NISO Z39.96-2015 JATS Journal Article Tag Suite

Some excruciating detail, a bit of history, and a “Help Wanted”

Jeffrey Beck
NCBI/NLM/NIH
beck@ncbi.nlm.nih.gov
What is JATS?

JATS is a NISO Standard (Z39.96-2015) that defines XML elements and attributes and three article models built from them.

XML is eXtensible Markup Language

- Became a W3C Recommendation in 1999
- Originally for “documents on the web”
- Optimized for machine processing of documents
DTD is Document Type Definition

- One of many (3 really) schema languages for defining XML documents
- Essentially a set of rules for what can be in your document, what must be in your document, and the order of things if you wish to enforce order
Important Memo

To: All Staff
From: The Boss
Date: May 17, 2016
Re: Refrigerator

Blah blah blah rotten food blah blah blah clean up after yourself blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah.

blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah you will be fired.
<memo type="important">
  <to>All Staff</to>
  <from>The Boss</from>
  <re>Refrigerator</re>
  <date day="17" month="05" year="2016"/>
  <p>Blah blah blah rotten food blah blah blah clean up after yourself blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah blah you will be fired. </p>
</memo>
<schema
xmlns:xs="http://www.w3.org/2001/XMLSchema"

elementFormDefault="qualified">
  <element name="memo">
    <complexType>
      <sequence>
        <element ref="to"/>
        <element ref="from"/>
        <element ref="re"/>
        <element ref="date"/>
        <element maxOccurs="unbounded" ref="p"/>
      </sequence>
      <attributeGroup ref="attlist.memo"/>
    </complexType>
  </element>
  <attributeGroup name="attlist.memo">
    <attribute name="type"/>
  </attributeGroup>
  <element name="to" type="xs:string"/>
  <element name="from" type="xs:string"/>
  <element name="re" type="xs:string"/>
  <element name="date">
    <complexType>
      <attributeGroup ref="attlist.date"/>
    </complexType>
  </element>
  <attributeGroup name="attlist.date">
    <attribute name="day"/>
    <attribute name="month"/>
    <attribute name="year"/>
  </attributeGroup>
  <element name="p" type="xs:string"/>
</schema>

memo = element memo { attlist.memo, to, from, re, date, p+ }
attlist.memo &*= attribute type { text }?
to = element to { attlist.to, text }
from = element from { attlist.from, text }
re = element re { attlist.re, text }
date = element date { attlist.date, empty }
attlist.date &=
  attribute day { text }?,
  attribute month { text }?,
  attribute year { text }?
p = element p { attlist.p, text }
What is JATS?

JATS is a NISO Standard (Z39.96-2015) that defines XML elements and attributes and three article models derived from them.
Where did it come from?

JATS is a continuation of the NLM DTDs (2003-2012) which were created in a collaboration between NLM (the Pubmed Central group) and the Harvard E-Journal Archiving project.
• The Melon Foundation funded the Harvard Library to study the feasibility of using one DTD for archiving journal articles.

• Harvard commissioned Inera, Inc. for the E-Journal Archive DTD Feasibility Study.
  • Conclusion – yes, it is feasible, but the right DTD does not exist.

• A meeting was held in April 2002 to discuss the changes needed to the PMC2 DTD to expand its range to include most any journal. Attendees included PMC, Mulberry Technologies, Inc. (consultant to PMC), The Mellon Foundation, The Harvard Library, and Inera (consultant to Harvard-Mellon).
The committee decided that this would be a good time to clean up those little things that we know were problems but we hadn’t fixed because we wanted all of the new models to be backward-compatible.
Backward-compatibility

Means that all existing XML instances will be valid according to the new model.

Mostly we had minor housekeeping issues that we had been putting off.

In version 1.0, the @id on <list-item> was defined as CDATA (when it obviously should have been defined as ID to allow ID/IDREF functionality).

So, any existing <list-item id="45qrt"> would be valid under version 1.0 but not valid when the attribute was properly defined as type=ID.
**NLM DTDs**

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>v 1.0</td>
<td>March 2003</td>
</tr>
<tr>
<td>v 1.1</td>
<td>November 2003</td>
</tr>
<tr>
<td>v 2.0</td>
<td>November 2004</td>
</tr>
<tr>
<td>v 2.1</td>
<td>September 2005</td>
</tr>
<tr>
<td>v 2.2</td>
<td>June 2006</td>
</tr>
<tr>
<td>v 2.3</td>
<td>March 2007</td>
</tr>
<tr>
<td>v 3.0</td>
<td>November 2008</td>
</tr>
<tr>
<td>v 3.1</td>
<td></td>
</tr>
</tbody>
</table>

- **Backward-incompatible release**
- NLM DTD Working Group is dissolved, and the NISO Journal Article Tag Suite Working Group is created.
<table>
<thead>
<tr>
<th>NLM DTDs v 1.0</th>
<th>NLM DTDs v 1.1</th>
<th>NLM DTDs v 2.0</th>
<th>NLM DTDs v 2.1</th>
<th>NLM DTDs v 2.2</th>
<th>NLM DTDs v 2.3</th>
<th>NLM DTDs v 3.0</th>
<th>NISO Z39.96 JATS v 0.4</th>
</tr>
</thead>
</table>

