



# XML STANDARDS FOR ARCHIVING LEGISLATIVE RECORDS

---

**NCSL LEGISLATIVE SUMMIT 2010**

DANIEL DODGE  
JULY 28, 2010



**THOMSON REUTERS**

# TODAY'S OVERVIEW

---

- Proposed XML Standard
  - A Working Group developed a proposed XML standard to support archiving legislative document.
- Proof-of-Concept Website
  - The Minnesota Revisor's Office developed a website to test the proposed XML standard.

# PROGRAM GOALS

---

- Program: NDIIPP (National Digital Information Infrastructure & Preservation Program)
  - Mission: The mission of the National Digital Information Infrastructure and Preservation Program is to develop a national strategy to collect, preserve and make available significant digital content, especially information that is created in digital form only, for current and future generations.
  - Website: <http://www.digitalpreservation.gov>
- Sponsor: Library of Congress

# PROJECT GOALS

---

- Project: A Model Technological and Social Architecture for the Preservation of State Government Digital Information
  - Mission: Explore methods to provide enhanced online access to legislative materials in digital form.
  - Website:  
<http://www.mnhs.org/preserve/records/legislative records>



# PROJECT ORGANIZATION

---

- Leader: Minnesota Historical Society
- Partners:
  - Minnesota Revisor's Office
  - Minnesota Legislative Reference Library
  - California Digital Library
  - California's State Library and State Archives
  - California Legislative Counsel
  - National Conference of State Legislatures
- Other participating states include Arkansas, Illinois, Kansas, Mississippi, Nebraska, North Dakota, Tennessee, and Vermont.



# XML SCHEMA WORKING GROUP

---

- Goal: Develop a data model to support archiving digital legislative artifacts

<http://www.mnhs.org/preserve/records/legislativerecords/xml1.htm>

- Members:

- Minnesota Office of Revisor of Statutes
- XMaLpha Technologies
- Thomson Reuters

- Started in June 2008



# REQUIREMENTS FOR SOURCE XML

---

- *Initial tentative goal:*

Develop schema that covers all requirements for all participating jurisdictions. (This also implies every jurisdiction would convert their source into the common format for archiving.)

- *Revised goal:*

Focus on archival goals: Develop schema to wrap "native" XML from jurisdiction

- no transformation to write
- no loss of semantic markup
- no mapping of structures to common structure



# SCHEMA GOALS

---

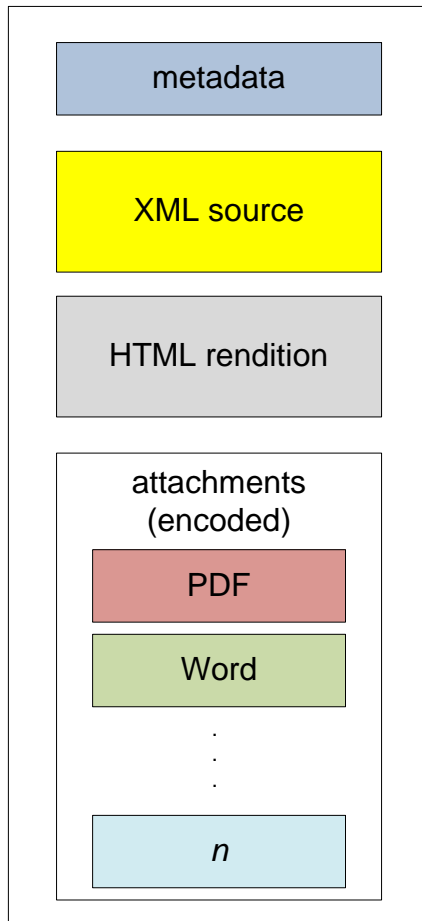
- SUPPORT STANDARD METADATA
- TRANSPORT DATA
  - PROVIDER XML
  - HTML RENDERING
  - BINARY ATTACHMENTS
- SUPPORT LONG-TERM ARCHIVAL AND RETRIEVAL





