JATS—Where's It Going, Where Has It Been?

An Interview with Bruce Rosenblum of Inera

September, 2017—NISO recently spoke to Bruce Rosenblum, CEO of NISO Voting member [Inera, Inc.](), about the development of [ANSI/NISO Z39.96, JATS: Journal Article Tag Suite](). JATS provides a standard XML format in which publishers, archives, and others in the journal publishing ecosystem can exchange metadata and full text of journal articles.

NISO: Can you give readers some background on why JATS was needed?

Bruce: By 2001, the ISO 12083 standard for journal articles [[ISO 12083:1994, Information and documentation—Electronic manuscript preparation and markup]()] was not widely adopted having been described as "way too complicated, yet not flexible enough." As a result, most publishers who had implemented SGML had created custom DTDs, and many publishers had not even started such projects due to their cost and complexity.

Inera was approached in 2001 by Harvard University and The Andrew W. Mellon Foundation to look at format issues for long-term archiving of electronic journal articles. At the time, some publishers were beginning to create online-only journals and people in libraries were beginning to ask how to archive these materials.

In our first conversation with Harvard, we determined two things: PDF was not a viable archive format, and XML was usable as an archive format, but each publisher had its own DTD. We were asked if it would be possible to create a single archive DTD that could be used as a common format to preserve the intellectual property of all journals.

We performed a month-long study that examined DTDs from 10 different publishers and determined that the model could be built, but the estimate cost was prohibitive. After further exploration, collaboration with PubMed Central seemed possible. PubMed's DTD wasn't quite where we needed it to be for Harvard's use, but their consultant was Mulberry [NISO Voting Member Mulberry Technologies, Inc.]. A meeting took place that included me, Debbie Lapeyre [Mulberry Technologies, Inc.], Jeff Beck [The National Center for Biotechnology Information (NCBI)], David Lipman [NCBI], Don Waters [Andrew W. Mellon Foundation], and Dale Flecker [Harvard University Libraries]. At the meeting's end, Debbie, Jeff, and I were asked to collaborate on a DTD that would meet Mellon and Harvard's requirements. (The archive was later built as Portico).

What was the result of that collaboration?

It took a year to develop NLM version 1.0 [the predecessor to JATS], and it became clear to us during that time that it would be pretty neat! It would be better documented than proprietary DTDs and really flexible. It would also be in the public domain as it was NLM- and Mellon-funded.

> "It became clear to us that NLM 1.0 … would be pretty neat!"

When we were getting close to the one-year mark, I was working with an Australian journal publisher, [CSIRO](#), and they needed a DTD. They didn't want to build their own, and they were considering Blackwell's and Elsevier's. I asked Harvard and Mellon if CSIRO could use what we were developing. Permission was granted, and CSIRO became the first organization to adopt it. They were even ahead of NLM and Portico (which didn't yet exist).

Between 2003 and 2007, a growing number of publishers and delivery platforms adopted the NLM DTD and as a vendor, Inera encouraged organizations to adopt it. Wider adoption meant we could develop our product in a less custom and more standard fashion, which allowed us to provide our eXtyles software to

customers at a lower cost. This proved critical for us as a vendor; we adopted the nascent standard early on and found it a great way to offer our product to a wider range of customers and build our business. And we're not the only company to benefit—the standard works for anyone touching journal article XML, including conversion vendors and online hosts, providing economies of scale that trickle to those companies' customers in turn.

It was not difficult to get vendors to take on this NISO standard, because vendors do what their customers tell them. If customers say they want JATS, that's what vendors will do. They will ask, for example, for XML that will be accepted by PubMed central, which requires JATS. In other cases, publishers just ask vendors for XML because they know they should have it, and vendors will default to JATS because they have an efficient workflow for it. In this regard, adoption of the standard is a happy accident.

How did NISO become involved?

Once the DTD began to have legs, we started to hear grumblings: "It's from NLM, it must be only for medicine." But the NLM DTD team was focused as much on non-science as on science content, meaning that we realized, for example, that we needed to account for Greek footnotes in an English language archeology journal. We also knew that significant adoption would only happen if

it were a real, not a de facto, standard. And since NISO had many other standards related to scholarly publishing, it made a logical home. So that's how it came to NISO.

