

ARTICLE
EXCERPTED
FROM:

INFORMATION STANDARDS QUARTERLY

SPRING 2013 | VOL 25 | ISSUE 1 | ISSN 1041-0031

ISQ

SPECIAL EDITION

YEAR IN REVIEW AND STATE OF THE STANDARDS

—
NISO 2012 YEAR IN REVIEW

—
TC46 2012 YEAR IN REVIEW

—
JATS—A NEW STANDARD FOR
AN OLD SPECIFICATION

—
STATE OF THE STANDARDS

NISO
How the information world
CONNECTS



Jeffrey
Beck

B. Tommie
Usdin

JEFFREY BECK AND B. TOMMIE USDIN

JATS—A New Standard for an Old Specification

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, cheaper, and more easily than if they had to create a model and persuade their vendors and publishing partners to use it.

An active and supportive community of users has developed around the JATS. The JATS List is a public forum for discussion of the tag suite; JATS applications, implementations, and customizations; and JATS user questions. The list is open to everyone: users and developers, experts and novices alike. An annual user group meeting, JATS-Con has been held in the fall at the National Library of Medicine on the NIH campus in Bethesda, MD since 2010. Proceedings include the articles, presentation materials, and video of the presentations.

History

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 intent of the project was 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 some 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) was written based on the two article models that were being submitted to PMC at the time, and its main focus was on representation of the articles online.

This 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. 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.

Universal DTD for Electronic Journal Articles

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.

The report prepared by Inera, Inc., Belmont, Massachusetts, was a survey of the journal article DTDs from the following publishers:

- » American Institute of Physics
- » BioOne
- » Blackwell Science
- » Elsevier Science
- » Highwire Press

CONTINUED »

- » Institute of Electrical and Electronics Engineers
- » Nature Publishing Group
- » PubMed Central
- » 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.

pmc-2.dtd

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, a broad target for conversion of any article content, and one for creating new content, a more prescriptive model that gave explicit rules for tagging content. The first model became the Archiving and Interchange Tag Set, and the second became the Journal Publishing Tag Set.
- 4 The new models should be easily extensible. For example, it should be easy to swap the OASIS CALS (Continuous Acquisition and Life-cycle Support) table model for the default HTML table model.

The NLM DTDs

The NLM DTDs were created based on this initial meeting. Version 1 of the NLM Archiving and Interchange Tag Suite was released in early 2003. It included two article models: the Archiving and Interchange DTD and the Journal Publishing DTD.

NLM created the “Archiving and Interchange Tag Suite Working Group” to advise on changes to the models and the tag suite based on public feedback and their own

usage. Several updated versions were released over the next few years.

Involvement of NISO

When the discussion started about formalizing the Archiving and Interchange Tag Suite with NISO, the plan was to submit the latest version of the Tag Suite and the article models and have them registered. However, the Working Group realized that standardization would bring a lot of attention and new users to the JATS and that this would be an ideal time to make the non-backwards-compatible improvements the Working Group had put on the back burner.

From the beginning of the project, the intent has always been to enable what publishers are doing with their content, not to try to define what they should do. Modifications are based on real user requirements, not on predictions of what may be needed at some time in the future. Both the NLM and the NISO JATS Working Groups saw their roles as normalizing and documenting existing practice to aid in the use, reuse, and interchange of existing and future article content and not to try to influence future directions of publishing.

All pending changes were incorporated into Version 3.0 of the NLM Tag Suite, and the three article models were released in November 2008. The work of the NLM Working Group was concluded, and the NISO Standardized Markup for Journal Articles Working Group was created.

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 all 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.

The Standard includes neither schemas nor much usage information. However, 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.

The Future of the 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.

ISP | doi: 10.3789/isqv25no1.2013.04

JEFFREY BECK (beck@ncbi.nlm.nih.gov) is Technical Information Specialist at the National Library of Medicine.

B. TOMMIE USDIN (btusdin@mulberrytech.com) is President with Mulberry Technologies, Inc.

Both authors were Co-chairs of the NISO Standardized Markup for Journal Articles Working Group.



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.

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

jats.niso.org

JATS supporting documentation

jats.nlm.nih.gov

JATS-Con

jats.nlm.nih.gov/jats-con/

JATS-Con Proceedings

www.ncbi.nlm.nih.gov/books/NBK65129/

JATS E-mail List

www.mulberrytech.com/JATS/JATS-List

Inera, Inc. E-Journal Archive DTD Feasibility Study. December 5, 2001.

www.diglib.org/preserve/hadtdfs.pdf

OASIS CALS table model

<https://www.oasis-open.org/specs/tablemodels.php>



RELEVANT
LINKS