



TOC: What happened in the past five years to divert XML from its original use, and how does this affect plans for your content today?

Content-centric XML: Coming Soon to an Intranet Near You?

ROBERT J. BOERI

Content-centric XML hasn't followed its original five-year script. Celebrating its fifth birthday as a standard last February, XML was supposed to supplant HTML, shift the burden of processing Web sites from servers to underutilized client PCs, and achieve the holy grail of "create once, reuse many times." Although use of XML to transfer information between applications was one of the World Wide Web Consortium's original goals, emphasis was on content-centric XML: Web pages and documents. What happened in the past five years to divert XML from its original use, and how does this affect plans for your content today?

Imagine a worldwide content network where every piece of content on one network node was understandable by every other node. To do this, all content must be intelligible to each node, even though the nature of the content in this networked universe varies widely. Each node's content would need its own vocabulary, with a common and shared grammar and syntax. XML provides content with this foundation. In its brief lifetime, XML has inspired and refined hundreds of specialized vocabularies or document models. Expressed in document type definitions or schemas, these models are often called "schemas," ranging from obviously practical models like MathML to niche esoterica. I recently found a niche XML model to analyze (put delicately) swine waste. (My pig-farmer grandfather must be chuckling in the hereafter.) The exponential growth of schemas has made XML difficult to assimilate, and dozens of related XML standards have also slowed product offerings. And most users still struggle with ordinary office tools like MS Word. If we still need to ask a buddy or call the HELP desk with MS Word problems, why spend a pixel's worth of ink on content-centric XML? Should you develop an XML strategy for your intranet, where most of your enterprise content probably resides?

For openers, the availability and cost of getting started has dropped dramatically. XML authoring tools have gone from pricey with expensive customization to low-priced or free and increasingly mainstream often working via familiar browsers. Microsoft, with its sixth sense for maturing markets, will support XML in its upcoming Office 11 suite. Lastly, although XML originally emphasized text content, multimedia use is also increasing, especially on

intranets. And now here's the XML-intranet connection.


Intranets often provide employees with external newfeeds. It's easy linking to these feeds, if you're satisfied with employees jumping outside the firewall or viewing them in a pop-up window. If that's all you want, then basic HTML (or XHTML) works fine. But if you'd like to store that news locally, index and search it, or—gasp—contribute new content expressed in XML, consider NewsML. "News Markup Language" is an XML schema conceived by Reuters, developed and ratified by the International Press Telecommunications Council, and increasingly a standard for composing and delivering news. NewsML provides a way to produce news and maintain its metadata, and it supports text and rich media. Intranets can use automated processes to deliver NewsML content to a wide variety of devices such as financial service desktops, Web sites, and mobile phones. Journalists can write news stories using standard XML authoring tools in several languages. Altova, for one, supports NewsML in its products, including its free "Authentic." CambridgeDocs recently announced that its xDoc XML Converter works with Altova's products to convert legacy content to XML. Although NewsML isn't a one-size-fits-all model (no model is), its adoption is growing both by news organizations creating syndicated content as well as intranets delivering that content.

NewsML work its magic with a carefully conceived schema that packages news items, regardless of language or media type, with robust metadata. News "envelopes" contain one or more news items, which in turn contain one or more components in one or more written or spoken languages. Envelopes describe items with attributes like date and time sent, news service, and priority. Text, images, video, and sound can be packaged in an item as hyperlinks.

And Gregor Geiermann, a consultant with NetFederation Interactive Media, actually uses NewsML and has success stories to tell: Aventis, a European pharmaceutical firm, upgraded its content

Why spend a **pixel's worth of ink**
on content-centric XML?

to NewsML and now uses a huge corporate intranet with dozens of portals, databases, and content management systems around the world working via one central corporate newsroom. A large European banking consortium uses NewsML to centralize its corporate communications and guarantee a seamless flow of information to affiliate company portals. Geiermann says, "Editors type their stories using an office-like WYSIWYG Web interface. The future of XML authoring will be browser-based editors."

Yet this is but one picture of XML success: Controlled, disciplined structure; multilingual; support for rich media; controlled metadata, without forcing content creators to become XML geeks. Happy fifth birthday, XML. You got your wish. 

ROBERT J. BOERI (bboeri@ieee.org) is a knowledge management analyst for a Boston-area biopharmaceutical firm.
Comments? Email letters to the editor to eclatters@infotoday.com.