# BASIC DOCUMENT STRUCTURE

---



← metadata essential to interchange recipients

← XML source in provider's native syntax

← HTML rendition of document being interchanged

← one or more attachments for other renderings  
(encoded as Base64 or other plain-text format)

# SCHEMA CORE

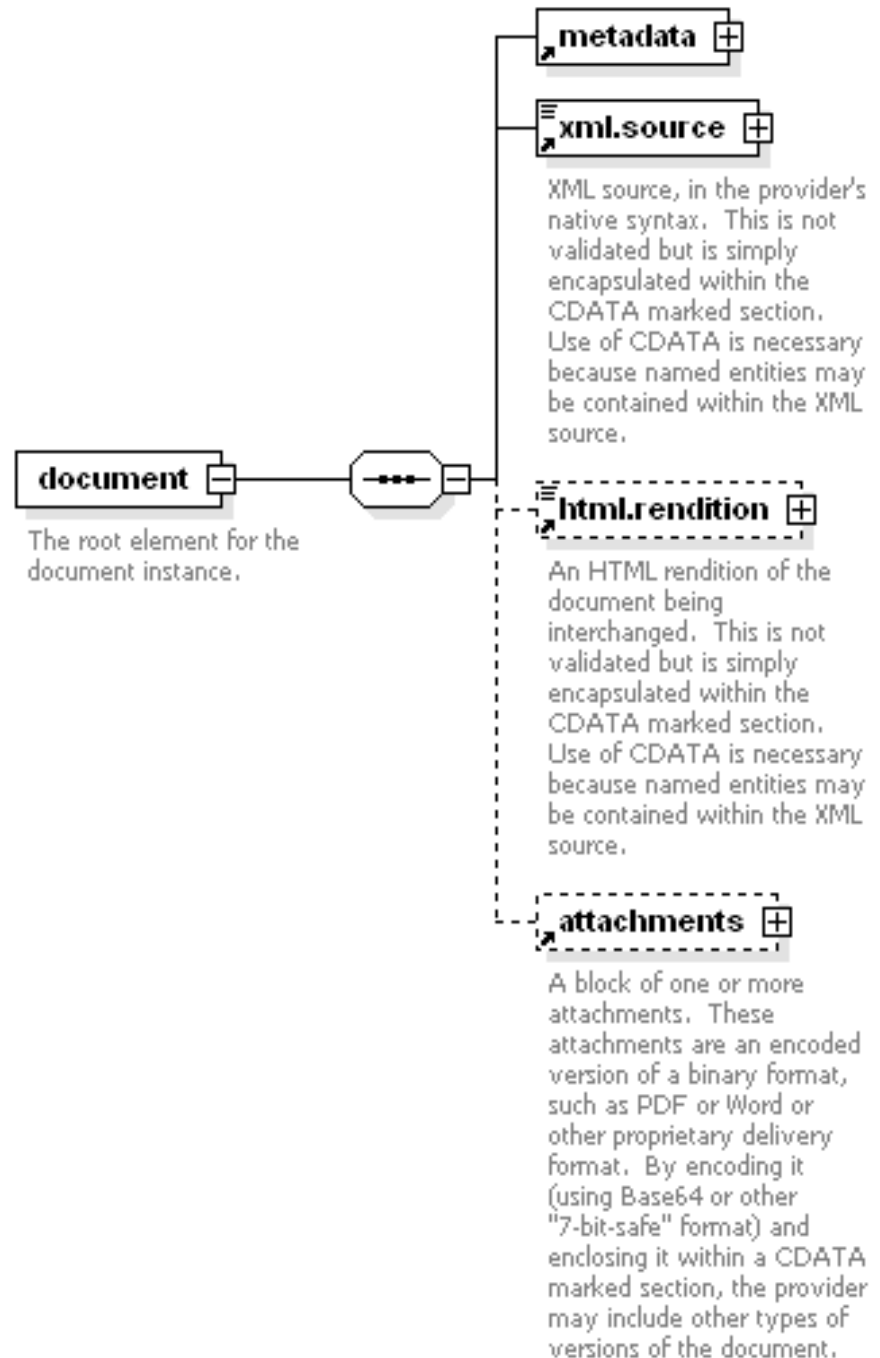
## Main elements:

