall" do. For example:

```
<text>The Attorney General shall ensure that the amendment made by paragraph (1) is implemented by not later than 6 months after the date of the enactment of this Act.</text>
```

Authorizations use the `entity-type` value of "auth-authorization". The above text would be marked up as follows:

```
<text><cato:entity entity-type="auth-authorization">The Attorney General shall ensure that the amendment made by paragraph (1) is implemented by not later than 6 months after the date of the enactment of this Act.</cato:entity></text>
```

"auth-regulation" **Regulations** consist of language that directs subjects of federal authority to act or refrain from acting in particular ways. This can include civilians, state governments, and private businesses. For example:

```
<text display-inline="yes-display-inline">Any employer who willfully violates the posting requirements of this section shall be subject to a civil fine in an amount not to exceed $100 for each separate offense.</text>
```

Regulations use the `entity-type` value of "auth-regulation". The above text would be marked up as follows:

```
<text display-inline="yes-display-inline"><cato:entity entity-type="auth-regulation">Any employer who willfully violates the posting requirements of this section shall be subject to a civil fine in an amount not to exceed $100 for each separate offense.</cato:entity></text>
```

"auth-auth-  
approp"

**Budget authorities: authorizations of appropriations** are exercises of congressional authority that authorize the enactment of appropriations for an agency or program. This language typically calls for chunks of funds to be spent in a given fiscal year (or years) for a given agency, bureau, or office. For example:

```
<text display-inline="no-display-inline">Funds are hereby authorized to be appropriated for the Department of Defense for fiscal year 2012 for expenses, not otherwise provided for, for Drug Interdiction and Counter-Drug Activities, Defense-wide, in the amount of $1,156,282,000.</text>
```

Authorizations of appropriations use the entity-type value of "auth-auth-approp". The above text would be marked up as follows:

```
<text display-inline="no-display-inline"><cato:entity entity-type="auth-auth-approp">Funds are hereby authorized to be appropriated for the <cato:entity-ref entity-type="federal-body" entity-id="9700">Department of Defense</cato:entity-ref> for fiscal year 2012 <cato:property name="purpose">for expenses, not otherwise provided for, for Drug Interdiction and Counter-Drug Activities, Defense-wide</cato:property>, in the amount of <cato:funds-and-year amount="1156282000" year="2012">$1,156,282,000</cato:funds-and-year>.</cato:entity></text>
```

"auth-approp"

**Budget authorities: appropriations** consist of language that directs the obligation and outlay of funds (previously authorized) for specific offices and purposes. This language is typically more specific than that of authorizations of appropriations and frequently mentions specific projects. For example:

```
<text>For expenses, not otherwise provided for, necessary for the operation and maintenance of the Air Force, as authorized by law; and not to exceed $7,699,000 can be used for emergencies and extraordinary expenses, to be expended on the approval or authority of the Secretary of the Air Force, and payments may be made on his certificate of necessity for confidential military purposes, $36,062,989,000.</text>
```

Appropriations use the entity-type value of "auth-approp". The above text would be marked up as follows:

```
<text><cato:entity entity-type="auth-approp">For expenses, not otherwise provided for, necessary for the operation and maintenance of the <cato:entity-ref entity-type="federal-body" entity-id="5700">Air Force</cato:entity-ref>, as authorized by law; and not to exceed $7,699,000 can be used for emergencies and extraordinary expenses, to be expended on the approval or authority of the <cato:entity-ref entity-type="federal-body" entity-id="5700">Secretary of the Air Force</cato:entity-ref>, and payments may be made on his certificate of necessity for confidential military purposes, <cato:funds-and-year value="36062989000" year="2011">$36,062,989,000</cato:funds-and-year>.</cato:entity></text>
```

### Budget Authority Properties

Budget authorities (both authorizations of appropriations and appropriations) have unique elements that define their properties. These elements are:

<property  
name=  
"purpose">

This captures the text node detailing the purpose of the authorization or appropriation. If no value attribute is provided, the value is assumed to be the text node. For example:

```
<text display-inline="no-display-inline"><cato:entity entity-
type="auth-approp"><cato:property name="purpose">For an additional
amount for <quote>Systems Acquisition</quote>, for necessary expenses
related to the consequences of Hurricane Sandy</cato:property>,
$3,869,000, to remain available until September 30, 2014:
<italic>Provided</italic>, That such amount is designated by the
Congress as being for an emergency requirement pursuant to section
251(b)(2)(A)(i) of the Balanced Budget and Emergency Deficit Control
Act of 1985.</cato:entity></text>
```

```
<property
name="funds-
source">
```