By about 2006 or 2007 we cleaned up our work and made it known that that our upcoming version would be the last fully backwards compatible NLM DTD one. We then moved the work to NISO, which released JATS 1.0 in 2012. Now the standard is at the point where if we break backwards compatibility, we would need to give a few years notice.

Many of the same people are working on JATS, BITS [an extension of JATS for books], and STS [NISO Standards Tag Suite]. Collectively we have now made ground rules for all of those standards so as to avoid mistakes. During a 2017 presentation at JATS-Con the JATS metamodel was presented, which includes rules for moving ahead and being strategic rather than tactical. Success has bred its own problems: we now get more requests than previously for additions and improvements. The JATS Standing Committee works to address new requirements. The JATS standard has created its own community, including the annual JATS-Con conference, the JATS email group, and the JATS4R group.

Has the NISO version of the standard been widely adopted?

There has been wide adoption of JATS in scholarly publishing. The only significant

exceptions are Elsevier, Springer, and Wiley-Blackwell, as they had proprietary XML models predating JATS. But they are all able to move into and out of JATS where necessary.

JATS has been more successful than we ever imagined. In many ways, it was an accident waiting to happen. By the time the NLM DTD got out the door, people were really looking for an off-the-shelf XML standard. A large part of the market was locked out of going toward XML without such a standard.

"JATS has been more successful than we ever imagined."

Now the success of JATS seems like a foregone conclusion, but it wasn't always. If any single publisher had said that they were going to make what they had done freely available as a standard, people would have wondered what they had up their sleeve. But our work came out of a skunkworks project—we didn't set out to create a standard, we tried to solve a problem. When others saw that work was open, well-documented, and extensible, they chose to adopt it rather than re-invent it.

What's next?

Even though publishing in the Internet age is changing daily, the JATS Standing Committee has decided to be

retrospective, meaning that we only make changes based on established needs. In order to add something to JATS, there has to be a documented use case. We also try to be thoughtful and not act too quickly. For example, in late 2007, I returned from Japan and realized that multi-script text, like author names in Japanese articles, wasn't supported. Initially, we thought: "Stop the presses! We have no way to markup an author's name in multiple scripts." But then the Standing Committee decided implementing this feature needed more thought as there was no simple way to add it. We decided not to squeeze this capability into NLM 3.0, but waited until the JATS 0.4 draft. The final result is a very flexible mechanism that supports more than Japanese names. But that support was critical to Japan's national commitment in 2011 that all journal articles are in JATS. Similar national commitments have come from Brazil and Mexico. So we're responding to needs, but doing it thoughtfully.

> ## "…we're responding to needs, but doing it thoughtfully."

Now we're working on JATS version 1.2. We've had requests for new elements, improved documentation, more use cases, and more best-practice information. We just encountered a real-world scenario that requires a tweak to the standard; it's CRediT, CASRAI's guidance on how to give authors credit for their work. This new taxonomy is used in tenure evaluation and is catching on quickly because of needs in publishing. A preprint went up recently on BioRxiv that recommended ORCID and CRediT should be used by all journals. While JATS 1.1 supports CRediT, the Standing Committee realized that we could provide a more sustainable model for CRediT and other taxonomies with an update. This new model has been proposed for JATS 1.2, and also appears in relation to subject taxonomies in the forthcoming ANSI/NISO STS standard. It was especially cool for me to watch this new idea being "baked" in both the STS and JATS Standing Committee simultaneously and to see how well it turned out for both groups.

Bruce Rosenblum is CEO of publishing technology company Inera, Inc. He served on NISO's Board of Directors from 2005 to 2013, is a member of the JATS Standing Committee, and is currently co-chair of the steering and technical working groups that are developing NISO Standards Tag Suite (STS).

NISO, the National Information Standards Organization, is an ANSI-accredited non-profit association that identifies, develops, maintains, and publishes technical standards to manage information in the digital environment.