XML: Why and How JATS

Deborah A Lapeyre
Mulberry Technologies, Inc.
17 West Jefferson Street
Rockville, MD 20851
USA
dalapeyre@mulberrytech.com

Version v1.0 (September 2018)

XML: Why and How JATS

The XML Journal Article Tag Suite (JATS) 1
JATS is a standard XML tag set (vocabulary) 1
JATS defined for *interchange* of journal articles 2
How JATS is being used 2
Why JATS is important? 3
Ubiquity: JATS depth of use/adoption 3
JATS naturalness of use 4
JATS helps to enable Open Science 4
Open Data 4
Enabling: Declarative and structural markup 5
Enabling: Semantics and structured metadata for search/data mining 5
JATS is how the world publishes journals 6

slide 1

The XML Journal Article Tag Suite (JATS)

- XML is a language for describing documents
  - XML markers (tags) are placed into the text of a document
  - these tags describe the document's structure and content
  - computer programs can then process the document in various ways
  - each kind of document (report, memo, recipe, standard) has different XML tags
- JATS is an XML tag set describing journal articles
  - JATS-specific tags are inserted into an article
  - used to produce and interchange articles
  - used to archive and host journal articles

slide 2

JATS is a standard XML tag set (vocabulary)

(the international language for describing journal articles)

JATS is a named collection of XML elements (marked with tags)
for describing
the structure of *one journal article*
called a "model" of the article

- Originally modeled STEM articles (Scientific, Technical, Engineering, Medical)
- Now covers all journal types (sociology, economics, soft sciences, etc.)
- Models only journal *articles*, not books, issues of journals, magazines, etc.

JATS is an ANSI/NISO standard:
ANSI/NISO Z39.96-2015 JATS: Journal Article Tag Suite (JATS V1.1)

slide 3

JATS defined for *interchange* of journal articles

(article content and article metadata interoperability)

- Each publisher/archive/library may use their own XML model
- They need to get their articles into *the same form* of XML:
  - to exchange information with each other
  - to put information into a single repository
  - to sell/display items on the same hosting platform
  - to share tools and resources
  - so conversion vendors do not need to learn their particular model

slide 4

How JATS is being used

- Publishing new content
- Converting backfiles
- Public archives such as libraries accept or require JATS deposit
- Hosting service and web vendor intake
- Private and commercial archives ingest JATS

slide 5

Why JATS is important?

slide 6

Ubiquity: JATS depth of use/adoption

- Most middle-sized/small publishers world-wide use JATS
- All of the huge publishers *can* (and do) make JATS for interchange
- SciELO Publishing Schema is a JATS superset
- Public and private archives accept (or require) JATS
  - PubMed Central (US and UK)
  - British National Library
  - Australian National Library
  - US Library of Congress
  - ITHAKA/JSTOR, and many many more
- Conversion vendors all know how to handle JATS
- Numerous web-hosting and service vendors require or support JATS
- Free, public transforms make CrossRef deposits from JATS

slide 7

JATS naturalness of use

Models what publishers are already doing

- Models *current* article publishing practices
  - does not lead practice, but reflects it
  - changes as practices change
- Tries to preserve current text order (reading sequence)
- Should be easy to encode *anyone's* journal articles in JATS
- Does not typically define or set "Best Practices"
- Escape hatch elements for metadata not defined by JATS

slide 8

JATS helps to enable Open Science

slide 9

Open Data

Open Data means that documents should be:

- Machine-readable (also hopefully human readable)
- Searchable (and mineable)
- In a non-proprietary format
- Freely available and open (able to be modified)
- Platform and vendor neutral

Only a publisher can enforce truly Open Data

But JATS is a key tool in that effort (providing XML, Unicode, and a public document model with clear semantics)

slide 10

Enabling: Declarative and structural markup

- Separates content from presentation/behavior
- Tags what something is, not what it looks like/behaves
- Therefore multiplicity of output styles/formats/behaviors possible from a single XML source
  - print in all sizes and styles (branding)
  - web pages and apps
  - eBooks
  - accessible and pronouncing material

slide 11

Enabling: Semantics and structured metadata for search/data mining

(extensive metadata that travels *with* the XML article)

- Bibliographic metadata for the article (title, authors, abstract, journal info)
- Unique identifiers for authors (e.g., ORCID)
- Unique identifiers for institutions, funders
- Detailed funding description with grant numbers, investigators
- Labeled Data Availability Statements
- Links that tie terminology to ontologies/taxonomies
- Bibliographic references (potentially very detailed)
  may include data citations
- Machine-readable context for efficient searching

JATS is how the world publishes journals

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

JATS is in use in *at least* 25 countries world-wide including: Australia, Belgium, Brazil, Bulgaria, Canada, Egypt, Finland, France, Germany, Italy, Japan, Norway, Russia, South Korea, Sweden, Switzerland, United Arab Emirates, United Kingdom, United States, etc.

And SciELO use of JATS adds Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Mexico, Peru, Portugal, South Africa, Spain, Uruguay, and Venezuela.
(Paraguay soon?)