Publish Faster. Publish Smarter.

Solving Problems with Standards

Subsetting the JATS DTD – So What?

NISO Plus 2021

Charles O'Connor
Business Systems Analyst
coconnor@ariessys.com

The JATS DTD

- Developed by the U.S. National Library of Medicine
- First version released in March 2003
- The Journal Article Tag Set became a NISO standard in July 2012
- JATS is the de facto standard for journal articles in scholarly publishing—not just science, technology, and medicine

***Solving Problems* with Standards**

or

Solving *Problems with Standards*?

Problem: JATS is Too Loose

- JATS is descriptive, not prescriptive, *by design*

- JATS allows
  - 11 different ways to associate authors and affiliations
  - 2 distinct bibliographic reference models
  - Different publication history capture methods <history> or <pub-history>?

- JATS comes in flavors
  - Archiving and Interchange: preserve existing content (rather loose)
  - Publishing: publishing, hosting, and archiving (a bit less loose)
  - Authoring: writing new articles (nobody uses it)

Machine Readable: JATS for Reuse

- Optimize the reusability of scholarly content by developing best-practice recommendations for tagging content in JATS XML

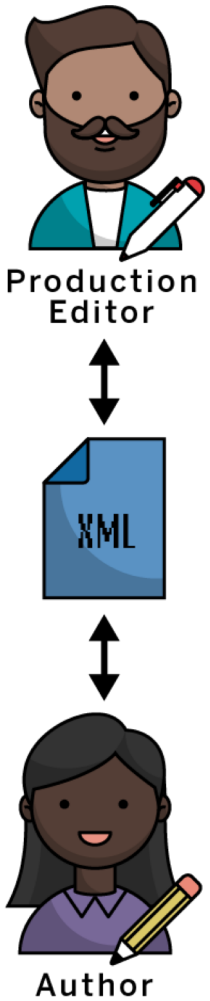
- Recommendations for
  - Authors and affiliations
  - Bibliographic references
  - Data availability statements
  - Funding
  - Permissions
  - etc.

Machine Readable JATS

- Where is the information I want (pull)?
  - (a few different places is OK)
- What does this information contain?
  - Example: complete affiliations
- How is this information formatted?

JATS XML-Through Workflows

- Reduce publication times
- Reduce introduction of errors
- Save money
- Greater focus on metadata
- Better author/staff experience
- More amenable to automated analysis
  - AI document analysis
  - QA with Schematron

