Math on the Web: A Status Report

January, 2003

Focus: Adding Value for STM Publishing

by Robert Miner and Paul Topping, Design Science, Inc.

To view this paper online, where the links and references are live, go to **http://www.dessci.com/webmath/status.**

We plan on updating this report as the world of Math on the Web changes. Join our Math on the Web mailing list and we'll notify you when the report is updated: **http://www.dessci.com/webmath.**

Design Science
How Science Communicates™

Design Science, Inc. • 4028 Broadway • Long Beach • California • 90803 • USA • 562.433.0685 • 562.433.6969 (fax) • info@dessci.com • www.dessci.com

Math on the Web: A Status Report

Focus: Adding Value for STM Publishing

by Robert Miner and Paul Topping

Over the last half year a tidal shift has taken place in the state of math on the web. High-quality math support is now available in Netscape, Mozilla and Internet Explorer browsers. In August, Netscape 7[1] was released with native math support on the Unix and Windows platforms. Math support on the Mac platform came about the same time in Mozilla 1.1[2], the open source sister browser of Netscape 7. MathPlayer[3], a free extension for Internet Explorer[4] from Design Science, completed the field in September by adding native-quality math support to the web's most widely deployed browser.

Better browser support for math has significant implications for Scientific, Technical and Medical (STM) publishers. Most scientific web publication is presently done using PDF[5]. Now it is also feasible to effectively publish technical content in HTML + MathML format, which may turn out to be cheaper and better in many situations. In response to increased interest, several new MathML-capable tools have been announced catering to the STM market. This edition of the Status Report takes a closer look at these tools.

The HTML + MathML Platform

The technology underlying the new browser support for math is MathML[6], a World Wide Web Consortium (W3C)[7] Recommendation for encoding mathematics in XML format. MathML is designed to be used in conjunction with some other document-level markup language, since it is only for encoding mathematical notation. For web publication, the natural document markup language to use is XHTML, the XML-compatible version of HTML. In the publishing arena, DocBook and a few other XML-based document markup languages are also common.

Of course, a modern web page consists of much more than static HTML code. Web pages are frequently dynamic, with features like menu highlighting implemented in JavaScript. Some pages utilize applets and plug-ins. Pages may also use style sheets to keep content and visual presentation separate for easier maintenance and management.

In past editions of the Status Report, we have dubbed HTML, MathML and related web technologies the HTML + MathML platform. Conceptually, the HTML + MathML platform is made up of three parts — markup languages for encoding content, stylesheets for controlling how content is displayed, and a programming model for making the page dynamic. The primary markup languages involved are HTML[8], XHTML[9], MathML, RDF[10] for metadata and SVG[11] for structured graphics. The most important stylesheet languages are XSL[12] for transforming XML into HTML, and CSS[13] for visual style information. For web programming, the Document Object Model (DOM) Recommendation[14] and JavaScript form the cornerstone, but many other technologies may be involved, such as applets, plug-ins and so on.

Standards work at W3C and browser implementation have been converging toward the HTML + MathML platform for some time. With MathML support in major browsers under Windows, Unix and MacOS, the HTML + MathML platform achieves a new level of viability. Moreover, because all the component technologies were designed around a common XML-based conceptual framework, they work well together in XML-based workflows. XHTML + MathML web pages now offer a strong alternative to PDF for many kinds of technical web publication. As a result, there has been a surge in MathML interest and activity as people take a closer look at the new developments.

MathML Gains Momentum

MathML is relatively venerable as web standards go. It was released in 1998, just a couple of months after XML itself was completed, and before CSS2 was available. In the intervening time, a number of scientific software packages have added support for MathML, and several applets and plug-ins have long been available for displaying MathML in browsers. However, these early display technologies were not robust enough or well enough integrated with browsers to be a viable option for demanding applications. So even though interest in better solutions for math on the web remained high, most individuals and organizations adopted a wait-and-see attitude after initially investigating MathML.

But now the wait is over, and people are liking what they see. Unlike earlier technologies, Netscape/Mozilla and MathPlayer are getting high marks in the marketplace. MathPlayer was chosen as a Hot Pick[15] at the Seybold 2002 conference for professional publishing. Judging by their initial reception, MathPlayer and Netscape/Mozilla promise to change the landscape for math on the web dramatically.

