

JATS 2020: Importance and Update

Deborah Aleyne Lapeyre

Mulberry Technologies, Inc.

17 West Jefferson St.

Suite 207

Rockville MD 20850

Phone: 301-315-9631

Fax: 301-315-8285

dalapeyre@mulberrytech.com

<http://www.mulberrytech.com>

Version v1.0 (February 2020)

Copyright ©2020 Mulberry Technologies, Inc.



JATS 2020: Importance and Update

The Journal Article Tag Suite (JATS)	1
What is JATS?	1
Aside: XML is a language for describing documents	2
How JATS is Being Used	2
Why is JATS important?	3
<i>Ubiquity</i> : JATS Depth of Use/Adoption	3
JATS Used Worldwide	3
<i>Standardization</i> : JATS is Maintained by a NISO Standing Committee	4
<i>Enabling</i> : Online Discoverability and Context-sensitive Search	4
<i>Foundational</i> : JATS Family of Standards	5
What's New Since JATS 1.1	5
New: Article Current Version	6
New: Semantic Enrichment for XML Tagging	6
Vocabulary Attributes Enable	7
New: Non-monetary Funding Information	7
New: Non-Latin Emphasis: @style-type	8
New: JATS 2.0 — Clearing our Technical Debt	8
Why are We Considering JATS 2.0?	9
Example of JATS Technical Cruft	9
Not Redesigning JATS from Scratch	10
JATS 2.0 will not be <i>Backward Compatible</i>	10
My Favorite JATS 2.0 Proposed Change	11
So where is JATS in 2020?	11
JATS	11
Colophon	12

The Journal Article Tag Suite (JATS)

JATS is the way publishers, libraries, and archives create, manage, interchange, and store journal articles.

What is JATS?

A named collection of XML elements (marked with tags) that describe a journal article

- Tags name article content, metadata, and bibliographic references
- Describes a single *article* (not an issue, book, magazine)
- Used for:
 - archives/libraries to store and exchange journal articles
 - publishers to produce journal articles
 - publishers/archives/webhosts to interchange journal articles
 - conversion vendors to convert proprietary formats

JATS: ANSI/NISO standard [ANSI/NISO JATS Version 1.2 (Z39.96-2019)]

Aside: XML is a language for describing documents

- XML markup (tags) are placed into the text of a document
- Tags name structural and content (semantic) components of the document
- Computer programs can then process the document in various ways
- Single source publishing: many outputs from one XML source
- Each kind of document (journal article, standard, book) may use different XML tags

```
<article-title>Standard Poodle Best in Show at Westminster</article-title>  
<pub-date iso-8601-date="2020-02-11">  
  <month>Feb</month><year>2020</year></pub-date>
```

How JATS is Being Used

- Public and commercial archives accept or require JATS deposit
- Hosting service and web vendors accept or require JATS
- CrossRef deposits made from JATS
(free public transforms available)
- Consortia develop JATS rules for enhanced interchange
- Publishing new content as well as converting backfiles
- Other tag sets based on JATS

Why is JATS important?

- Ubiquity
- Standardization
- Enabling
- Foundational

Ubiquity: JATS Depth of Use/Adoption

- **Nearly all middle-sized/small journal publishers world-wide use JATS**
- Public and private archives accept (or require) JATS
 - PubMed Central (US and UK)
 - British National Library and Australian National Library
 - US Library of Congress
 - ITHAKA/JSTOR, and many many more
- Conversion vendors can make/use JATS
- The *huge* publishers
 - publish using their own proprietary tag sets
 - *can* (and do) make JATS for interchange/archiving

JATS Used Worldwide

Australia, Belgium, Brazil, Bulgaria, Canada, China, Egypt,

Finland, France, Germany, Italy, Japan, Mexico, Norway,

Russia, South Korea, South Africa, Sweden, Switzerland,

United Arab Emirates, United Kingdom, United States

Standardization: JATS is Maintained by a NISO Standing Committee

- JATS is a living specification
- It grows to accommodate:
 - changes in publishing
 - new use cases
- Users submit requests for changes (on a NISO web form)
- Committee discusses requests and makes recommendations
- When enough changes have accumulated, a new version is voted and released

Enabling: Online Discoverability and Context-sensitive Search

(metadata that travels *with* the XML article)

- Machine-readable context for efficient searching
- Provides semantics and text/metadata structure for data mining/inference
 - Bibliographic metadata for the article
 - title, authors, abstract, journal metadata, ...
 - unique identifiers for authors (e.g., ORCID)
 - unique identifiers for institutions, funders (e.g., ISNI, Ringgold)
 - funding description with grant numbers, investigators
 - Bibliographic references (potentially very detailed) may include data citations
 - Labeled Data Availability Statements

Foundational: JATS Family of Standards

JATS forms the basis for several other standardized document vocabularies

- JATS (ANSI/NISO JATS Version 1.2 (Z39.96-2019) Journal Article Tag Suite) for journal articles
 - Journal Article Archiving and Interchange
 - Journal Article Publishing
 - Journal Article Authoring
- NISO STS (ANSI/NISO Z39.102-2017 Standards Tag Suite) for standards
 - Extended
 - Interchange
- BITS (Book Interchange Tag Set) for books and parts of books, including collection metadata