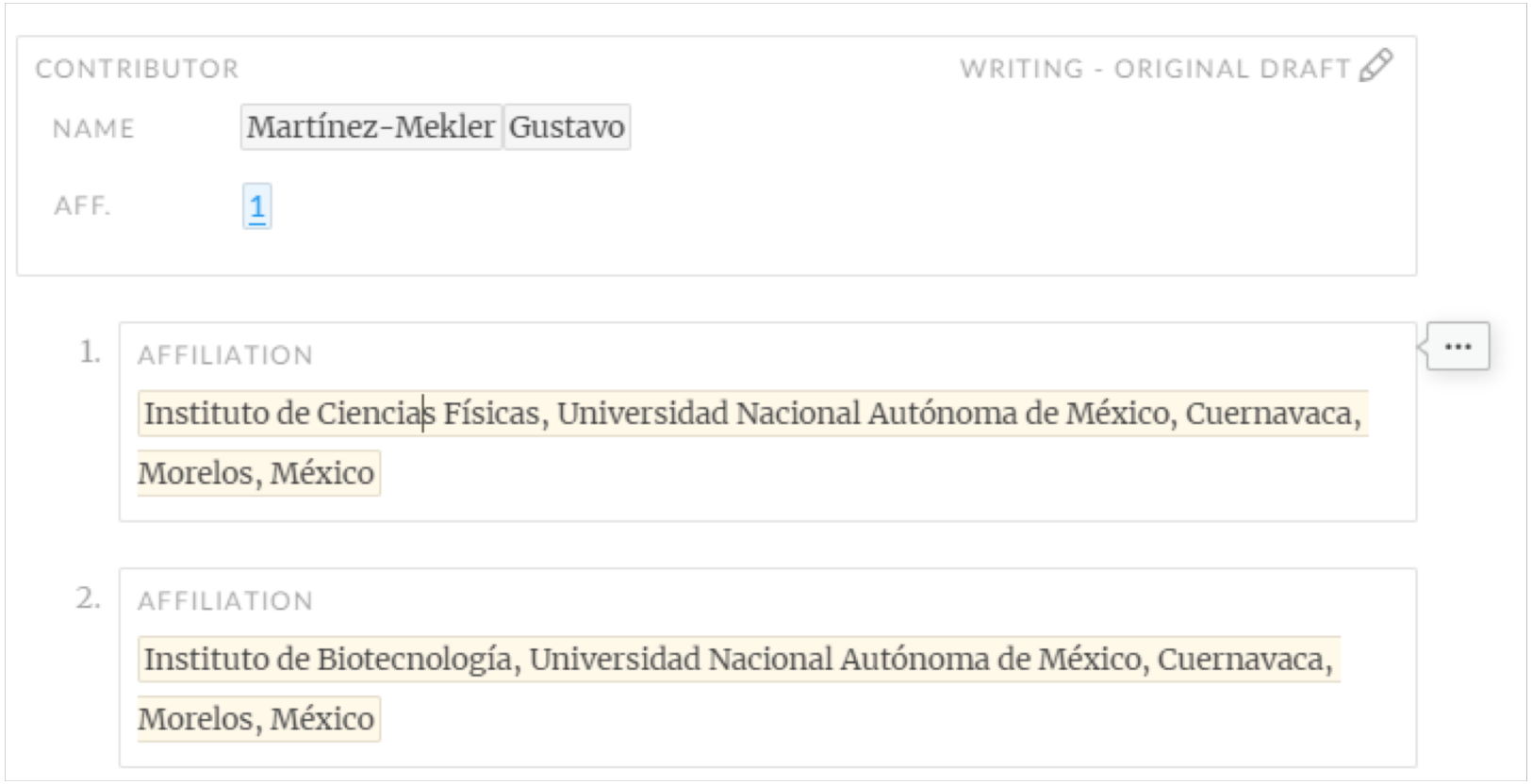


Machine Writeable JATS

- Where does this information go (push)?
  - (only one place, hopefully)
- How does this information interact with the rest of the article?
- How am I entering this information?
- Solution: Subset the JATS DTD

Subsetting Example 1: <aff>

- Only allowed one place: within <contrib-group>
- Must be associated to authors via an <xref> cross reference
- No <label>



Subsetting Example 2: <ref>

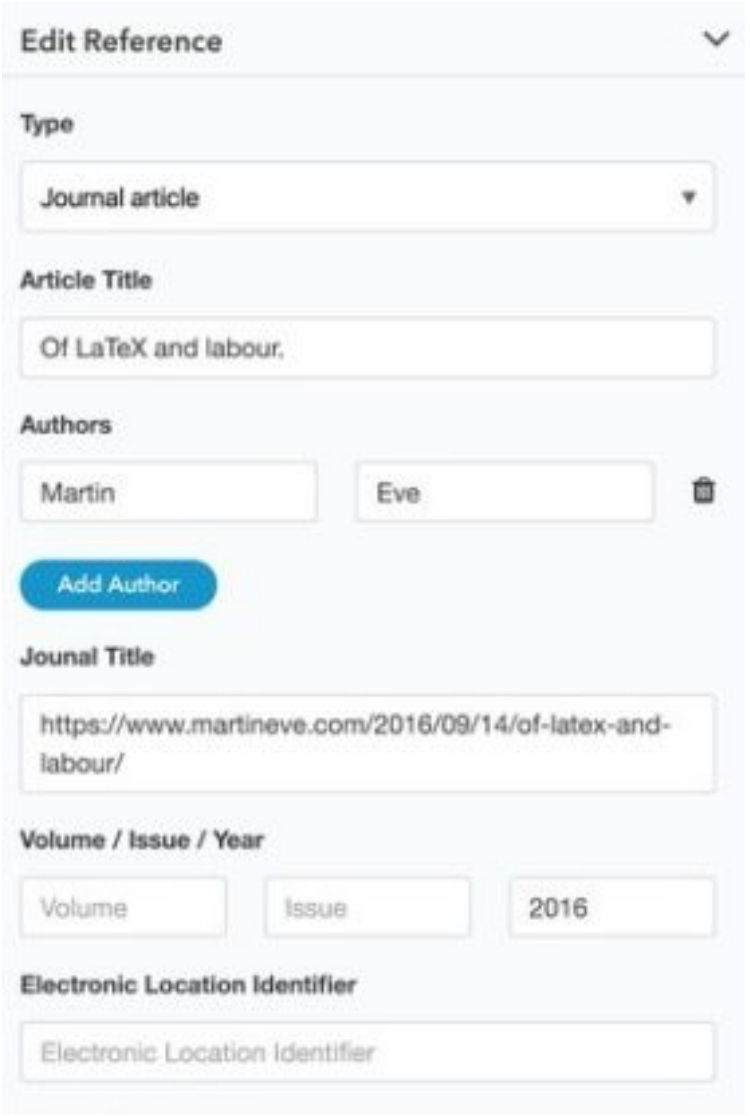
- No <label>
- and um, uh

Both versatility and specificity of this ion as a physiological signal lie in how it is temporally and spatially organised within a cell [1]. Calcium oscillations are widely described in both excitable and non-excitable cells, and lie at the core of several signalling processes, such as neuronal firing, embryonic cell differentiation, immune cell activation and rhythmic beating of the heart [2-5]. A good example is the calcium induced calcium release, which is generated by an inositol trisphosphate and calcium sig This link text is not editable because it is generated automatically. ses [6], for which its operation has been understood with the help of mathematical models [7, 8]. Furthermore, in olfactory neurons, calcium influx is regulated directly by cyclic nucleotides [9 , 5].

