



The image shows the top navigation bar of the EMedia website. It features the 'EMEDIA' logo in a large, bold, black font. Below the logo is a horizontal menu with several colored buttons: 'CURRENT ISSUE' (orange), 'SUBSCRIPTIONS' (yellow), 'EDITOR'S CHOICE' (yellow), 'EDITORIAL CALENDAR' (yellow), 'INFOLINK' (yellow), and 'CONTACT' (yellow). Above the main menu are four smaller buttons: 'EMEDIA LIVE!' (green), 'EMEDIA MAGAZINE' (yellow), 'DVD PRO' (orange), and 'ONLINE INC.' (blue). To the right of the main menu is a search box with a 'Sea' button. Below the search box is a link that says 'Browse by Subject'.

information insider

Making the Commitment to XML

Robert J. Boeri
Martin Hensel

EMedia, March 2000
[Copyright © Online Inc.](#)

Let's face it: If you knew SGML, you probably didn't like it. The "mother" of HTML and XML, SGML was huge (a 300-page specification), it had all the charm of an alpha personality, and was complicated, difficult, uncompromising--SGML never did learn to put on a stylish face. DSSSL, the standard applying style to SGML content, never met with commercial success. Since then, you've had a long-term relationship with HTML, but are finding its superficiality no longer satisfies. Recently you've been attracted to XML, flirting with the idea of getting to know it better. Now you're ready to take the plunge. XML seems to combine the best elements of SGML and HTML. XML publishing is far more disciplined and more powerful than HTML, but it seems less demanding than SGML. What would a commitment to XML be getting you into? Just this: XML requires a commitment to serious discipline, and likely a fundamental change in the way you do business. But with everybody wooing XML, those prenuptial agreements may be getting easier to accept.

Before we put down HTML, let's not forget: HTML's simplicity and flexibility spawned one of the biggest inventions in history--the World Wide Web. HTML won't go away any time soon, and HTML (like SGML) is based on Document Type Definitions (three to be exact). HTML was the dating phase of markup languages, and it charmed browser vendors into forgiving Web pages their sloppy code. Nonetheless, HTML's teen blemishes are becoming obvious. Just look at all the plug-ins and proprietary extensions demanded by a world that wants to go beyond brochure-ware and just another pretty HTML face. Enter XML, much leaner than SGML (only a 30-page spec), yet promising discipline and flexibility impossible in HTML. XML insists that valid documents conform to a Document Type Definition (DTD) or document model, yet retains the flexibility you want and need. Follow the 30 pages of the XML specification, and you can develop a model with practically any combination of elements, attributes, and entities you choose.

Once XML invites you home for dinner, you're surprised to see how many standards sit at the table. You recognize SGML and HTML immediately, but who are all these others? Why there's the Document Object Models (level 1 and level 2), the Multimedia cousins (HTML+Time, Synchronized Multimedia Integration

Language, SMIL-Boston Integration Language), and Namespaces. There are also standards for linking, querying, pointers, and even XHTML (Extensible HTML) to help you transition from dating HTML to an engagement with XML.

Not far down the table, and obviously uncomfortable, you notice XSL, the Extensible Stylesheet Language, who introduces itself as "just like DSSSL, only simpler." The other specifications snicker that XSL is "just like DSSSL, only later; it's still just a working draft." At this you shudder, wondering if SGML's troubles will recur. It's true that XML is only about 30 pages long, but if you count up all the pages in all the other specifications, it's even bigger than SGML's 300-page count. And you don't even want to go into the next room where an even larger number of XML cousins await: Dozens of Markup Languages (MLs, really DTDs) that every industry on the planet seems to be incubating. It's going to be really tough getting acquainted with this extended family.

What about life after infatuation? You're all for discipline, but will you really have to give up your word processor and WYSIWYG Web development tools?

Some vendors promise you can keep your old, favorite word processors in the new XML world. Right, and you'll be able to move away from all your crazy XML in-laws and never have to go visit them either. Sure, you can use MS Word and just map the styles to Docbook DTD elements (or pay extra for customizing to another DTD). But nobody really pays attention to styles, and word processors let you use (or omit) any style anywhere. In XML publishing, documents need DTDs. And yes--XML DTDs may be simpler than in the old SGML days, but DTDs are essentially the same. Every XML authoring tool--some very reasonably priced and relatively quick to deploy--needs a DTD. And DTDs have their own lifecycle, also requiring documentation and training. You'll need to analyze your documents, assure they're consistently structured, then either buy a DTD development tool or pay someone else both to create and then maintain your DTDs. You may even need a document management system to keep track of all the versions, models, and content. And how will you get all your legacy content into XML?

By now, you're probably getting typical commitment cold feet. Was XML such a great idea after all? One thing to remember: the "good old days" were not so good either: tossing WP files over to typesetting, having to manage and proofread the results in a cycle that often ended only because you had to ship the product. And there is no easy way to unlock the potential of multiple Web and CD products from that content that you know could be paying big dividends. As CAP Ventures predicted, "The Web content market topped \$1 billion for software development tools in 1998, and is expected to grow at 94 percent per year for the next several years." XML is central to that growth. XML is irresistible, and if you don't link up with XML soon, you may be left behind. Nonetheless, remember Shakespeare's warning: "The course of true love never did run smooth."

Robert J. Boeri (bboeri@world.std.com) and **Martin Hensel** (mhensel@texterity.com) are co-columnists for Information Insider. Boeri is an Information Systems Publishing consultant at a Boston-area insurance company. Hensel is president of Texterity, Inc., a Newton, Massachusetts-based consulting firm that builds SGML-based editorial and production systems for publishers, corporations, ecommerce services, and type-setters.

Comments? Email us at letters@onlineinc.com.

[\[EMedia Home\]](#) [\[Subscriptions\]](#)

Copyright © 2000, Online Inc. All rights reserved.
info@onlineinc.com