XML, CMS, and other Acronyms A plain language - crash course to our web publishing.

XML: Stands for "Extensible Markup Language" and is a flexible away to create standard information formats for the World Wide Web. While XML is based on the same parent technology as HTML, the true value of XML comes from its ability to organize and manage information. In other words, an XML file not only contains the content of an article (body copy, images, listed items, etc), it also contains information about how to manage that information.

DTD: A DTD or "Document Type Definition" is a required component found within each XML file. It's purpose is to define what information should be and can be included in the XML file. We create this DTD based on our needs as a publisher. For instance, in our DTD there is a requirement that the "Issue Date" of the article must be defined somewhere in the XML file. If the Issue Date is not declared, the XML file is considered "invalid" and cannot be used.

CMS: CMS stands for "Content Management System" and is the tool we use to manage and create our web sites. Each of these web sites are comprised of multiple "components" and can be added, subtracted, or modified from within our CMS. An example of a component would be the Navigation bar on the left hand side of the page or the Banner Ad at the top of the page. The advantage of using such a system is that anyone can modify a site without prior knowledge computer coding.

So why XML and not HTML? HTML is a very powerful and common computer language used to create web pages. However, most of the codes used to create an HTML page deal with how the page should be displayed and doesn't give any information about what the page contains or how the data should be managed. XML on the other hand was designed for the specific purpose of Managing Data and has very little to do with how the information should be displayed. We use other programs and processes to determine how the information contained in the XML file should be displayed. (See "XML and our CMS" for more information)

Another drawback to HTML is that it is governed and limited to a Global Standard maintained by an organization called the W3C. This means that all HTML code must follow the W3C standards and will not work if we deviate from it. On the other hand, we can customize all the code found in an XML file. These rules are located in the DTD that we have created for our system and can be changed to fit our needs at any time. Another advantage to using XML is its ability to be re-purposed to other programs and organizations. The XML that is used in our web sites can also be used in a database for quick searches or another company that wants to use our data for their web sites. Who is RSB and what do they do for us? The ultimate goal is to have a computer program that will convert our print articles to XML articles at the click of a button. There are programs showing promise, such as Quark 6 and Adobe Indesign. But at this time, there is no "silver bullet" to tame that beast. In the mean time, we're using RSB Systems to manually convert the articles from Quark Files to XML files based on our DTD. Here's the process:

- 1. We receive a Specification sheet from the editor of the magazines that contains what articles and information should be converted to the web.
- 2. Web ops then double checks the info on the spec sheet and then forwards the spec sheet to digital pre-press.
- 3. Digital pre-press then collects the files on the spec sheet and sends them to RSB via Mass Transit.
- 4. RSB then converts the files from Quark files to XML files based on our DTD.
- 5. After RSB has converted the files, they are uploaded in bulk into our CMS system on a pending status.
- 6. Web ops then goes over the files, checks for technical errors, and uploads any missing articles.
- 7. We then notify the editor of the magazine that when all files are ready for editorial corrections

XML and our CMS? So, now that we have our XML articles created and our web site created in CMS... how do they work together to display the content? Answer: MAGIC! Well, maybe not magic, but the process is actually pretty cool. I've simplified the steps below to help clarify the process:

- 1. An XML article file has been created using our DTD as a guideline.
- 2. We then load the article into our CMS one by one or in large groups.
- 3. At the time of loading the article, the CMS looks at the different elements within an XML file and pulls out the information it needs for display and storage. These elements would include the Article title, Author, Issue Date, Content, etc.
- 4. Once the CMS has pulled all the information it needs from the XML file, it then places that information into a component of the CMS for display on the web.
- 5. Now, since the XML files contain very little information about how the information is to be displayed, the data needs to go through one more process before someone can view it. The information such as the title, the body copy is filtered through something called a CSS or "Cascading Style Sheet". The CSS tells the your browser how the information should look. For example, the title should be bold and at 14 pt, and the body copy should be 10 pt and italicized.
- 6. Finally the content is displayed when the visitor to the site opens up an article.