

Update on the NLM Journal Article Tag Suite (the NLM DTDs)

Jeffrey Beck
beck@ncbi.nlm.nih.gov

A Brief History

- Version 1 was released in December 2002 with the **Archiving and Interchange DTD** and the **Journal Publishing DTD**.
- Version 1 was based on work at NCBI to upgrade the PubMed Central DTD and a project at Harvard University funded by the Mellon Foundation to address the problems of archiving scholarly journals in electronic form (E-journals).

- The initial meeting included participants from NCBI, Harvard, and the Mellon Foundation along with NCBI's consultants, Mulberry Technologies, and Harvard's consultants, Inera, Inc.

But there was confusion about what the model should be.

Easy Target for Conversion?

- Should the new DTD be a broad, descriptive target that would be easy to translate articles from other SGML or XML models into?

A model like this would have many optional elements with few things in a prescribed order, and different ways to tag the same object.

Easy model to create content in?

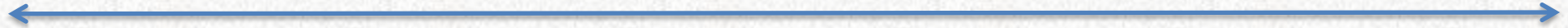
- Or should the new DTD be a narrower, prescriptive target that would give creators of new XML articles guidance about how to make a valid article?

A model like this would have more required elements with fewer choices on how to tag the same object.

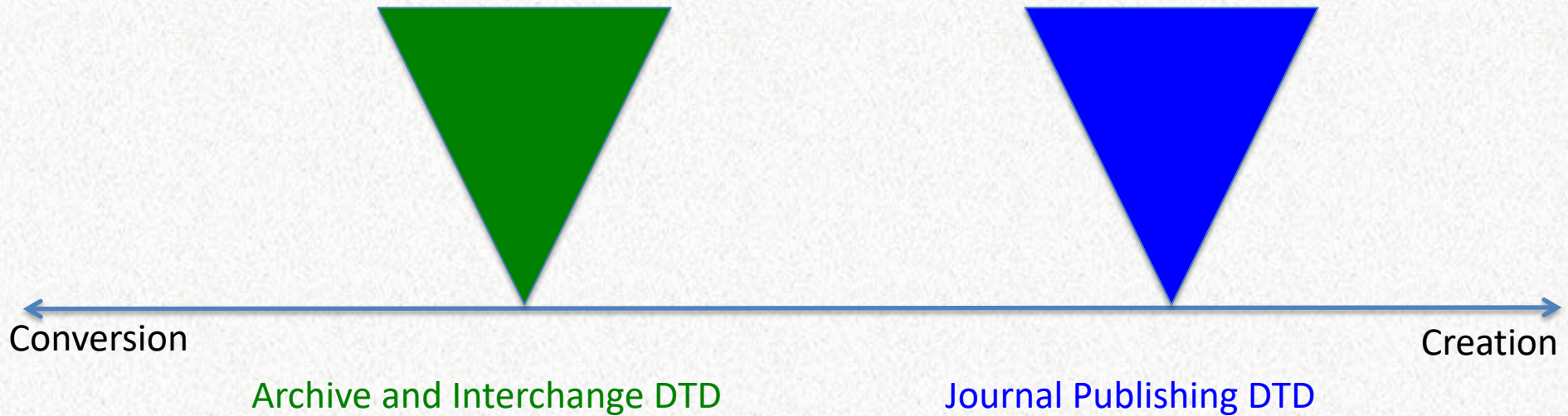
The DTD Spectrum

Optimized for Conversion to

Optimized to Create Content in



The DTD Spectrum



- Journal Archiving and Interchange Tag Set Tag Library version 3.0
- How To Use (Read Me First)
- General Introduction
- ▶ Elements
- ▶ Attributes
- ▶ Parameter Entities
- ▶ Document Hierarchy Diagrams
- ▶ Common Tagging Practice
- ▶ Implementing This Tag Set
- ▶ Change Report: Version 3.0
- DTD, XSD, and RNG Downloads
- Element Context Table
- Index
- Archiving 3.0 Home

Journal Archiving and Interchange Tag Set Tag Library version 3.0

November 2008

Digital Archive of Journal Articles
National Center for Biotechnology Information (NCBI)
National Library of Medicine (NLM)

Developer:
Mulberry Technologies, Inc.
17 West Jefferson Street, Suite 207
Rockville, MD 20850
Phone: 301/315-9631
Fax: 301/315-8285

- Journal Publishing Tag Set Tag Library version 3.0
- How To Use (Read Me First)
- General Introduction
- ▶ Elements
- ▶ Attributes
- ▶ Parameter Entities
- ▶ Document Hierarchy Diagrams
- Full Article Samples
- ▶ Common Tagging Practice
- ▶ Implementing This Tag Set
- ▶ Change Report: Version 3.0
- DTD, XSD, and RNG Downloads
- Element Context Table
- Index
- Publishing 3.0 Home

Journal Publishing Tag Set Tag Library version 3.0

November 2008

Digital Archive of Journal Articles
National Center for Biotechnology Information (NCBI)
National Library of Medicine (NLM)

Developer:
Mulberry Technologies, Inc.
17 West Jefferson Street, Suite 207
Rockville, MD 20850
Phone: 301/315-9631
Fax: 301/315-8285

Everything was fine, until

<X>

The two archiving strategies

Archiving the intellectual content of the article?

Or

Archiving the article file?


If you need to archive the entire file, you need a way to keep those items in the file that the **Archiving and Interchange DTD** did not worry about.

Punctuation in Keywords.