In Netscape 7 and Mozilla 1.1, MathML support is built into the rendering engine. In speed and quality, it is comparable to the rest of the browser text. Because it is built in, users don't need to download a separate plug-in. However, many users find they need to download and install math fonts.

In Internet Explorer, MathPlayer provides MathML support. MathPlayer utilizes powerful, low-level extension capabilities called *behaviors* only available in the Windows version of Internet Explorer. However, by utilizing behaviors, MathPlayer achieves high-performance, native-quality rendering and seamless browser integration. MathPlayer is installed by downloading a standard Windows installer. The installer also includes the fonts needed by MathPlayer.

Even with much improved support for MathML in browsers, some technical challenges still remain. Because of differences in the way in which Netscape/Mozilla and Internet Explorer handle XML documents, in practice many people find it necessary to publish HTML + MathML documents with an XSL stylesheet that customizes the document to the browser. Nonetheless, the new math support has recently spawned a variety of interesting, experimental projects. Two representative examples are an online formula finder[16] and an open source MathML stylesheet archive.[17] Postings to newsgroups indicate that although dealing with browser differences is still a painful subject, the technical problems are surmountable.

Of course the acid test for MathML support in browsers is whether it is adequate for large-scale publication of technical information, such as scientific journals. To be credible as a solution in that arena, a candidate technology has to demonstrate that it looks good, renders fast, and prints well in a browser, even for long, dense, research articles. But indications are good that MathML support in browsers now largely achieves that goal.

Conventional wisdom holds that investment in new technology drops off in a down economy. Consequently, the mere fact that the HTML + MathML platform in now feasible is not what has attracted attention from STM publishers. Rather, it is because MathML is both information-rich and XML-friendly, and thus it presents a number of enticing possibilities for cutting production costs and adding value to technical documents.

New Interest in MathML from STM Publishers

All for-profit businesses seek to cut costs and increase sales. For STM publishers, MathML has appeal on both fronts. Publishing is a very mature industry, and thus finding ways to innovate and differentiate a brand or product is a major challenge. On the surface, the prevalence of the web offers publishers a fertile new arena to work in. However, for many STM publishers dealing with highly technical material, the web has been a problematic medium in practice. Users easily come to take for granted online versions of articles and books, and are reluctant to pay extra for them. At the same time, because of lack of browser support, producing online versions of articles with a lot of math in them is very expensive, frequently involving a second, independent workflow for web publishing in parallel to the main print workflow.

In order to address the high cost of web publication, many publishers are moving toward XML-based workflows, where the same document can be composed as PDF for print and as HTML for web publication. In this context, the appeal of MathML for STM publishers is obvious. By using MathML to encode equations, XML documents can be self-contained. There is no need to generate and store hundreds of images of equations along with a document. Further, since MathML is an XML application, documents can be uniformly processed using industry standard tools such as XSL stylesheets. In the past, it was often necessary to somehow extract the math for separate processing, and then merge it back into the text later in the composition process. For a more detailed analysis, see the Design Science white paper, *MathML Workflows in STM Publishing*.[18]

The short term cost-cutting benefit of unifying workflows is noteworthy. But MathML's potential for adding value to web publication may be even more significant. In addition to fitting nicely into XML workflows, MathML is an information-rich way to encode mathematics. It takes pains to insure the hierarchical structure of the markup coincides with the mathematical structure of the expression. As an example, in the expression (x+2)^2, the MathML markup structure makes it clear that the exponent applies to the entire expression, not just the final parenthesis. MathML also provides a means of directly specifying the mathematical content of an equation in markup in addition to the presentation markup that describes how an equation should be typeset.

Because there is so much information in a MathML expression, it can be used in ways that are impossible for the equivalent print expression. For example, MathML equations can transfer between applications using cut and paste. A researcher might cut a MathML equation from a web browser, and paste it into a computer algebra system such as Mathematica[19] or Maple.[20] Or a student could paste an equation into an interactive graphing applet like the WebEQ Graph Control. Accessibility is another area where information-rich MathML might play a significant role. MathML was designed with a view to voice rendering for the vision impaired. Accessibility legislation requires many commercial and governmental organizations to publish material in accessible format when possible. There have been a few prototype projects, and there is considerable synergy with other technologies such as VoiceXML.[21] For a more comprehensive look at the possibilities MathML offers, see the Design Science white paper, *MathML Adds Value to STM Publishing*.[22]

Before you can do any of these slick new things with an equation, you have to find it. Fortunately, because the presentation and meaning of an equation are tied together by the markup structure in MathML, it has great potential for improved searching and indexing of technical material. As increasing numbers of documents containing MathML appear on the web, metadata for math will become increasingly important as well. The timing is good for increase math metadata activity, particularly since there are signs that standards and technologies for handling metadata in general are beginning to stabilize. For example, Adobe has begun a major initiative to deploy a common way of storing and accessing metadata. Also, several metadata standards have been successfully employed for some time in particular vertical markets such as NewsML[23] for newspapers and PRISM[24] for magazine articles.

