

Content-centric XML: Where We've Been, Where We're Going in 2003

ROBERT J. BOERI

Genuine XML skills are so uncommon—and XML use still relatively rare because of a combination of factors, none having to do with the **benefits of XML** itself.

as *EContent* readers, you likely develop, manage or deliver content. In doing so, increasingly you encounter the three-letter acronym XML, short for “eXtensible Markup Language.” This is true whether your content is only text, graphics, streaming media, voice, or any combination thereof. Read resumes today and you’re likely to see claims of XML skills, but when you press candidates to learn what they mean, most inadvertently cite “data-centric” XML experience—the use of XML to interchange information between systems, often for ecommerce or Web applications. One way to determine if candidates are familiar with XML for content is to ask how they viewed or developed XML models. If you get a blank stare or a response like “Notepad,” the candidate is probably familiar only with data-centric XML, if that. The use of data-centric XML has in fact taken the software world by storm but is only one of the originally intended uses of XML. *Content-centric XML*—or simply “XML” in the rest of this column—is likely to be of more interest to you, and such XML requires conformance to a document model expressed as a document type definition (DTD) or, increasingly, as an XML schema. Schemas can express and enforce far richer models, like “date,” which DTDs express only as text.

Genuine XML skills are so uncommon—and XML use still relatively rare because of a combination of factors, none having to do with the benefits of XML itself. The biggest hurdles are first, “legacy” content—all those MS Word or FrameMaker files to convert to XML for a specific DTD or schema; the second, organizational resistance to change. Data-centric applications are new, and therefore usually do not have the same burden of legacy conversion. Yet most who understand XML freely admit that its benefits are compelling: Create content that doesn’t depend on any one vendor, and produce many products—Web, ebook, print, or even voice—from a one-time investment in single-source content. If committing to XML is so compelling yet so rare, what happened this year to change that, and will 2003 be the year to explore upgrading your content and business processes to XML? Let’s look at what happened this year in the premier XML standards body, the World Wide Web Consortium (W3C) and others like Organization for the Advancement of Structured Information Standards (OASIS). Then we’ll review how vendors are responding to W3C standards with new product initiatives and hazard a guess about 2003.

SETTING STANDARDS

The lag from standards to products usually takes several years, and the first XML stan-

dard was released in February 1998. Vendors also work on W3C committees, so looking at W3C activity provides hints of vendor product directions, with a time lag. This year's W3C XML standards cluster into roughly four areas: XHTML (modularization, multimedia), format independence (upgrades to the Web formatting capability called Cascading Style Sheets), semantic Web, and multimedia (speech and speech recognition, voice browsers). XHTML is essentially HTML following XML syntax; this year, standards efforts emphasized developing subsets of XHTML for use in devices like wireless phones that can't support a desktop Web browser. The semantic Web, a strong interest of the W3C's leader Tim Berners-Lee, aims to facilitate automated collaboration between Web-connected devices.

Other standards organizations, most notably OASIS, focused on designing and delivering XML standards for interoperability. One example is DocBook, designed for content describing computer hardware and software although by no means excluding other uses. Other standards deal with special industry needs like legal contracts and automotive repair information. XrML is a digital rights management grammar for expressing rights and conditions associated with digital content or services. In April, ContentGuard Inc., owned by Xerox with a minority Microsoft interest, submitted XrML to OASIS and other consortiums. ContentGuard wants to encourage development of an open digital rights standard that meets everyone's needs regardless of industry, platform, or type of media.

PRODUCTS

This year vendors began offering powerful XML management tools, targeting less sophisticated users, often with an Internet Explorer browser interface. In the past, Tibco and Altova offered integrated XML development environments. Until 2002, Altova's only product, "XML Spy," was geared towards sophisticated users. In 2002, Altova upgraded and renamed its flagship product xmlspy 5, and began

offering two additional products for business users: authentic 5, a browser-based authoring tool for creating XML content, and stylevision 5 for Web developers to migrate their Web sites from traditional HTML to XML, often via XHTML. stylevision is the first such XML-centric tool for client-Web developers, and this could be revolutionary. To gain traction beyond early adopters, stylevision must make clear what XML offers that is unavailable with traditional client technologies like JavaScript.

Ektron Inc. offers "business-user-driven" Web content management solutions. Ektron also offers browser-based authoring for both HTML and XML content; its newest XML authoring product is Ektron eWebEditPro+XML. These products too are geared towards business users, not HTML or XML experts. Adobe's FrameMaker is geared towards technical writers and, in 2002, Adobe upgraded this product to full XML support. Corel, a vendor with consumer experience, acquired SoftQuad and its XMetaL tool. Finally, Arbortext, long a leader in SGML and XML, delivered a major upgrade of its Epic XML multichannel publishing system with full support for both DTDs and schemas.

LOOK INTO THE FUTURE


"Predictions are difficult—especially about the future." Whether it was Yogi Berra or Niels Bohr who said that, it certainly is true of XML. Waiting for XML's adoption has been like waiting for Godot. With increasing XML support for business users, and general awareness of XML's benefits, 2003 may be the year when content-centric XML gets as much support as does the data-centric variety. If this happens, it will be in part because vendors understand that general business users constitute a far larger market than the early technical adopters.

"Semantic Web" is less likely. My hunch says it fails the 80-20 rule. Give us 80% (or a little less) of the capabilities with 20% or less of the effort or complexity, and that's the solution most will choose. Search engines—especially Google—are

getting increasingly clever at sensing the best results for search requests, and Google is software sans "Semantic Web." Moreover, the sheer number of legacy Web pages, and the complexity and number of XML vocabularies, makes me say: Not in 2003 and, except for certain vertical industries, maybe never.

Consumer resistance to paying for Web content and for restrictions on their ability to make "fair use" copies of content from one medium to another suggest that XrML for consumer applications will still face tough sledding in 2003. XrML for business applications may come sooner, but as features of suites, which may explain Microsoft's stake in ContentGuard.

Vendors think 2003 may be a watershed year for XML authoring systems. In late September, PG Bartlett, Arbortext's vice president of marketing, told me "Reuse and automation are the key requirements driving most businesses to commit to XML content management solutions." Initiatives such as Arbortext's focus on the life sciences are key. Arbortext is targeting new drug applications, product information (package inserts, leaflets and labels), and written documentation of procedures (so-called SOPs) used in the preparation, processing and packaging of pharmaceutical and medical products. The benefits? Not only faster time to market, but less exposure to FDA penalties. One large company expects a benefit of 30 times the cost of this solution. This focused application of XML with specific benefits is what I believe will drive industry acceptance.

That's the view of content in late 2002. Arguably, content without knowledge is data without information, and content is useful only as a foundation for knowledge, the ultimate enterprise intellectual asset. But knowledge management, itself increasingly exploiting XML, is another column for another day. 

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