This is used to capture the source of the appropriated funds. If omitted, it is assumed to be the U.S. Treasury, but occasionally it can be a trust fund or some other source. If no value attribute is provided, the value is assumed to be the text node. For example:

```
<text display-inline="no-display-inline"><cato:entity entity-
type="auth-approp">For expenses necessary for the Armed Forces
Retirement Home to operate and maintain the Armed Forces Retirement
Home—Washington, District of Columbia, and the Armed Forces Retirement
Home—Gulfport, Mississippi, to be paid from funds available in the
<cato:property name="funds-source">Armed Forces Retirement Home Trust
Fund</cato:property>, $67,700,000, of which $2,000,000 shall remain
available until expended for construction and renovation of the
physical plants at the Armed Forces Retirement Home—Washington,
District of Columbia, and the Armed Forces Retirement Home—Gulfport,
Mississippi.</cato:entity></text>
```

```
<funds-and-
year>
```

Captures the U.S. dollar value of an appropriated amount as well as the year or years in which the funds are available. For example:

```
<text display-inline="no-display-inline"><cato:entity entity-
type="auth-approp">For an additional amount for
<quote>Construction</quote> for necessary expenses related to the
consequences of Hurricane Sandy, <cato:funds-and-year
amount="49875000" year="indefinite">$49,875,000</cato:funds-and-year>,
to remain available until expended:<italic>Provided</italic>, That
such amount is designated by the Congress as being for an emergency
requirement pursuant to section 251(b)(2)(A)(i) of the Balanced Budget
and Emergency Deficit Control Act of 1985.</cato:entity></text>
```

Unlike "fund-source" and "purpose", <funds-and-year> does not use the <property> element. This is because one <entity> authorization may have multiple appropriations for multiple years.

Attributes for <funds-and-year> include:

- amount The amount appropriated or authorized to be appropriated for the stated time period. This can be valued in either U.S. dollars (such as "200000000") or as "indefinite".
- year The fiscal year or years in which the dollar amount is relevant. This can be a single year (such as "2013"), a list of years (such as "2014, 2015, 2016"), a year range (such as "2014-2020"), or "indefinite" when there is no clear end to the funding.

These elements must be contained within the <entity> element with the entity-type value of either "auth-approp" or "auth-auth-approp".

## Lookup Tables

As noted above, whenever a Cato XML element has a unique identifier, the reference is tied to a lookup table. Lookup tables are XML formatted documents that assign uniform values to the various names an entity may go by.

Each table identifies a different set of entity types, such as federal agencies and bureaus, elected officials, and committee and subcommittee names. Lookup table entries look like the following:

```
<entity id="SSEG00">
 <name role="official">Senate Committee on Energy and Natural
Resources</name>
 <name>Senate Energy and Natural Resources Committee</name>
</entity>
<entity id="SSEG01" parent-id="SSEG00">
 <name role="official">Subcommittee on Energy</name>
 <name>Energy Subcommittee</name>
</entity>
<entity id="SSEG03" parent-id="SSEG00">
 <name role="official">Subcommittee on Public Lands and Forests</name>
 <name>Public Lands and Forests Subcommittee</name>
</entity>
```

In this example, the table entries cover a committee and two of its subcommittees. The attributes under `<entity>` tie all of its child elements (in this case, just `<name>`) to one value. While one entity (in this case, either a committee or subcommittee) may have multiple names in legislative documents, the varying names refer to one unique entity. The `<name>` element with the attribute `role="official"` denotes that it is the official name assigned to the entity, as distinct from colloquial names that may appear in government documents and in the table.

### The `<entity>` and `<entities>` Elements

All lookup tables have a root element of `<entities>`. This has the required attribute of `type`, which denotes the content of the lookup table. This attribute also matches the `entity-type` value used in `<entity>` and `<entity-ref>` elements used throughout Cato XML documents. The `<entities>` element also has the required attribute `updated`, which details the last time and date the lookup table was updated. For example:

```
<entities type="act" updated="2013-01-09T12:31:52">
```

The `<entities>` element can also use a `version` attribute. This is used when there may be multiple different versions of a `type` or `entity-type`. For example, agencies and bureaus, which use the `type` value "federal-body", can change over time, so multiple versions of a federal entities lookup table would be necessary to capture the entirety of the information.

