Article

The Birth of XML

By Jon Bosak
Solaris Global Engineering and Information Services

XML arose from the recognition that key components of the original web infrastructure -- HTML tagging, simple hypertext linking, and hardcoded presentation -- would not scale up to meet the future needs of the web. This awareness started with people like me who were involved in industrial-strength electronic publishing before the web came into existence.

I learned the shape of the future by supervising the transition of Novell's NetWare documentation from print to online delivery. This transition, which took from 1990 through 1994 to implement and perfect, was based on SGML. The decision to use SGML paid off in 1995 when I was able single-handedly to put 150,000 pages of Novell technical manuals on the web. This is the kind of thing that an SGML-based system will let you do. A more advanced and heavily customized version of the same system, built on technology from Inso Corporation, is used today for Solaris documentation under the name AnswerBook2. You can see it running at *http://docs.sun.com*, which looks like an HTML web site but in fact contains no HTML; all of the HTML is generated the moment it's needed from an SGML database. (You can get XML from this site if you know how -- but that's another story.)

Like many of my colleagues in industry, I had learned the hard way that nothing substantially less powerful than SGML was going to work over the long run. So from the very earliest days of the World Wide Web Consortium, there was a small group of us who kept saying, "You have to put SGML on the web. HTML just won't work for the kinds of things we've been doing in industry."

Now, the people in charge of the W3C were far from ignorant about SGML. Dan Connolly, in particular, saw very early the need to standardize HTML itself as a proper SGML language, and by the beginning of 1996, he had created a placeholder for some future SGML work within the W3C. But W3C didn't have the resources to pursue this direction, and outside of the few of us who had already been through the development of large-scale electronic publishing systems, no one else really understood the problem.

I had been pestering W3C about SGML and about DSSSL, the SGML stylesheet language, right from the beginning, while I was still working at Novell, and I kept this up after I went to work for Sun. Finally, in early May of 1996, Dan challenged me to put Sun's money where my mouth was -- to organize and lead a W3C working group to put SGML on the web. This was an unprecedented offer, because up until then, all W3C working groups had been organized and run by W3C staff. Dan's willingness to go beyond established practice was the first key development in the process that led to XML.

Dan's offer came just as I was beginning a three-week series of WWW, SGML, and ISO conferences in Europe. This tour put me in touch with just about everyone I needed to talk to about the idea, and by the time I got back home, I had managed to recruit some of the world's leading SGML experts for the "Web SGML" initiative and had secured funding from my management at Solaris Global Engineering and Information Services to carry out the work. This was the second critical turn in the path to XML. Many people know that XML grew out of the expertise of the SGML community, but few people realize even today that the whole two-year effort to develop XML was organized, led, and underwritten by Sun.

It was obvious from the beginning of what was originally called the Web SGML Activity (the name XML was suggested by our technical lead, SGML/DSSSL guru James Clark, several months later) that it would need the support of at least one of the two major vendors of web browsers. In June of 1996 I succeeded in persuading Jean Paoli of Microsoft to join the working group. This turned out to be especially important, because in addition to his SGML expertise, Jean was eventually able to convince Microsoft to adopt the technology.

The basic design of XML was accomplished in eleven weeks of feverish activity under the guidance of editors Tim Bray and C. M. Sperberg-McQueen. The work started in the last few days of August, 1996, and ended with the release of the first XML draft at the SGML '96 conference in November. While it took another year to finish working out all the details, virtually every basic feature of XML as we know it today was specified in that first published draft. This remarkable achievement is a tribute to the team spirit and world-class expertise of the original design group. I am proud to have had the honor of leading this group and proud of my management at Sun for having had the vision to underwrite the effort.