What's New Since JATS 1.1

(highlights version)

New: Article Current Version

- **Why** : Many articles “published” more than once (preprints, revised article, continuous maintenance article)
- **Content**: A status type, version number, lifecycle stage, etc.
- **Caution**: JATS is agnostic about the exact definition of “version”

```
<article-version
  @article-version-type="preprint">XJDAL-pre-34752</article-version>

<article-version
  vocab="JAV"
  vocab-identifier="http://www.niso.org/publications/rp/RP-8-2008.pdf"
  article-version-type="VoR"
  vocab-term="Version of Record">version-of-record</article-version>
```

New: Semantic Enrichment for XML Tagging

(My favorite you-can't-do-without new thing)

- Semantic linking for JATS
- Tie a JATS element (keyword, subject, id, etc.) to specific semantics with attributes that name:
 - ontology
 - taxonomy
 - vocabulary
 - thesaurus
 - term list or code list
 - industry glossary
 - lexicon
 - database

Vocabulary Attributes Enable

- Associate keywords and subjects with the vocabulary they were taken from
- Record one or more defined “contributor roles” for each author
- Mark an institution identifier as coming from Ringgold or ISNI
- Open Funder Registry Identification

```
<funding-source id="gs1" country="US">
  <institution-wrap>
    <institution>National Institutes of Health</institution>
    <institution-id institution-id-type="doi"
      vocab="open-funder-registry"
      vocab-identifier="10.13039/open_funder_registry"
      >10.13039/1000000002</institution-id>
    </institution-wrap>
  </funding-source>
```

and lots lots more!

New: Non-monetary Funding Information

- JATS metadata <funding-group> describes funding sources, awards, contracts, principals, etc
- But not all support is *money*:
 - facilities
 - computer resources
 - data sets
 - photographs and illustrations
 - genome sequencing
 - deep sea sampling

New: Non-Latin Emphasis: @style-type

- Western emphasis (bold, italic) cannot describe practices of some non-Latin languages
- In Japanese, authors use emphasizing marks associated with a Japanese character called, “Kenten” (dots or sesamis)
- The JATS solution: Added attribute @style-detail
 - to the <styled-content> element
 - used in conjunction with the @style, @style-type, @specific-use
 - provides 3 pieces of styling information and a rationale

```
<styled-content
  specific-use="emphasis"
  style="sesami"
  style-type="dot"
  style-detail="open" >some content...</styled-content>
```

New: JATS 2.0 — Clearing our Technical Debt

Short Answer: A clearer, leaner, tidied-up, more modern version of JATS

- An ordinary JATS release is a dot release (1.1, 1.2, 1.3 ...)
- JATS 2.0 changes numbers because it will be a **big**, ground-breaking release

Currently, JATS 2.0 is a vision, an exploration, a JATS SC discussion

Why are We Considering JATS 2.0?

- We have been accruing “Technical Debt” for years
 - World has changed: Best practice in 1998 may not be best now
 - We guessed on how to model something, but the world has show us a better way
 - Cruft has been accumulating for years
- Opportunity to make JATS better
(for example, replace Authoring Tag Set — rarely used, of little value)

Example of JATS Technical Cruft

- JATS has long described funding information (<funding-group>)
- Added a new group to hold all funding: monetary *and* non-monetary
- Problem: now <funding-group> can be in two places:
 - inside the group (logical place), but also
 - outside the group (where it has always been)
 - could not remove old location because not backward compatible
- In JATS 2.0, will allow only *inside* the group

Not Redesigning JATS from Scratch

- A maintenance release of JATS
- Not a totally new vocabulary
- A cleanup not a demolition
- Still several years away
 - Notice was given at JATS-Con in May 2019
 - JATS 2.0 will not become a standard before 2021
(and probably not that soon)

JATS 2.0 will *not* be *Backward Compatible*

Since the beginning, all JATS versions have been backward compatible

- Backward Compatible means:
 - Any document valid to any JATS tag set
 - Will be valid to ALL subsequent JATS tag sets
 - ***No changes to content or XML tagging are necessary!***
 - *If you change:*
 - the @dtd-version attribute on <article>
 - the DOCTYPE Declaration
- JATS 2.0 makes no such promise, just promises better

My Favorite JATS 2.0 Proposed Change

Make a new simplified JATS subset

- Specifically designed for XML tool use and schema-driven editing
- Intended audience:
 - publication departments
 - production staff
 - editors
 - Quality Assurance staff

So where is JATS in 2020?

- JATS is the *present and immediate* future of journal publishing
- JATS Standing Committee hard at work on next version
 - based on latest user requests
 - probably released this year
- We are considering the first *backward incompatible version* (JATS 2.0)

JATS

**JATS is no longer one of the cool kids;
it's just what you do if you have journal articles.**

— Jeff Beck of PubMed Central at JATS-Con 2017

Colophon

- Slides and handouts created from single XML source
- Slides projected from HTML generated from XML using XSLT
- Print copy created from the same XML source
 - XSLT transform generates XHTML
 - Antenna House Formatter makes PDF from:
 - XHTML
 - CSS3 (slightly extended)
 - Graphics sizing table