The `<entity>` element is used for all unique listings in a lookup table. Attributes include `id`, which is the unique identifier for the individual listing and is required for all entries, and `parent-id`, which is only required when the entity is the child of another entity. In the committees lookup table, the `parent-id` is used to denote which committee is the parent entity to a subcommittee; in the federal entities lookup table, the `parent-id` is used to denote which agency or bureau the given entity resides under. For example:

```
<entity id="1100">
 <name role="official">Executive Office of the President</name>
</entity>
<entity id="1101" parent-id="1100">
 <name role="official">President of the United States</name>
</entity>
<entity id="1102" parent-id="1100">
 <name role="official">National Security Council</name>
</entity>
<entity id="1103" parent-id="1100">
 <name role="official">Office of Management and Budget</name>
 <abbr role="official">OMB</abbr>
</entity>
```

Under the `<entity>` element are the `<name>` and `<abbr>` elements. The `<name>` element is used for all full names of entities, while the `<abbr>` element is used for any abbreviations. Both of these elements use the `role` attribute, which explains the status of the given name or abbreviation. The `role` attribute is typically valued at either

---

"official" (for the current name or abbreviation of the office) or "historical" (for any names or abbreviations a federal entity may have gone by in the past).



## Appendix

**Table 1: Committee and Subcommittee Codes**

House XML ID	Official Committee ID	Committee Name
HAG00	HAG00	Committee on Agriculture
(none)	HAG03	Nutrition and Horticulture Subcommittee
(none)	HAG14	Rural Development, Research, Biotechnology, and Foreign Agriculture Subcommittee
(none)	HAG15	Conservation, Energy, and Forestry Subcommittee
(none)	HAG16	General Farm Commodities and Risk Management Subcommittee
(none)	HAG22	Department Operations, Oversight, and Credit Subcommittee
(none)	HAG29	Livestock, Dairy, and Poultry Subcommittee
HAP00	HAP00	Committee on Appropriations
(none)	HAP01	Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Subcommittee
(none)	HAP02	Defense Subcommittee
(none)	HAP04	State, Foreign Operations, and Related Programs Subcommittee
(none)	HAP06	Interior, Environment, and Related Agencies Subcommittee
(none)	HAP07	Labor, Health and Human Services, Education, and Related Agencies Subcommittee
(none)	HAP10	Energy and Water Development, and Related Agencies Subcommittee
(none)	HAP15	Homeland Security Subcommittee
(none)	HAP18	Military Construction, Veterans Affairs, and Related Agencies Subcommittee
(none)	HAP19	Commerce, Justice, Science, and Related Agencies Subcommittee
(none)	HAP20	Transportation, and Housing and Urban Development, and Related Agencies Subcommittee
(none)	HAP23	Financial Services and General Government Subcommittee
(none)	HAP24	Legislative Branch Subcommittee
HAS00	HAS00	Committee on Armed Services
(none)	HAS02	Military Personnel Subcommittee
(none)	HAS03	Readiness Subcommittee
(none)	HAS06	Oversight and Investigations Subcommittee
(none)	HAS25	Tactical Air and Land Forces Subcommittee
(none)	HAS26	Emerging Threats and Capabilities Subcommittee
(none)	HAS28	Seapower and Projection Forces Subcommittee
(none)	HAS29	Strategic Forces Subcommittee
HBA00	HBA00	Committee on Financial Services
(none)	HBA04	Insurance, Housing and Community Opportunity Subcommittee
(none)	HBA09	Oversight and Investigations Subcommittee
(none)	HBA15	Financial Institutions and Consumer Credit Subcommittee
(none)	HBA16	Capital Markets and Government Sponsored Enterprises
(none)	HBA19	Domestic Monetary Policy and Technology Subcommittee
(none)	HBA20	International Monetary Policy and Trade Subcommittee
HBU00	HBU00	Committee on the Budget

