

Is Schema.org about a technical standard or about something else?

Martin Fenner, Gobbledygook

June 19, 2011

At the beginning of the month Google, Bing and Yahoo announced schema.org, a new initiative for structured markup on the web. Richard MacManus responded with a critical piece at ReadWriteWeb (Is Schema.org really a Google Land Grab?), mainly criticizing that the initiative didn't use RDFa and didn't seem to have consulted with the web standards body W3C. The best discussion about schema.org that I found so far (via Eric Hellman's post) is by Henri Sivonen (Schema.org and Pre-Existing Communities). He explains why he thinks it makes sense that schema.org picked microdata over RDFa or microformats as markup standard. And he discusses how the web community has developed standards in the past and what characterizes the successful efforts. I personally believe that schema.org will bring us closer to the semantic web, and the discussion reminds me a little bit of the discussion around HTML 5 vs. XHTML 2 (where the pragmatic solution supported by the major browser vendors won over the ideal but dogmatic solution discussed for many years in standards bodies). Schema.org is will certainly be discussed at the Science Online London Conference in September.

Whether schema.org becomes a success depends in large part on the adoption by the community. Schema.org is backed by the three largest search engines (Google, Bing, Yahoo), so everybody who wants to be found by them will take a close look at the standard. Scholarly content is for the most part collected and curated in specialized databases. This makes it not only a technical decision whether or not to adopt schema.org for markup (we need more and better defined scholarly data types in the standard), but also a business decision. If scholarly publishers start marking up their journal articles, they will be easier to find via Google, Bing, etc. But will this mean that institutions would start re-evaluating their subscriptions to commercial services like Scopus or Web of Science? And will schema.org help or hinder the very large market of discovery tools for scholarly content?

At this point this is very much a theoretical discussion. We first need better scholarly data types in the schema.org standard. The Association of Educational Publishers and Creative Commons has announced an initiative to create a

metadata framework for educational resources based on schema.org. And we hopefully see a similar initiative from scholarly publishers. I hope that the group behind schema.org is open to this input.