<metadata>

<xml.source>

<html.rendition>

<attachments>



# STANDARD METADATA

---

- Identifier
- Title
- Type
- Jurisdiction
- Agent
- Date
- Session
- Description
- Subject
- Relation
- Governor Action
- Management History
- Rights



# REQUIREMENTS FOR SOURCE XML

---

- Preserve original text content
- Preserve original markup



# REQUIREMENTS FOR HTML

---

- Allow optional HTML rendition to be included
  - HTML
  - XHTML
- Data with and without named entities (e.g. &sect;)



# REQUIREMENTS FOR ATTACHMENTS

---

- Encode binary file to allow to be included
- Format types (very extensible)
  - Microsoft Word (application/msword)
  - PDF (application/pdf)
  - compressed (application/zip)
  - Postscript (application/eps)
  - RTF (text/rtf)
  - Text (text/plain)
  - Image (image/jpg)



# METADATA EXAMPLE 1

---

```
<meta.identifier.block>
  <meta.identifier>200800SF2377-1</meta.identifier>
  <!--          YYYYssBT####-E
  where:
  YYYY      = year
  ss        = session number.  0 is regular session
  B         = body.  H(ouse) S(enate)
  T         = type.  F(ile)  R(esolution)
  ####     = document number
  E         = engrossment/version number
  -->
</meta.identifier.block>

<meta.title.block>
  <meta.title>SF2377-1</meta.title>
</meta.title.block>

<meta.type.block>
  <meta.type>bill</meta.type>
</meta.type.block>
```



# METADATA EXAMPLE 2

---

```
<meta.subject.block>
  <meta.subject>Emergency and 911 Services</meta.subject>
  <meta.subject>Health and Health Department</meta.subject>
  <meta.subject>Hospitals and Health Facilities</meta.subject>
  <meta.subject>Boards</meta.subject>
</meta.subject.block>

<meta.relation.block>
  <meta.relation>
    <meta.version.of>HF1111</meta.version.of>
    <!-- 'version of' another document -->
    <meta.ingrossment>1</meta.ingrossment>
    <meta.companion.bill>HF2591</meta.companion.bill>
    <meta.chapter>2008 0 22</meta.chapter>
    <!-- Session Law: Year Session_Number Chapter -->
  </meta.relation>
</meta.relation.block>
```





# SOURCE XML EXAMPLE

---

**<![CDATA[**

```
<bill engrossed = "y" type = "bill" xmlns:xsi =  
  "http://www.w3.org/2001/XMLSchema-instance" xmlns:atipl =  
  "http://www.arbortext.com/namespace/PageLayout"  
  xsi:noNamespaceSchemaLocation = "bill.xsd">  
  . . .  
  <bttitle>  
    <bttitle_prolog>A bill for an act</bttitle_prolog>  
    <bttitle_summary>relating to health; modifying basic life  
support ambulance staffing requirements under certain  
circumstances; allowing a hardship waiver of advanced life  
support ambulance staffing requirements; modifying advanced life  
support ambulance staffing requirements under certain  
circumstances;</bttitle_summary>  
    <bttitle_action>amending Minnesota Statutes 2006, section  
144E.101, subdivision 7; Minnesota Statutes 2007 Supplement,  
section 144E.101, subdivision 6.</bttitle_action>  
  </bttitle>  
  . . .  
</bill>
```