(none)	HED02	Health, Employment, Labor, and Pensions Subcommittee
(none)	HED10	Workforce Protections Subcommittee
(none)	HED13	Higher Education and Workforce Training Subcommittee
(none)	HED14	Early Childhood, Elementary, and Secondary Education Subcommittee
HFA00	HFA00	Committee on Foreign Affairs
(none)	HFA05	Asia and the Pacific Subcommittee
(none)	HFA07	The Western Hemisphere Subcommittee
(none)	HFA13	The Middle East and South Asia Subcommittee
(none)	HFA14	Europe and Eurasia Subcommittee
(none)	HFA16	Africa, Global Health, and Human Rights Subcommittee
(none)	HFA17	Oversight and Investigations Subcommittee
(none)	HFA18	Terrorism, Nonproliferation, and Trade Subcommittee
HGO00	HGO00	Committee on Oversight and Government Reform
(none)	HGO06	National Security, Homeland Defense and Foreign Operations Subcommittee
(none)	HGO24	Government Organization, Efficiency and Financial Management Subcommittee
(none)	HGO25	Federal Workforce, U.S. Postal Service and Labor Policy Subcommittee
(none)	HGO27	Health Care, District of Columbia, Census and the National Archives Subcommittee
(none)	HGO28	Regulatory Affairs, Stimulus Oversight and Government Spending Subcommittee
(none)	HGO29	TARP, Financial Services and Bailouts of Public and Private Programs Subcommittee
(none)	HGO30	Technology, Information Policy, Intergovernmental Relations and Procurement Reform Subcommittee
HHA00	HHA00	Committee on House Administration
(none)	HHA06	Oversight Subcommittee
(none)	HHA08	Elections Subcommittee
HHM00	HHM00	Committee on Homeland Security
(none)	HHM05	Counterterrorism and Intelligence Subcommittee
(none)	HHM07	Transportation Security Subcommittee
(none)	HHM08	Cybersecurity, Infrastructure Protection, and Security Technologies Subcommittee
(none)	HHM09	Oversight, Investigations, and Management Subcommittee
(none)	HHM11	Border and Maritime Security Subcommittee
(none)	HHM12	Emergency Preparedness, Response and Communications Subcommittee
HIF00	HIF00	Committee on Energy and Commerce
(none)	HIF02	Oversight and Investigations Subcommittee
(none)	HIF03	Energy and Power Subcommittee
(none)	HIF14	Health Subcommittee
(none)	HIF16	Communications and Technology Subcommittee
(none)	HIF17	Commerce, Manufacturing, and Trade Subcommittee
(none)	HIF18	Environment and the Economy Subcommittee
HIG00	HIG00	Permanent Select Committee on Intelligence
(none)	HIG05	Technical and Tactical Intelligence Subcommittee

(none)	HIG08	Terrorism, Human Intelligence, Analysis, and Counterintelligence Subcommittee
(none)	HIG09	Oversight and Investigations Subcommittee
HII00	HII00	Committee on Natural Resources
(none)	HII06	Energy and Mineral Resources Subcommittee
(none)	HII10	National Parks, Forests and Public Lands Subcommittee
(none)	HII13	Water and Power Subcommittee
(none)	HII22	Fisheries, Wildlife, Oceans and Insular Affairs Subcommittee
(none)	HII24	Indian and Alaska Native Affairs Subcommittee
HJU00	HJU00	Committee on the Judiciary
(none)	HJU01	Immigration Policy and Enforcement Subcommittee
(none)	HJU03	Courts, Commercial and Administrative Law Subcommittee
(none)	HJU08	Crime, Terrorism, and Homeland Security Subcommittee
(none)	HJU10	The Constitution Subcommittee
(none)	HJU12	Intellectual Property, Competition, and the Internet Subcommittee
HPW00	HPW00	Committee on Transportation and Infrastructure
(none)	HPW02	Water Resources and Environment Subcommittee
(none)	HPW05	Aviation Subcommittee
(none)	HPW07	Coast Guard and Maritime Transportation Subcommittee
(none)	HPW12	Highways and Transit Subcommittee
(none)	HPW13	Economic Development, Public Buildings, and Emergency Management Subcommittee
(none)	HPW14	Railroads, Pipelines, and Hazardous Materials Subcommittee
HRU00	HRU00	Committee on Rules
(none)	HRU02	Legislative and Budget Process Subcommittee
(none)	HRU04	Rules and Organization of the House Subcommittee
HSM00	HSM00	Committee on Small Business
(none)	HSM23	Contracting and Workforce Subcommittee
(none)	HSM24	Investigations, Oversight and Regulations Subcommittee
(none)	HSM25	Agriculture, Energy and Trade Subcommittee
(none)	HSM26	Healthcare and Technology Subcommittee
(none)	HSM27	Economic Growth, Tax and Capital Access Subcommittee
HSO00	HSO00	Committee on Ethics
(none)	HSY14	Research and Science Education Subcommittee
(none)	HSY16	Space and Aeronautics Subcommittee
(none)	HSY19	Technology and Innovation Subcommittee
(none)	HSY20	Energy and Environment Subcommittee
(none)	HSY21	Investigations and Oversight Subcommittee
HVR00	HVR00	Committee on Veterans' Affairs
(none)	HVR03	Health Subcommittee
(none)	HVR08	Oversight and Investigations Subcommittee
(none)	HVR09	Disability Assistance and Memorial Affairs Subcommittee
(none)	HVR10	Economic Opportunity Subcommittee
HWM00	HWM00	Committee on Ways and Means

