



Working with XML

This column focuses on XML and its use in the technical communication field. It explores the benefits/promises of XML, but mostly how to get XML to deliver on these promises. Please send your comments, questions, and suggestions to france.baril@ixiasoft.com.

FRANCE BARIL, Column Editor

A Brief Tour of XML

BY FRANCE BARIL

Before we begin digging into XML, I want to make sure that we start with the basics: What is XML? Why should we care?

Next, we will explore the “how” in order to develop a critical approach, build realistic expectations, know where to spend our time and energy, and know how to discuss features with our development team or with the external folks we hire to build our in-house system. We might also discover that we can get some of XML’s main advantages with our current tools. So let’s start with the basics and the best-known advantages of this technology.

Reading this column should help you understand what XML can do. Future columns will help you understand the technical aspects, but always with a high-level approach.

The goal is not to make you a programmer, but to help you understand how XML relates to what we, technical communicators, do every day.

What Is XML?

XML (eXtensible Markup Language) is a markup language designed to describe information. It is said to be extensible because users can define the markup used to describe content. XML is not made of predefined tags; rather, it helps you build your own tag sets. Basically, XML helps you build your own information/document model based on semantics (the content’s meaning). Moreover, because of its predictable form (XML has a specific syntax that

you must follow), it is easy for computers to manipulate.

Why Use XML?

XML describes data. People use this ability for different purposes. The eBusiness industry has adopted XML because it eases the process of sharing and exchanging data between otherwise incompatible systems. Transporting and exchanging data via XML saves developers a lot of time and money. For the technical publication industry, XML is useful for the following reasons:

Separating content from presentation. XML lets the content creator identify content by its meaning rather than by the way it should be displayed.

In contrast, HTML is based on presentation. The tags used in HTML define how the document *appears* to the user. For example, `text` means bold, `<i>text</i>` means italic, `<h1>text</h1>` means the content will be displayed as a level 1 title—usually bigger and bolder than the regular text.

In XML, elements usually identify what the content *is*. For example, a button name may be identified as `<button>button name</button>`. Only later, when it is time to display the content, will the designer decide whether buttons will be displayed in bold, italic, or regular text. Further, the same `<title>` element may be displayed as a heading 1, 2, or 3, depending on where it appears in the document. The value of this capability will be obvious to everyone who has added a level to their document hierarchy, only to realize that they

have to apply new headings to all titles below that new hierarchical level.

Separating content from presentation is the basis for many other advantages of XML, like the ability to reuse and repurpose content, or to publish information in multiple output types.

Reusing content. XML enhances the writer’s ability to create chunks of information once and use them in multiple documents. There are multiple strategies for content reuse, and most of them are not technology-related; they depend on your ability to develop a content structure that allows for reuse at different levels. XML offers the mechanisms to implement these strategies because it is easy for computers to manipulate.

Repurposing content. XML enables writers to use content in different contexts—that is, to use content for a deliverable with a different purpose than the original one. For example, the same content may be used in a reference guide, a training manual, and a marketing brochure. Repurposing depends greatly on your ability to develop a content structure that facilitates content manipulation in all projected contexts.

Creating content automatically. XML facilitates automatic content creation by supporting the creation of systematic content models or content structures and then reinforcing them. Automatic content creation usually requires planning ahead to organize information in a way that will allow the system to find it and put it together with other information to create new pieces of content.

The simplest and most popular generated documentation pieces are overviews, summaries, learning objectives, and sometimes application program interfaces (think of something similar to *Javadoc*).

Creating multiple output types. Multiple output creation is easier because XML may be transformed, provided you have a proper tool set, into almost any format. The most popular output types are HTML, XHTML, CHM, PDF, RTE, TXT, XLS, or even different XML structures.

Enhancing consistency in form and content. Consistency in content is easier because you can use a predefined tag structure or a document model that reinforces your information model. Moreover, formatting is done by the computer, following predefined formatting rules.

Because formatting is left out of the creation process, nobody gets to tweak the presentation and different people cannot apply different formatting styles to elements. Enhanced consistency facilitates not only computer processing and authoring, but also usability, as consistent layouts and information structures usually mean that readers can find information more quickly.

Enhancing time to market. This advantage of XML results from other advantages, such as content reuse, automatic content generation, machine formatting, workflow support, and process automation.

Enhancing search capabilities. Because XML documents are usually tagged semantically, search engines may be integrated at multiple levels: in the authoring environment or in Web deliverables, provided you get an XML-compatible search engine.

How is search enhanced? Well, just imagine searching for a button name in an HTML document. You can search for the name of the button, or if you have applied bold tags to all button names, you can search for the button name in a bold tag. You'll find the button name, as well as a few other things. But if the information was tagged semantically, you can target your search for the button name in the button tag—and find it much more quickly. Your document

structure should include tags for the information that you or end users might want to search for.

What It Comes Down To

As we have just seen, XML offers many advantages in our industry, and we have only scratched the surface. We have not talked about how reuse impacts costs in content creation, translation, and content update. We have not discussed how XML can facilitate collaboration, workflow management, and metadata management.

You might have noticed that the recurrent theme in all these explanations was *planning*. Moving to XML implies knowing what you want to get from the technology. Once you do, you can choose or create information models and build processes that will actually bring you where you want to go. The technology itself won't do it alone.

There is an overhead to using XML. You need to plan your information struc-

ture, the reuse and repurposing that you want to do, and so on. You also will need, at some point, to develop the content model in a language understood by the machine (DTD, XML schema), to learn to create content in XML, and to get some transformation processes to create output.

In the future, we will explore each of the advantages more closely, to learn what is needed to get the promised benefits and where we fit in as technical communicators. ❶

France Baril has a unique background with a B.A. in communication from the University of Ottawa and a B.Sc. in computer science from L'Université de Sherbrooke. She is a documentation architect at Ixiasoft where she helps various clients enhance their documentation processes with XML-based solutions. France is vice-president of STC's Montreal chapter and a founding member of the DITA technical committee at OASIS. You may reach her at france.baril@ixiasoft.com.

Translating Your Company Into a Global Success

Software Localization

Documentation Translation

Web Site Localization

Technical Writing and Editing

Desktop Publishing

Additional Customized Services

International Communication by Design, Inc.

1726 NORTH 1ST STREET MILWAUKEE, WI 53212
TEL: 414.265.2171 • FAX: 414.265.2101
EMAIL: INFO@ICDTRANSLATION.COM
WWW.ICDTRANSLATION.COM

ata
TRADOS