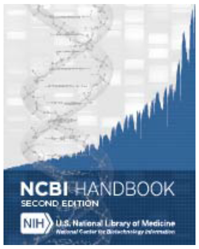
NLM Citation: Beck J, Randall L. NLM DTD to NISO JATS Z39.96-2012. 2013 Nov 14. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



NLM DTD to NISO JATS Z39.96-2012

Jeffrey Beck[1] and Laura Randall[1]

Created: November 14, 2013.

Scope

The Journal Article Tag Suite (JATS) is a description of a set of elements and attributes that is used to build XML models of journal articles for archiving, publishing, and authoring. JATS became an American National Standard (ANSI/NISO Z39.96-2012) in August 2012, but it was already a well-established specification (known by the colloquial name “NLM DTD”) by the time work began on standardization in late 2009.

Normalizing the structure of journal articles enables interchange of articles among publishers, authors, data conversion vendors, and aggregators such as archives and indexing services. An existing, well used, and freely available article model also allows new, small journal publishers to start creating articles in XML significantly faster, more easily, and at less cost than if they had to create a model and persuade their vendors and publishing partners to use it.

History

PMC DTD

PubMed Central (PMC), developed and maintained by the National Center for Biotechnology Information (NCBI), is the NLM’s digital library of full-text life sciences journal literature. The project’s mission is to make full-text article content (submitted by participating publishers) available through a public database. The only technical requirement when PMC started in 1999 was that publishers supply the articles in either SGML or XML format and include all images.

It quickly became obvious that article content needed to be normalized into a single article model on ingest to reduce the stress on the database and the software that rendered the articles on the Web. The PMC Document Type Definition (DTD), known as pmc-1.dtd, was written based on the two article models that were being submitted to PMC at the time, and its main focus was the online representation of the articles.

The original article model was built based on a small sample set, and as publishers submitted new formats for inclusion in PMC, the pmc-1.dtd grew to handle new article structures. This approach did not scale, so NCBI contacted Mulberry Technologies, Inc., in Rockville, Maryland, to perform an independent review of the pmc-1.dtd and to work on a replacement model.

Author Affiliation: 1 NCBI; Email: beck@ncbi.nlm.nih.gov; Email: lrandall@ncbi.nlm.nih.gov.

Universal DTD for Electronic Journals

In 2001, the Harvard University Library E-Journal Archiving Project (using funds from the Mellon Foundation) commissioned a study into the feasibility of having one DTD that could be used to archive all electronic journals (6). The report prepared by Inera, Inc., Belmont, Massachusetts, was a survey of the journal article DTDs from PubMed Central and the following publishers:

- American Institute of Physics
- BioOne
- Blackwell Science
- Elsevier Science
- Highwire Press
- Institute of Electrical and Electronics Engineers
- Nature Publishing Group
- University of Chicago Press
- John Wiley & Sons

The report concluded that there could be a single DTD that could accommodate any electronic journal article, but none of the existing DTDs in the study met all of the requirements.

At this point, the modification of the pmc-1.dtd was well under way. Many of the suggestions from the study were incorporated into the modified PMC article model. When the modified model was shared with Bruce Rosenblum from Inera, he determined that the pmc-2.dtd was almost the one model that they had been looking for during the feasibility study.

A meeting was held in the spring of 2002 at the NLM that included representatives of NCBI/NLM, the Harvard Library, the Mellon Foundation, Mulberry Technologies, and Inera to try to work out the details of adopting the new pmc-2.dtd to general use for archiving any electronic journal article.

At this meeting it was decided that:

1. The project would be a set of “standard” XML elements and attributes that could be used to build article models.
2. Work should continue on the new models to expand them to handle any journal article content, including a survey of articles across many disciplines, to ensure that all article objects could be accommodated in the new model.
3. There should be two initial article models: one for existing content, providing a broad target for conversion of any article content, and one for creating new content, having a more prescriptive model to provide explicit rules for tagging content.
4. The new models should be easily extensible. For example, it should be easy to swap the OASIS CALS (**C**ontinuous **A**cquisition and **L**ife-cycle **S**upport) table model for the default XHTML table model.

The NLM DTDs

The National Library of Medicine (NLM) DTDs were created based on that initial meeting. Version 1 of the NLM Archiving and Interchange Tag Suite was released in early 2003 and included two article tag sets: the Archiving and Interchange model and the Journal Publishing model. The Archiving model was intended for tagging existing content, and the more prescriptive Publishing model was intended for authoring and tagging new article content (or article content that would be marked up in XML for the first time).

The intention of the NLM DTD project was to enable what publishers are already doing with their content rather than to define what they should be doing. In order to keep the suite relevant, because publishing practices are not static, the NLM assembled the Archiving and Interchange Tag Suite Working Group—a group of individuals who advised the NLM or recommended changes to the suite. The Working Group, responding to public feedback and, drawing from their experience, released several updated versions of the Tag Suite and the individual models over the next several years.

In 2005, with the release of the Tag Suite version 2.1, a new article model was introduced: the Article Authoring model. Between versions 1.0 and 2.0, the modifications to the Tag Sets had made models far more permissive, and the Working Group realized the Journal Publishing set was no longer suited for authoring new content. The Article Authoring model is the most prescriptive of the sets and is targeted toward new content creation.