(none)	HWM01	Social Security Subcommittee
(none)	HWM02	Health Subcommittee
(none)	HWM03	Human Resources Subcommittee
(none)	HWM04	Trade Subcommittee
(none)	HWM05	Select Revenue Measures Subcommittee
(none)	HWM06	Oversight Subcommittee
SSAF00	SSAF00	Senate Committee on Agriculture, Nutrition, and Forestry
(none)	SSAF13	Subcommittee on Commodities, Markets, Trade and Risk Management
(none)	SSAF14	Subcommittee on Conservation, Forestry and Natural Resources
(none)	SSAF15	Subcommittee on Jobs, Rural Economic Growth and Energy Innovation
(none)	SSAF16	Subcommittee on Nutrition, Specialty Crops, Food and Agricultural Research
(none)	SSAF17	Subcommittee on Livestock, Dairy, Poultry, Marketing and Agriculture Security
SSAP00	SSAP00	Senate Committee on Appropriations
(none)	SSAP01	Subcommittee on Agriculture, Rural Development, Food and Drug Administration, and Related Agencies
(none)	SSAP02	Subcommittee on Department of Defense
(none)	SSAP08	Subcommittee on Legislative Branch
(none)	SSAP14	Subcommittee on Department of Homeland Security
(none)	SSAP16	Subcommittee on Commerce, Justice, Science, and Related Agencies
(none)	SSAP17	Subcommittee on Department of the Interior, Environment, and Related Agencies
(none)	SSAP18	Subcommittee on Departments of Labor, Health and Human Services, and Education, and Related Agencies
(none)	SSAP19	Subcommittee on Military Construction and Veterans Affairs, and Related Agencies
(none)	SSAP20	Subcommittee on State, Foreign Operations, and Related Programs
(none)	SSAP22	Subcommittee on Energy and Water Development
(none)	SSAP23	Subcommittee on Financial Service and General Government
(none)	SSAP24	Subcommittee on Transportation and Housing and Urban Development, and Related Agencies
SSAS00	SSAS00	Committee on Armed Services
(none)	SSAS13	Subcommittee on SeaPower
(none)	SSAS14	Subcommittee on Airland
(none)	SSAS15	Subcommittee on Readiness and Management Support
(none)	SSAS16	Subcommittee on Strategic Forces
(none)	SSAS17	Subcommittee on Personnel
(none)	SSAS20	Subcommittee on Emerging Threats and Capabilities
SSBK00	SSBK00	Senate Committee on Banking, Housing, and Urban Affairs Budget
(none)	SSBK04	Subcommittee on Securities, Insurance and Investment
(none)	SSBK05	Subcommittee on Security and International Trade and Finance
(none)	SSBK08	Subcommittee on Financial Institutions and Consumer Protection
(none)	SSBK09	Subcommittee on Housing, Transportation and Community Development
(none)	SSBK12	Subcommittee on Economic Policy
SSBU00	SSBU00	Senate Committee on the Budget