Subsetting Example 2: <ref>

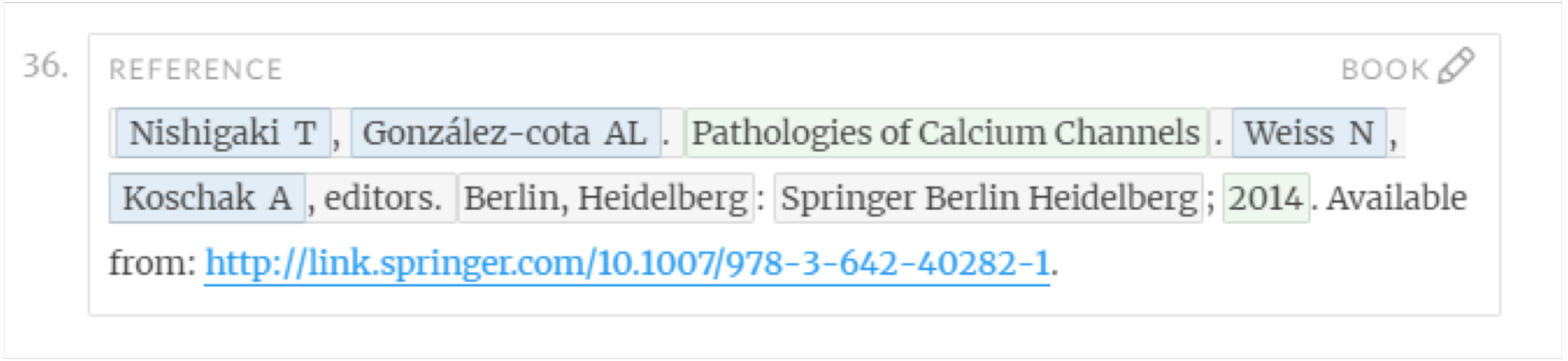
- Form-based?
- Only <element-citation>
- Contains reference elements only
- No punctuation, spaces, or untagged text

```
<element-citation publication-type="journal">
    <person-group>
        <name><surname>Eve</surname><given-names>Martin</given-names></name>
    </person-group>
    <article-title>Of LaTeX and Labor</article-title>
    <source>https://martineve.com</source>
    <year>2006</year>
</element-citation>
```