Backward-incompatible release
Backward-incompatible release

NLM DTDs v 1.0
March 2003

NLM DTDs v 1.1
November 2003

NLM DTDs v 2.0
November 2004

NLM DTDs v 2.1
September 2005

NLM DTDs v 2.2
June 2006

NLM DTDs v 2.3
March 2007

NLM DTDs v 3.0
November 2008

NISO Z39.96 JATS v 0.4
March 2011

NISO Z39.96-2012 is official
August 2012
<table>
<thead>
<tr>
<th>NLM DTDs</th>
<th>NISO Z39.96</th>
<th>JATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>v 1.0</td>
<td>v 0.4</td>
<td>v 1.1</td>
</tr>
<tr>
<td>March 2003</td>
<td>November 2008</td>
<td>December 2015</td>
</tr>
<tr>
<td>v 1.1</td>
<td>v 1.1d1</td>
<td></td>
</tr>
<tr>
<td>November 2004</td>
<td>August 2012</td>
<td></td>
</tr>
<tr>
<td>v 2.0</td>
<td>v 1.1d2</td>
<td></td>
</tr>
<tr>
<td>November 2005</td>
<td>December 2014</td>
<td></td>
</tr>
<tr>
<td>v 2.1</td>
<td>v 1.1d3</td>
<td></td>
</tr>
<tr>
<td>September 2005</td>
<td>April 2015</td>
<td></td>
</tr>
<tr>
<td>v 2.2</td>
<td>v 1.1</td>
<td></td>
</tr>
<tr>
<td>June 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>v 2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 2007</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Backward-incompatible release

NISO Z39.96-2012 is official

NISO Z39.96-2015 is official
I know what you are thinking

Jeff, are there any advantages to using a standard and well used article model for my content?
As a matter of fact …

- Ease of article interchange
- Common Tools – from jats.nlm.nih.gov and vendors
- $$$ - Creating and maintaining a custom article model is expensive
- Simplify reuse of your content *
- Tagging vendors know the models **
- There is a community of users
  - JATS wiki - http://jatswiki.org/wiki/Main_Page
http://jats4r.org

JATS for Reuse is a group of publishers, tagging vendors, and aggregators that define best practices for tagging articles in JATS to ease machine-readability.

Editors are active in the group.
Tagging vendors know the models **

This can be a great advantage.

But don’t hand over the keys. Make sure you (as the content owner) have a say in how your articles are tagged.

This means that Editors need to understand their XML.
HELP WANTED

• Editorial or Library background
• Exhibited understanding of document structure and document modeling
• Experience with document transformations

The Alphabet Soup

• SGML
• XML
• XSLT
• XSL-FO
• DTD
• XSD
• RNG
• Schematron
• XPATH

NATIONAL LIBRARY OF MEDICINE
What if I want more control?

Write your publisher- or journal-specific rules in a separate validation layer.

This means that you can control your content very closely within your production systems. But when you send your files out, they are still JATS compliant. So anyone can use them without knowing your business rules.

I would suggest Schematron.

Schematron is a rule-based validation language, written in XML, that allows you to write tests about the presence or absence of patterns in XML documents.
<xml version="1.0" encoding="UTF-8"/>

<sch:schema xmlns:sch="http://purl.oclc.org/dsdl/schematron" queryBinding="xslt2"
xmlns:sqf="http://www.schematron-quickfix.com/validator/process">
  <sch:pattern id="memo">
    <sch:rule context="memo">
      <sch:assert test="count(p) > 1">ERROR: memo must have more than one paragraph.</sch:assert>
    </sch:rule>
    <sch:rule context="date">
      <sch:report test="@year &gt; 2000">ERROR: This memo format must not be used after 2000.</sch:report>
    </sch:rule>
  </sch:pattern>
</sch:schema>
<!DOCTYPE memo SYSTEM "memo.dtd">
<memo type="important">
  <to>All Staff</to>
  <from>The Boss</from>
  <re>Refrigerator</re>
  <date day="17" month="05" year="2016"/>
  <p>Blah blah blah rotten food blah blah clean up after yourself blah blah blah</p>
  <p>blah blah blah blah blah blah blah blah blah blah blah blah blah blah</p>
  <p>you will be fired.</p>
</memo>
JATS Extensions

BITS – Book Interchange Tag Suite
  NCBI project
  chapter ~ article

NISO STS – Standards Tag Suite
  In development at NISO
  standard ~ article

TaxPub – article extension for Taxonomic Treatments
Thank you

Z39.96-2015  JATS Standard document

JATS Non-Normative Supporting Information

BITS 2.0 home

JATS for Reuse
http://jats4r.org

Schematron
http://www.schematron.com/
https://en.wikipedia.org/wiki/Schematron