Backward-compatibility is a significant factor in adoption of a new version of any tag set, so to facilitate the adoption of updated versions, the Working Group tabled all non-backward-compatible changes through the version 2.3 release. Concurrent with the release of version 2.3, the Working Group made the decision that the next major version release, version 3.0, would incorporate all of the non-backward-compatible changes that had been accumulating.

Involvement of NISO

The decision to make version 3.0 non-backward-compatible was part of the discussion about formalizing the Tag Suite with the National Information Standards Organization (NISO). The original plan had been to submit the latest version of the suite and models for registration, but because standardization would bring a lot of attention and new users to the suite, the Working Group chose to make the non-backward-compatible changes prior to registration.

Once version 3.0 was released in November 2008, the work of the NLM Archiving and Interchange Tag Suite Working Group concluded and the NISO Standardized Markup for Journal Articles Working Group was created. Like the NLM Working Group before it, the NISO Working Group saw its role as normalizing and documented existing practices rather than dictating what should be done.

On March 30, 2011, after approval by the NISO Standardized Markup for Journal Articles Working Group and the NISO Content and Collection Management Topic Committee that oversaw the Working Group, NISO released NISO Z39.96, JATS: Journal Article Tag Suite, as a Draft Standard for Trial Use. Officially, this was NISO JATS version 0.4, but in essence it was a minor update to the NLM version 3.0 Tag Suite and article models. The draft standard was available for public comment until September 30, 2011.

The Working Group responded to each of the comments received and created JATS version 1.0, which was approved by NISO voting members and the American National Standards Institute as ANSI/NISO Z39.96-2012 in August 2012.

The Standard and the Supporting Information

ANSI/NISO Z39.96-2012 defines elements and attributes that describe metadata and full content of scholarly journal articles. It is not designed to describe magazines, books, or other publishing formats that may have some similar structures to journal articles but could also have significantly different structures.

The Tag Suite is the complete set of elements and attributes described in the standard. Along with these descriptions the standard includes three article models, or Tag Sets:

- The Journal Archive and Interchange Tag Set
- The Journal Publishing Tag Set
- The Article Authoring Tag Set

The Tag Suite has been designed to be extensible. Any of the tag sets may be extended or restricted to meet the needs of a given project. Also, new tag sets can be built from the elements and attributes in the Tag Suite and should be considered conforming to the standard.

Non-normative Information

The standard includes neither schemas nor much usage information. Non-normative supporting information, available from the NLM site, includes:

1. Schemas for each of the Tag Sets described above in three schema languages: DTD, W3C Schema (XSD), and RELAX NG.
2. Detailed “Tag Libraries” for each Tag Set that include the element and attribute definitions from the standard, remarks on usage, tagged examples, and detailed discussions of topics ranging from customizing a tag set to tagging names and dates.
3. A basic set of style sheets for rendering articles in HTML or in PDF through XSL-FO. These style sheets are intended as “starters” to be modified and personalized by each user.

Additional Schemas

The article models for NLM version 1.0 in 2003 were released only as DTDs. Beginning with version 1.1, WC3 Schema expressions of the Tag Sets were released along with the DTDs, and RelaxNG schema versions were added beginning with version 2.1. The additional schema languages were created from the DTD versions. Because the three languages have different features and limitations, the DTD version was declared as the version intended for maintenance and the other two as derivatives. This ensured that data tagged in one of the tag sets would be valid according to all of that Set’s schemas.

Tools

The NLM (NCBI) released tools for use with the NLM DTDs to the public. These tools include an XSL conversion to HTML for previewing NLM DTD content and an NLM DTD-to-XSL-FO conversion for creating PDFs. These basic tools are intended to be launching points for groups and it is expected that groups will customize these basic stylesheets for their own uses.

The public tools also include an XSL stylesheet that will transform data from any version prior to 3.0 into 3.0. This was released to help ease the transition to the non-backward-compatible version.

The Future of JATS

The plan with NISO is to maintain JATS continuously. Continuous maintenance is an option for American National Standards that allows comments and requests for enhancements to be submitted at any time, with a published regular schedule of when a Standing Committee will meet to evaluate such requests. When a sufficient number of substantive changes have been approved, a revision is balloted for approval and publication. (The alternative default option of periodic maintenance provides for a five-year review of the standard and, if a revision is deemed to be needed after such a review, a revision working group is initiated.) Continuous maintenance will allow revisions to be issued on a more timely basis and ensure ongoing interaction with the community that is using the standard. We look forward to working with users as the JATS grows to accommodate the needs of its growing user community.

References

JATS standard (ANSI/NISO Z39.96-2012) jats.niso.org

JATS supporting documentation http://jats.nlm.nih.gov

JATS-Con Available at: http://jats.nlm.nih.gov/jats-con/

JATS-Con Proceedings Available at: http://www.ncbi.nlm.nih.gov/books/NBK65129/

JATS E-mail List Available at: http://www.mulberrytech.com/JATS/JATS-List

Inera, Inc. E-Journal Archive DTD Feasibility Study. December 5, 2001. http://www.diglib.org/preserve/hadtdfs.pdf

NLM Archiving and Interchange DTD, version 1.0 http://dtd.nlm.nih.gov/archiving/1.0/

NLM Journal Publishing DTD, version 1.0 Available at: http://dtd.nlm.nih.gov/publishing/1.0/

OASIS CALS table model Available at: https://www.oasis-open.org/specs/tablemodels.php