Interview with Pablo Fernicola

Martin Fenner, Gobbledygook

November 7, 2008

The Article Authoring Tag Set of the National Library of Medicine (NLM DTD for short) creates a standardized format for new journal articles that can be used by authors to submit publications to journals and to archives such as PubMed Central.[1] The Microsoft Word Article Authoring Add-in that was released earlier this year reads and writes this format. Pablo Fernicola from Microsoft first explains the Add-in in a video and then answers a few questions.

1. Can you describe what the Microsoft Word Article Authoring Add-in is and does?

There is an already ongoing transition to digital workflows for journals, as well as a nascent transition to digital distribution and consumption. Generating content that is best suited for digital distribution, as well as for archival, search, and semantic analysis in the future, is going to be essential. As a community, today we are not taking full advantage of the potential that the tools and formats currently in use present to us, and trying to bolt these capabilities to existing print centric processes is costly, inefficient, and does not allow for exercising the benefits afforded by the digital medium. Authoring for print delivery still dominates many of the processes, and constrains the final outcome, even if print distribution is discontinued.

It is important to realize that the content we generate today is what we will be accessing, and relying on as reference material, in the future. It is imperative that we generate content that is best suited for the way it will be consumed. Also, it would be ideal if this transition could be done in a non intrusive and low cost approach as possible.

The Article Authoring Add-in for Word 2007[2] is a free download, which provides new capabilities to the Word application, focused on Scholarly Publishing. The overall goal for the Authoring Add-in project is to help improve the scholarly publishing process for workflows that rely on Microsoft Word or Word generated content, covering the authoring experience, editorial workflow, and archiving of STM articles, with an eye towards generating content that is best suited for preservation, digital consumption, and search. A core capability provided by the add-in is the ability to open and save files in the NLM XML format.

**2. How is the Authoring Add-in different from commercial tools such as
eXtyles from Inera?**

The add-in differs in at least the following ways:

- It is targeted at both authors and editorial staff audiences
- It focuses on enhancing the experience within Word, both for authors and editorial staff, and not just for content but also for metadata
- Enables providing a consistent experience for authors, in relation to templates and the entering of metadata, across journals
- Enables two way interaction with the NLM formats (article and book formats), saving and opening NLM files within Word
- Provides a platform for software vendors to build on top of
- Builds on the transition to XML as the underlying native format in Word 2007

3. What advantages do you see for authors that submit their manuscripts in the NLM Journal Publishing XML format?

The add-in doesn't force authors to submit their articles in the NLM format, but makes the conversion to that format a lot easier as part of the workflow. The add-in provides a way for authors to enter information in their articles so that this content (semantics and metadata) is preserved through the publishing workflow, in a way that is ready to save as a valid NLM document, while still using the Word user interface. I would expect that the more common scenario will be that of journals providing templates and authors submitting docx files, augmented with NLM data through the add-in, and that the journals or repository staff will be the ones doing the conversion to the NLM format.

4. How do you think the author submission process will change in relation to formats?

Many journals already use the NLM format as part of their publishing workflow, as well as for their archival format. Some publishers are moving to the format now. And, certainly for NIH funded research, all articles eventually end up in the NLM format as part of the submission to PubMed Central. But I don't know of any journals that take in the NLM format directly from authors.

We have tried to take a very end user centric approach in the project. We would like for authors not to have to be aware of the underlying format. We don't want authors to think of XML, for example, as formats should be something that happens in the background, which the authoring tools handle for the authors, and just make the authors' work easier, and their content more easily searchable and relevant.

In our work with the journals, publishers, and repositories we focus a lot on interoperability based on formats and protocols commonly used in the community,

not only in the form of the NLM format itself, but also by incorporating technologies such as SWORD[4] and OAI-ORE[5].

5. What is your job at Microsoft? What did you do before working on the Authoring Add-in?

I have always been involved in software development, working and managing both small and large teams. Currently I am a group manager at Microsoft, in charge of running this overall project, which also includes an online service focused on the peer review process. I drive the development, the technology direction and architecture, and community engagement as it relates to scholarly publishing[6]. Before starting this effort, I worked for many years on developer platform technologies related to text, reading, graphics, and multimedia, both at Microsoft and at Apple, as well as being the Program Manager in charge of the web developer platform in Internet Explorer for a couple of versions of the browser.

6. Do you plan to also release an Add-in for Microsoft Word 2008 for Macintosh?

We are investigating how we can best bring the authoring focused features to Word 2008 users. On Windows, Word 2007 is now in many ways a developer platform in its own right, there are even software development kits for it, and Word 2007 provides a lot of extensibility and programmability to developers. The equivalent developer support is not provided in Word 2008, with the Macintosh offering's strength being on providing a great end user experience to its end user audience.

7. Do you want to talk about future plans for the Authoring Add-in?

We got a lot of feedback on our version 1 of the add-in, which we made available this past July, from folks involved in the back-end of the publishing workflow, such as folks at journals, publishers, and repositories, as well as from companies that provide the software tools and services in support of this work (of note is the integration work we did with Design Science for their MathType package7).

Development work on version 2 is currently underway and we expect to make available a Technology Preview soon, with the final release of version 2 in 2009. In version 1 we focused quite a bit on the architecture and getting the basic infrastructure in place to provide support for the NLM format. In version 2 there is a greater focus on the author experience, as well as on continuing to improve the support for the format.

Some of the driving questions that we would like to address are:

- in which ways can we make the submission/upload process easier for authors

- Can we make the author and article metadata more reliable and consistent, thereby reducing the roundtrips between authors and the journals, as well as reducing the cost for cleaning up the data?

And overall, the philosophy is to simplify, simplify, simplify. Especially for authors, help them get the content into the article, and keep the technology in the background. For the staff at journals and repositories, provide them with access to all the richness of the NLM format, and the flexibility that they will need to build their own solutions.

fn1. Article Authoring Tag Set

fn2. Article Authoring Add-in for Microsoft Office Word 2007

fn3. eXtyles Product Information

fn4. SWORD

fn5. Open Archives Initiative Object Reuse and Exchange

fn6. ex Scientia

fn7. MathType