Keyword Group in Archiving 1.0:

```
<!ELEMENT   kwd-group   (title?, kwd+)   >
```

Keywords: DNA analysis; gene expression; parallel cloning; fluid microarray.



```
<kwd-group>  
  <kwd>DNA analysis</kwd>  
  <kwd>gene expression</kwd>  
  <kwd>parallel cloning</kwd>  
  <kwd>fluid microarray</kwd>  
</kwd-group>
```


Punctuation in Keywords.

Keyword Group in Archiving 2.0:

```
<!ELEMENT kwd-group (title?, (kwd | x)+)
>
```

Keywords: DNA analysis; gene expression; parallel cloning; fluid microarray.

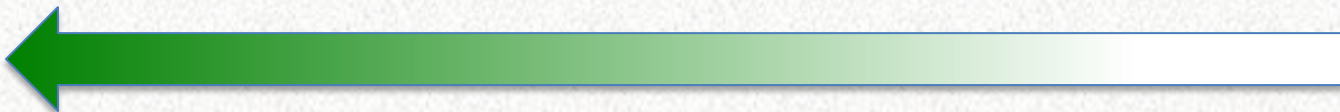
```
<kwd-group>
  <title>Keywords: </title>
  <kwd>DNA analysis</kwd><x>; </x>
  <kwd>gene expression</kwd><x>; </x>
  <kwd>parallel cloning</kwd><x>; </x>
  <kwd>fluid microarray</kwd><x>.</x>
</kwd-group>
```

The DTD Spectrum

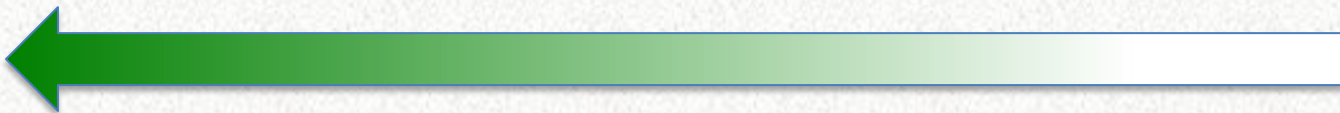


Is Journal Publishing meeting our needs?

Not really. It is too restrictive for some users,
and not prescriptive enough to be a good
Authoring model.



The DTD Spectrum



JATS?

Journal Article Tag Suite

The Tag Suite is the collection of all Elements and Attributes.

Each model (Archiving, Publishing, Authoring) is a Tag Set.

Each schema (DTD, XSD, RELAX NG) represents a model or Tag Set.

NLM DTD vs PubMed Central

PubMed Central (PMC) is a user of the NLM DTD.

The JATS Survey

- Half of the respondents said that they impose rules other than schema validation on their content.
- To the question on using the DTDs as published or modified in some way:
 - 9 as published
 - 5 subset
 - 5 superset
 - 4 informed

To the question, “What form(s) of the Tag Set are you using?”

- 18 DTD
- 4 XSD
- 1 RELAX NG

Dear Software manager, I am very upset that my tax money is supporting obsolete technology. Please translate your DTDs into XML schemas. XMLSpy can be used for this purpose. The schemas will then work with Microsoft Word 2003 and other modern software. Thank you.

-- May 31, 2003

The NISO Standard

In late 2009, a new NISO Working Group was formed to address the “Standardized Markup for Journal Articles” project.

The first task was to address the list of change requests that had accumulated since we released version 3.0 at the end of 2008.

v 3.1

- A minor, backward-compatible update.
- Increased support for multi-language documents.
- Increased support for tagging documents accessibility. Much of this will be improved and explicit documentation on how to use the accessibility elements and attributes.

Back to NISO

- The Standard will describe the elements and attributes in the JATS.
- And then it will describe each of the three article models:
 - Archiving and Interchange
 - Journal Publishing
 - Article Authoring

- The standard will not include the schema files, DTDs, XSDs, or RELAX NGs.
- These will be non-normative supporting documents - as the Tag Libraries will be.

Book Review

- Assembling a group to review the current NCBI Book Tag Set and build a Tag Set for books from the JATS.
- Content should be scoped similar to the journal work. Not trying to model every book.

New Discussion List

A new open discussion list for all things JATS.

You can read about the JATS–List at:

<http://www.mulberrytech.com/JATS/JATS-List/index.html>

and subscribe at:

<http://www.mulberrytech.com/JATS/JATS-List/subscribe-unsubscribe.html>

Or Google “jats-list”

<JATS-Con>

November 1&2, 2010

- The first User Group meeting for the NLM DTDs.
- Free conference held on NIH campus.
- 2-day meeting with 125 attendees and 13 presentations.
- Some technical, some not so technical.

Presenters from

- Inera, Inc
- The American Chemical Society (ACS)
- Mulberry Technologies
- NCBI (NLM)
- Federation of Animal Science Societies (FASS)
- Ontario Council of University Libraries (OCUL)
- Portico (ITHAKA)

- Columbia University
- The World Health Organization (WHO)
- The American Geophysical Union (AGU)
- The American Institute of Physics (AIP)

Papers on

- DTD flexibility and usage trends
- Use cases in journals, magazines, and books
- Using the free Journal Publishing 3.0 Preview Stylesheets
- Creating EPUB from JATS
- Use cases in archives
- An extension for Taxonomic information
- Accessible publishing
- Subsets vs Supersets

- Details available on the JATS-Con website
<http://dtd.nlm.nih.gov/jats-con/>
- All papers, presentations, and videos of the presentations are available in the online proceedings
<http://www.ncbi.nlm.nih.gov/books/NBK47086/>