Monday, February 16, 2026

[8:45 am–9:00 am] **Welcome and Introductions**
Speaker: Jeffrey Beck, *JATS-Con Chair*

[9:00 am–9:45 am] **JATS Black-and-White?**
Speaker: B. Tommie Usdin, *Mulberry Technologies, Inc.*

JATS variations are nick-named Green, Blue, and Pumpkin. The models vary in how permissive/flexible they are. Even the most restrictive is, by many people's standards, excessively loose. In the loosest model there can be a dozen ways to encode the same information. This flexibility is there to accommodate the needs and wants of the creators of JATS documents and comes at the expense of the recipients and users of those documents. Perhaps it is time to recognize that JATS is a success; in many communities publishers and other content creators are now obligated to create JATS. Given that power, perhaps we should clamp down on this nonsense, build a JATS that forces publishers to do it right, and give the archives and database vendors the support they have been quietly asking for. Can we build a version of JATS that has only one way to encode each type of information? That requires documents to contain the metadata that should be there? Is it time to require, accessibility information, math and tables in one and only one encoding, clear ownership and permissions information, sensible style rules such as titles on sections? Let's discuss how we got here and if we can clean up this mess!

[9:45 am–10:30 am] **Merging full-text query with data sets: a perspective from compiler theory**
Speaker: Nathaniel Christensen, *Linguistic Technology Systems*

Data transparency and Minimum Information standards have become an important part of scientific research. Structured data sets -- sometimes called "research objects" -- often accompany publications, but document-preparation software is still largely separated from dataset tools. Partly this may reflect copyright concerns, but the "diamond" open-access model simultaneously allows authors to retain all rights to their work and readers to access documents in multiple available formats (PDF, XML, LaTeX, etc.) -- in short, data sets and text publications effectively fall under the same licenses. As such, there is no non-technical restriction on computationally tractable text encodings for associated publications to be present in open-access research objects, alongside raw data files and utility code. We can thus envision supporting code (perhaps itself included in the research object) to work with both full-text and dataset representations, providing functionality such as deserializing data files and searching against natural-language content. This might be achieved via tools similar to XQuery or DOM traversal, but they should be specialized for the unique facets of encoding written text (with details about sentences, block/inline quotes, footnotes, annotations, citations, and other features specific to academic writing) as well as parsers sufficient for many (often discipline-specific) scientific data formats. In this presentation, based on experience developing an existing software package that has been used for multiple books and articles, I will discuss via compiler theory some problems or techniques in the contexts of top-level parsers, intermediate representations, virtual machines, and runtime code libraries, that are specific to full-text query evaluation in conjunction with published data sets. I will summarize dimensions of regular-expression-based parsing, instruction sets, overload-by-contract, lifetime management, native interop/dynamic libraries, dependent types, template metaprogramming, and techniques for encapsulating compiler/runtime stacks (such that they are self-contained and embeddable) insofar as these topics specifically overlap with full-text search and dataset management.

[10:30 am–10:45 am] **Coffee Break**

[10:45 am–11:30 am] **XML conversion isn’t the hard part: Creating JATS in a publishing environment**

Speakers: Caleb Clauset, Chandi Perera, and Robin Dunford, *Typefi Systems*

There have been many attempts to build editorial tools that support XML publishing workflows for JATS, BITS and STS (and many have been presented at this conference). For most, generating XML is the primary focus and the biggest challenge—but it's actually the easiest problem to solve.

Inera eXtyles has long been considered the gold standard in JATS, BITS, and STS publishing workflows, but its end-of-life in August 2026 is fast approaching. Typefi took this opportunity to build a replacement called Typefi Orion, with extensive input and feedback from actual editors during every stage of development. During this collaborative process, it became clear that the editorial workflow is actually the most challenging piece of the JATS puzzle.

This paper will explore the unique development process used for Typefi Orion and reveal some of the knowledge gained during the process, including the critical requirements for JATS editorial tools that are often ignored. We believe in a diverse ecosystem of JATS, BITS, and STS tools, and these insights will help other developers build better JATS, BITS, and STS tools.

[11:30 am–12:15 pm] **Deep XML Modeling at Scale: Consolidating 80 Years of Legacy Content into a Unified, Standards-Driven Platform**

Speaker: Mark Gross, *Data Conversion Laboratory, Inc.*

The Association for Materials Protection and Performance (AMPP) was formed through the consolidation of three organizations, bringing together more than 80 years of technical papers, standards, books, training materials, and articles that were created in different formats, governed by different workflows, and tagged (if tagged at all) according to different rules. To support unified member services and prepare for modern digital delivery, AMPP needed a way to normalize this heterogeneous archive and express it in a single, coherent, deeply modeled XML format.

This case study examines how AMPP transformed decades of disparate, unstructured, and inconsistently structured documents into a richly tagged JATS/BITS corpus. We dive deep into granular levels that were required to represent the technical nuance, including detailed metadata tagging, normalization of terminology, extraction and structuring of entities, and more.

This paper presents how applying text analytics (NLP, entity extraction, topic modeling) accelerates discovery and tagging and prepares AMPP for future AI initiatives.

[12:15 am–1:45 pm] **Lunch [On Your Own]**

[1:45 pm–2:30 pm] **Open tools and technologies in PubNote supporting XML for OLSPub and PubMed users**

Speaker: G. Ken Holman, *Réalta Online Publishing Solutions Ltd.*

Announced in mid-2025, the PubNote GitHub repository (https://github.com/realtaonline/PubNote) was created by Réalta Online Publishing Solutions Limited as a set of simple tools to assist OLSPub and PubMed users of XML metadata files. These tools recognize that having to deal with raw XML markup can be discomfiting to those allergic to angle brackets.

This presentation overviews the free and open technologies used in deploying PubNote as an instantiation of the sustainable single source publishing model described formally in https://www.ncbi.nlm.nih.gov/books/NBK591971/ at JATS-Con 2023.

Also included is a status on the ongoing development of the manual creation of PubMed metadata XML using a non-XML syntax.

This open project was developed expressly to support the Open Life Science publication database OLSPub initiated in 2025 under the auspices of ZB Med, the German Medical Library.

***[2:30 pm–3:15 pm] **STS [NIST]**

Speaker: TBD

[3:15 pm–3:30 pm] **Coffee Break**

[3:30 pm–4:45 pm] **Open Session**

Speaker: Debbie Lapeyre, *Mulberry Technologies*

[4:45 pm–5:00 pm] **Conference Closing**

Speaker: Jeffrey Beck*, JATS-Con Chair*

****