SSCM00	SSCM00	Senate Committee on Commerce, Science, and Transportation
(none)	SSCM01	Subcommittee on Aviation Operations, Safety, and Security
(none)	SSCM20	Subcommittee on Consumer Protection, Product Safety, and Insurance
(none)	SSCM22	Subcommittee on Oceans, Atmosphere, Fisheries, and Coast Guard
(none)	SSCM24	Subcommittee on Science and Space
(none)	SSCM25	Subcommittee on Surface Transportation and Merchant Marine Infrastructure, Safety, and Security
(none)	SSCM26	Subcommittee on Communications, Technology, and the Internet
(none)	SSCM27	Subcommittee on Competitiveness, Innovation, and Export Promotion
SSEG00	SSEG00	Senate Committee on Energy and Natural Resources
(none)	SSEG01	Subcommittee on Energy
(none)	SSEG03	Subcommittee on Public Lands and Forests
(none)	SSEG04	Subcommittee on National Parks
(none)	SSEG07	Subcommittee on Water and Power
SSFI00	SSFI00	Senate Committee on Finance
(none)	SSFI02	Subcommittee on Social Security, Pensions and Family Policy
(none)	SSFI10	Subcommittee on Health Care
(none)	SSFI11	Subcommittee on Taxation and IRS Oversight
(none)	SSFI12	Subcommittee on Energy, Natural Resources, and Infrastructure
(none)	SSFI13	Subcommittee on International Trade, Customs, and Global Competitiveness
(none)	SSFI14	Subcommittee on Fiscal Responsibility and Economic Growth
SSFR00	SSFR00	Committee on Foreign Relations of the Senate
(none)	SSFR01	Subcommittee on European Affairs
(none)	SSFR02	Subcommittee on East Asian and Pacific Affairs
(none)	SSFR06	Subcommittee on Western Hemisphere, Peace Corps and Global Narcotics Affairs
(none)	SSFR07	Subcommittee on Near Eastern and South and Central Asian Affairs
(none)	SSFR09	Subcommittee on African Affairs
(none)	SSFR12	Subcommittee on International Development and Foreign Assistance, Economic Affairs and International Environmental Protection
(none)	SSFR13	Subcommittee on International Operations and Organizations, Human Rights, Democracy and Global Women's Issues
SSGA00	SSGA00	Committee on Homeland Security and Governmental Affairs of the Senate
(none)	SSGA01	Permanent Subcommittee on Investigations
(none)	SSGA03	Subcommittee on Oversight of Government Management, the Federal Workforce, and the District of Columbia
(none)	SSGA09	Subcommittee on Federal Financial Management, Government Information, Federal Services, and International Security
(none)	SSGA13	Ad Hoc Subcommittee on Contracting Oversight
(none)	SSGA14	Ad Hoc Subcommittee on Disaster Recovery and Intergovernmental Affairs
SSHR00	SSHR00	Senate Committee on Health, Education, Labor, and Pensions
(none)	SSHR09	Subcommittee on Children and Families
(none)	SSHR11	Subcommittee on Employment and Workplace Safety
(none)	SSHR12	Subcommittee on Primary Health and Aging

SLIA00	SLIA00	Senate Committee on Indian Affairs
SSJU00	SSJU00	Committee on the Judiciary
(none)	SSJU01	Subcommittee on Antitrust, Competition Policy and Consumer Rights
(none)	SSJU02	Subcommittee on Administrative Oversight and the Courts
(none)	SSJU04	Subcommittee on Immigration, Refugees and Border Security
(none)	SSJU21	Subcommittee on the Constitution, Civil Rights and Human Rights
(none)	SSJU22	Subcommittee on Crime and Terrorism
(none)	SSJU23	Subcommittee on Privacy, Technology and the Law
SSRA00	SSRA00	Senate Committee on Rules and Administration
SSSB00	SSSB00	Senate Committee on Small Business and Entrepreneurship
SSVA00	SSVA00	Senate Committee on Veterans' Affairs
SSEV00	SSEV00	Senate Committee on Environment and Public Works
(none)	SSEV08	Subcommittee on Transportation and Infrastructure
(none)	SSEV09	Subcommittee on Superfund, Toxics and Environmental Health
(none)	SSEV10	Subcommittee on Clean Air and Nuclear Safety
(none)	SSEV15	Subcommittee on Water and Wildlife
(none)	SSEV16	Subcommittee on Green Jobs and the New Economy
(none)	SSEV17	Subcommittee on Children's Health and Environmental Responsibility
(none)	SSEV18	Subcommittee on Oversight
(none)	JSEC00	Joint Economic Committee
(none)	JSPR00	Joint Committee on Printing
(none)	JSLC00	Joint Committee on the Library
(none)	JSTX00	Joint Committee on Taxation
(none)	JSDF00	Joint Select Committee on Deficit Reduction