**]]>**



# EXAMPLE DATA FOR ATTACHMENT

---

<attachments>

```
<binary ID="thing1" encoding="Base64"
      source="application/pdf">
  <![CDATA[
    JVBERi0xLjQKJeTjz9IKMSAwIG9iagpbL1BERi9JbWFnZUIvSW1
    hZ2VDL01tYWdlSS9UZXBh0XQplbmRv . . . DEzOTUKJSVFT0YK
  ]]>
</binary>
```

```
<binary ID="thing2" ... ><![CDATA[ ... ]]></binary>
```

</attachments>



# EXAMPLE ATTACHMENT RENDERING

S2377-1.pdf - Adobe Reader

File Edit View Document Tools Window Help

1 / 3 75%

Find

1.1 A bill for an act  
1.2 relating to health; modifying basic life support ambulance staffing requirements  
1.3 under certain circumstances; allowing a hardship waiver of advanced life support  
1.4 ambulance staffing requirements; modifying advanced life support ambulance  
1.5 staffing requirements under certain circumstances; amending Minnesota Statutes  
1.6 2006, section 144E.101, subdivision 7; Minnesota Statutes 2007 Supplement,  
1.7 section 144E.101, subdivision 6.

1.8 BE IT ENACTED BY THE LEGISLATURE OF THE STATE OF MINNESOTA:

1.9 Section 1. Minnesota Statutes 2007 Supplement, section 144E.101, subdivision 6,  
1.10 is amended to read:

1.11 Subd. 6. **Basic life support.** (a) Except as provided in ~~paragraph~~ paragraphs  
1.12 (e) and (f), a basic life support ambulance shall be staffed by at least two ambulance  
1.13 ~~service personnel, at least one of which must be an EMT, who~~ EMTs, one of whom must  
1.14 accompany the patient and provide a level of care so as to ensure that:

1.15 (1) life-threatening situations and potentially serious injuries are recognized;  
1.16 (2) patients are protected from additional hazards;  
1.17 (3) basic treatment to reduce the seriousness of emergency situations is administered;

# PROOF OF CONCEPT

---

- Minnesota Revisor's Office

- Tim Orr
- Isaac Holmlund

- Online on November 11, 2009

- Documentation:

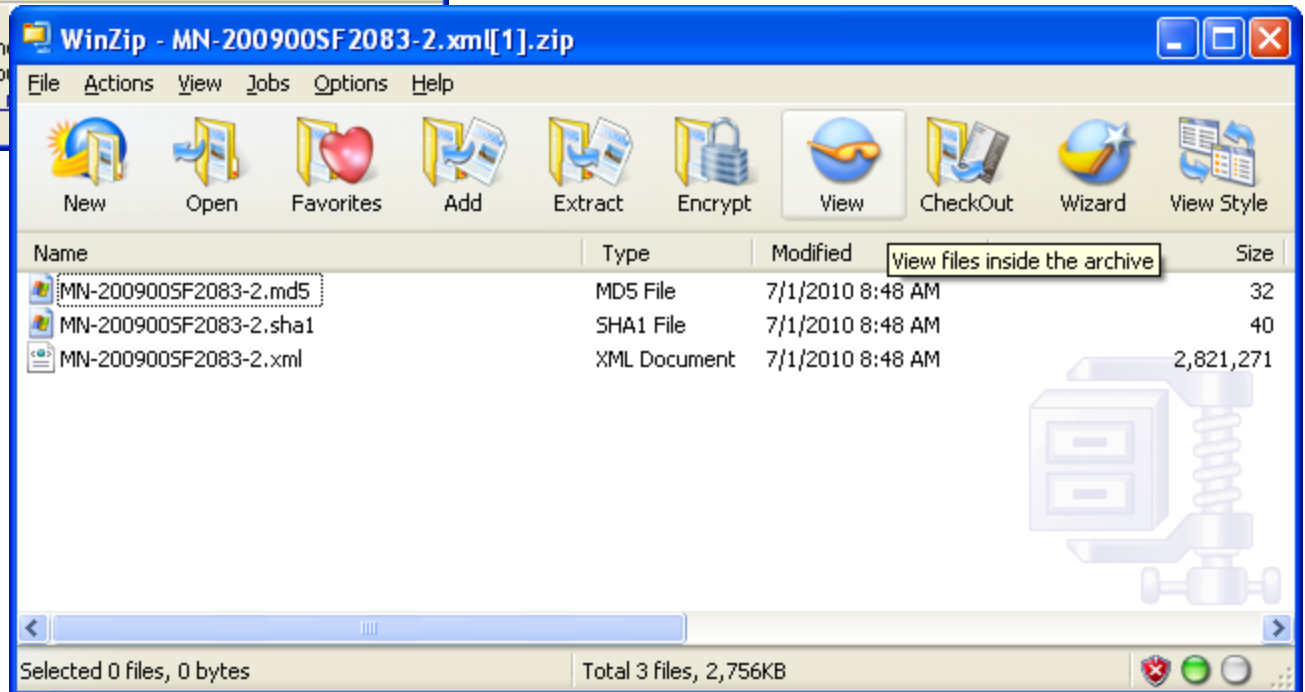
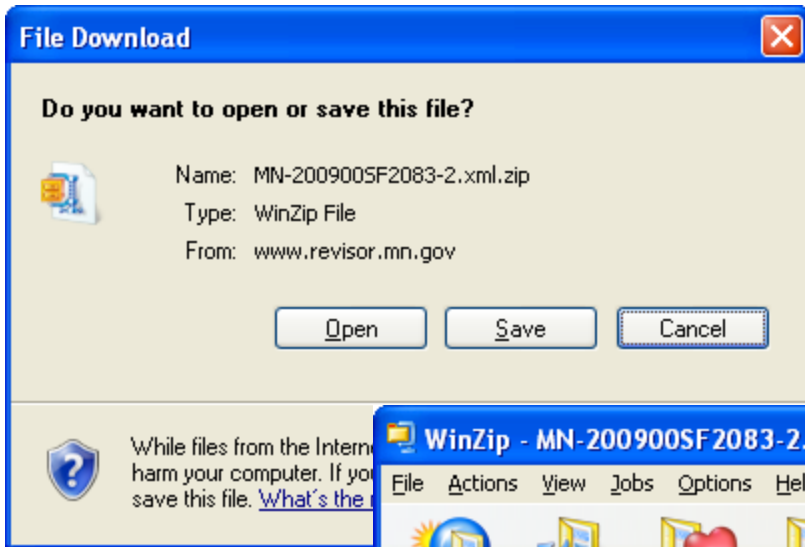
<https://www.revisor.mn.gov/beta/mhs/xmlwrapper.php>

- Interface:

[xmlwrapper.php?year=2009&sess=0&body=Senate&numb=2083&enr=2&formats=xml,html,pdf](https://www.revisor.mn.gov/beta/mhs/xmlwrapper.php?year=2009&sess=0&body=Senate&numb=2083&enr=2&formats=xml,html,pdf)



# XML FILE DOWNLOAD



## NEXT STEPS

---

- Complete additional features for proof-of-concept with MN Revisor's Office
  - Secure interface for aggregation proof-of-concept
  - Centralized archival repository
  - Authentication/verification of transmitted data
- Engage other states for next proof-of-concept



# POSSIBILITIES

---

- Toolkit for jump-starting website interfaces for additional states
- Toolkit for viewing contents of archivable package
- Proof-of-Concept for an archive repository (centralized aggregation process)
- Legislative content XML schema template (for easier adoption of XML by states)



# QUESTIONS?

---

Email:

[daniel.dodge@thomsonreuters.com](mailto:daniel.dodge@thomsonreuters.com)

References:

<http://www.digitalpreservation.gov>

<http://www.mnhs.org/preserve/records/legislativerecords>

<http://www.mnhs.org/preserve/records/legislativerecords/xml1.htm>