

Meeting Report: The NLM DTD

London, ALPSP Technology Update 21 Sept 2006

Is there a need for a standard XML DTD? A recent ALPSP Technology Update looked at this issue, and looked at the case for the emerging NLM DTD as a standard for journals. XML is one of those strange technologies that seems to be a standard, but which in practice proves remarkably problematic to implement systematically – after all, XML enables every user to create their own element labels. Interestingly, it was Greg Suprock of Nature Publishing Group (not currently using the NLM DTD) who pointed out the benefits of a single DTD. Many different publishers may implement many different DTDs and then integrators have a major conversion job to bring all the content together. Of course, you can convert from one DTD to another, but in practice, that's not a very elegant solution. A whole host of conversion filters is required, and even then, the result might not be what publishers want. They might not achieve inter-searchable metadata, for example.

So a standard such as the NLM DTD is potentially welcome, and since it is being progressively more widely adopted, it becomes still more valuable. The DTD is not copyrighted, can be freely downloaded, and is maintained by a wide-ranging board. Although derived initially from life-science journals, the standard was designed to be used for all types of journals – there isn't, for example, a specific element for "gene". With a take-up of a wide range of organisations, including PubMed Central, Atypon, Portico, HighWire Press, and Ingenta, plus the British Library and Library of Congress endorsing it as the archiving standard for electronic content, the NLM DTD is clearly here to stay. Although the NLM DTD is currently only an unofficial standard, it is now starting to move towards NISO registration over the coming few years.

The conclusion drawn from the meeting is clear. The answer to the question "Should publishers adopt the NLM?" is:

- yes, if you publish any kind of journal, even in the humanities
- Possibly not for books – in that case, you might find the TEI more suitable.
- Although created for archiving (the "green" NLM), the NLM is now widely used for content creation (the "blue" NLM) as well.

The day included not just advocates from organisations involved in setting up the NLM DTD – Bruce Rosenblum of Inera, and Bill Kasdorf of Apex CoVantage – but also had time for a fascinating case study on the digitisation of the archive of The Institute of Mechanical Engineers, which uses the DTD for the 22,000+ articles in its repository, dating back to 1847. This project faced the typical challenges that any organisation would

have when attempting the commendable task of making the archive more widely available. In an age of increasingly tightly specified metadata, how do you shoehorn content from a different era into a 21st-century framework? Chris Baker, who managed the project, revealed that even scholarly journal articles presented problems in digitisation, since before 1945 they had no abstracts. Eventually the decision was taken to add abstracts to all of them. The archive can be viewed at <http://www.imeche.org.uk/ils/archive.asp>.

Michael Upshall