Subsetting Example 2: <ref>

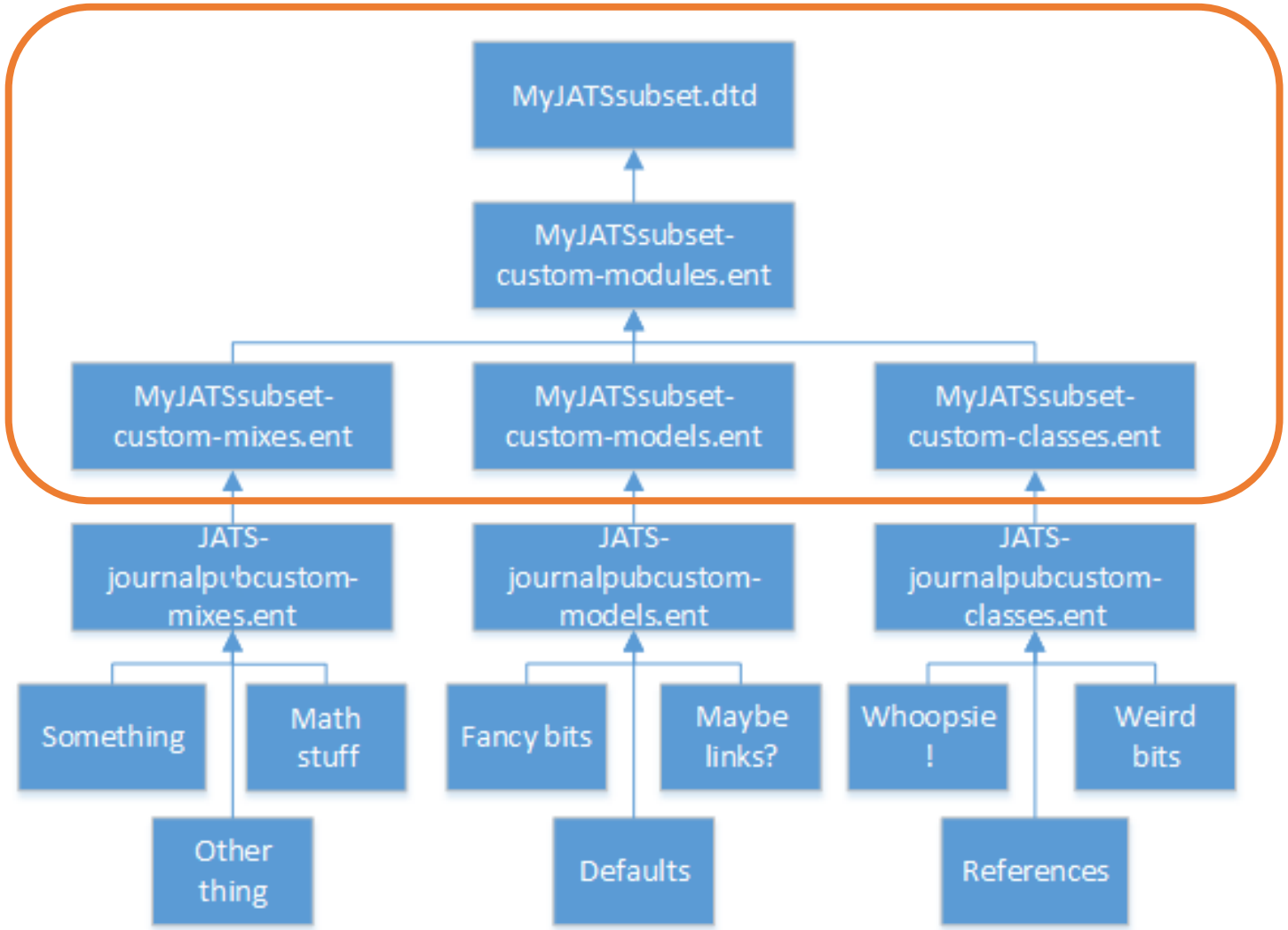
- Text-entry based?
- Only <mixed-citation>
- Punctuation, spacing, and untagged text allowed



Solving Problems with Standards

JATS is Made to Be Customized

- JATS is modular
- You don't need to touch common files
- Just override content models in your custom, high-level files



JATS Compatibility Meta-Model

- For anyone interested in JATS customization, we recommend they read the JATS compatibility meta-model, which provides key insights into how JATS is structured.

https://groups.niso.org/apps/group_public/download.php/16764/JATS-Compatibility-Model-v0-7.pdf

oh, and also Schematron

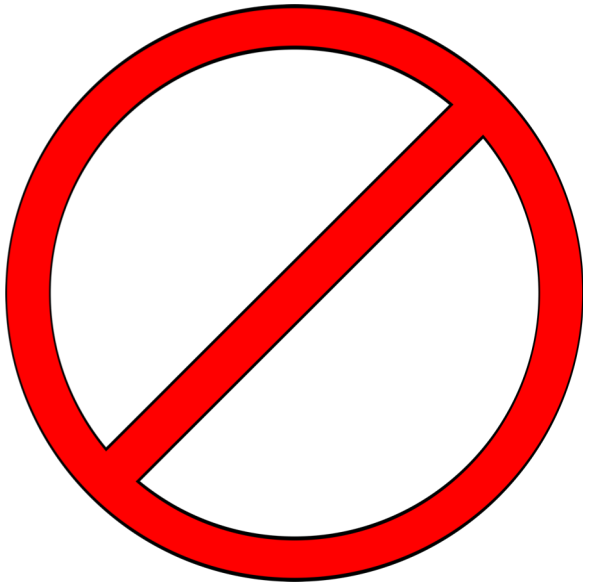
- Schematron is a rule-based validation language
- You can validate content as well as structure
- You can write you're your own error messages
- Not all of the restrictions that you want to enforce can be expressed via DTD

What Does Subsetting Get You?

- Tools for creating and editing XML are easier to build, easier to maintain, and more robust
- Predictable content for rendering and transforming
- Clear expectations for suppliers and other partners
- A handy means for enforcing these expectations

What to Avoid When Subsetting

- Don’t get rid of things that are mandatory in the parent DTD
  - Your subset should be valid to the parent DTD
- Don’t get rid of anything if it doesn’t get you a “win”
  - Me sowing . . .
  - Me reaping . . .
- Don’t forget to document everything



Aries Systems Corporation
50 High Street, Suite 21
North Andover, MA 01845 USA

www.ariessys.com