**Table 2: <act-name parsable-cite=" "> values**

Abbreviation	Common Name
ADA75	Age Discrimination Act of 1975
AEA54	Atomic Energy Act of 1954
AECA	Arms Export Control Act
BBA97	Balanced Budget Act of 1997
CA34	Communications Act of 1934
CAA	Clean Air Act
CAMA72	Coastal Zone Management Act of 1972
CAPTA	Child Abuse Prevention and Treatment Act
CCDBG90	Child Care and Development Block Grant Act of 1990
CCPA	Consumer Credit Protection Act
CERCL	Comprehensive Environmental Response, Compensation, and Liability Act of 1980
CFRDA	Consolidated Farm and Rural Development Act
CLAY	Clayton Act
CNA66	Child Nutrition Act of 1966
CNCA	Columbia Nonprofit Corporation Act

COMEX	Commodity Exchange Act
CPSA	Consumer Product Safety Act
CRA64	Civil Rights Act of 1964
CSA	Controlled Substances Act
DDA	Developmental Disabilities Assistance and Bill of Rights Act
EAA79	Export Administration Act of 1979
EFTA	Electronic Fund Transfer Act
EIBA45	Export-Import Bank Act of 1945
ERISA	Employee Retirement Income Security Act of 1974
ESA	Endangered Species Act of 1973
ESEA	Elementary and Secondary Education Act of 1965
FAA61	Foreign Assistance Act of 1961
FACA	Federal Advisory Committee Act
FCA71	Farm Credit Act of 1971
FCRA	Fair Credit Reporting Act
FCRA90	Federal Credit Reform Act of 1990
FCUA	Federal Credit Union Act
FDIA	Federal Deposit Insurance Act
FECA	Federal Election Campaign Act of 1971
FFDCA	Federal Food, Drug, and Cosmetic Act
FHA	Fair Housing Act
FICA	Federal Insurance Contributions Act
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FLPMA	Federal Land Policy and Management Act of 1976
FPASA	Federal Property and Administrative Services Act of 1949
FPRA93**	Government Performance and Results Act of 1993
FRA	Federal Reserve Act
FRAA	Foreign Relations Authorization Act
FRRRP	Forest and Rangeland Renewable Resources Planning Act of 1974
FSA77*	Food and Nutrition Act of 2008
FSA80	Foreign Service Act of 1980
FTCA	Federal Trade Commission Act
FWPCA	Federal Water Pollution Control Act
HA49	Housing Act of 1949
HCDA	Housing and Community Development Act of 1974
HEA65	Higher Education Act of 1965
HIPAA	Health Insurance Portability and Accountability Act of 1996
HSA	Head Start Act
IBA78	International Banking Act of 1978
IHCIA	Indian Health Care Improvement Act
INA	Immigration and Nationality Act
ISDA	Indian Self-Determination and Education Assistance Act

ISDEAA	Indian Self-Determination Act
JTPA	Job Training Partnership Act
LWCFA	Land and Water Conservation Fund Act of 1965
MLA	Mineral Leasing Act
NEPA69	National Environmental Policy Act of 1969
NFA	National Firearms Act
NHA	National Housing Act
NSLA	National School Lunch Act
NWPA	Nuclear Waste Policy Act of 1982
OAA65	Older Americans Act of 1965
OBRA93	Omnibus Budget Reconciliation Act of 1993
OCCSS	Omnibus Crime Control and Safe Streets Act of 1968
OTCA	Omnibus Trade and Competitiveness Act of 1988
PHSA	Public Health Service Act
REH	Rehabilitation Act of 1973
RFPA78	Right to Financial Privacy Act of 1978
RHYA	Runaway and Homeless Youth Act
RRTA	Railroad Retirement Tax Act
SBA	Small Business Act
SBIA	Small Business Investment Act of 1958
SDBAA	State Department Basic Authorities Act of 1956
SEA34	Securities Exchange Act of 1934
SHERMAN	Sherman Act
SSA	Social Security Act
SWDA	Solid Waste Disposal Act
TA30	Tariff Act of 1930
TA74	Trade Act of 1974
USHA	United States Housing Act of 1937
VCCLEA94	Violent Crime Control and Law Enforcement Act of 1994
VRA65	Voting Rights Act of 1965

\* The Food and Nutrition Act of 2008 is an amendment to the Food Stamp Act of 1977. The House XML looks like `<act-name parsable-cite="FSA77">Food and Nutrition Act of 2008</act-name>`

\*\* The common abbreviation for the Government Performance and Results Act of 1993 is GPRA. Regardless, the House XML looks like `<act-name parsable-cite="FPRA93">Government Performance and Results Act of 1993</act-name>`