While MathML offers significant potential both for cost reduction and adding value, one might justifiably counter that it is unwise to count your chickens before they hatch. While a number of STM publishers are working on MathML-based projects, there are not yet many large, high-volume, integrated XML workflows incorporating MathML. To a large extent, this is a matter of inadequate support in the high-end tools.

Significantly, the tool situation has begun to change in response to customer demand. Since demand ultimately determines the success or failure of a technology, new demand for tools is worth a closer look. In the following section, we will focus on forthcoming MathML support in several important XML and HTML tools.

Focus: MathML Support in Web Publishing Tools

MathFlow and Arbortext

The most ambitious integration of MathML support with high-end publishing tools announced [25] to date is a partnership between Design Science and Arbortext. MathFlow™ for Arbortext combines aspects of Design Science's MathType,[26] WebEQ [27] and MathPlayer products to provide comprehensive MathML functionality for Arbortext's Epic XML editor[28] and backend E3 e-content engine. MathFlow for Arbortext was announced at XML 2002 in December, and is currently in beta testing.

MathFlow utilizes a combined DocBook and MathML markup language called AxDocBook + MathML, which extends Arbortext's standard DocBook support. MathFlow consists of three parts. MathFlow Exchange works with Epic's Interchange module to import documents from Microsoft Word containing MathType equations. Equations are converted to MathML while the surrounding document is converted into AxDocBook.

Once an AxDocBook + MathML document has been opened in the Epic Editor, both the math and the document can be edited naturally. Equations appear in typeset form. Clicking on an equation opens it in the MathFlow Editor. Closing the Editor reinserts the typeset equation into the document. In general, editing in Epic is reminiscent of word processors, and the feel of the Epic/MathFlow integration will be familiar to Word/MathType users.

One a document is finished, MathFlow and Epic work together to generate both PDF and web output. To compose a document as PDF, users can choose from either XSL or FOSI stylesheets, which are used to transform AxDocBook + MathML into a low-level composition language used for formatting documents. The math equations are rendered into PostScript by the MathFlow Composer. The typeset equations and the remaining formatting code are then combined and converted to PDF by the Epic Composer.

For web output, there are two options. Users can use an XSL stylesheet to convert AxDocBook + MathML into XHTML + MathML for use in new browsers. Alternatively, Epic/MathFlow can generate Design Science image-based MathPage format, which uses CSS, JavaScript and images at several resolutions to create good looking web pages that print at 300dpi. MathPage documents extend accessibility back to the older 4.x browsers.

Dreamweaver and WebEQ Author

While MathFlow and Epic are high-end tools aimed primarily at corporate users, Macromedia Dreamweaver[29] and WebEQ Author are for a more mainstream audience. Dreamweaver is a widely-used HTML editor and site development tool. WebEQ Author adds MathML support to Dreamweaver in a way analogous to that in which MathFlow works with Epic.

Installing WebEQ Author adds an equation editor button to the Dreamweaver toolbar. Clicking the button opens the WebEQ Editor where an author creates an equation. Closing the editor inserts a preview of the equation in the Dreamweaver editing window. Double clicking the preview reopens the equation in the WebEQ Editor.

Equations are encoded as MathML code in the HTML configured to display properly in Internet Explorer with MathPlayer. However, because Dreamweaver doesn't support XHTML, only HTML, web pages created with Dreamweaver can't take advantage of the MathML support in Netscape/Mozilla without further editing. Since cross-browser compatibility is often important, WebEQ Author also lets authors generate web pages where equations use the image-based MathPage format described above.

Although XHTML has an advantage for cross-platform interoperability, HTML has advantages of its own. Most notably, HTML pages have much better support for interactivity in browsers. To take advantage of that, WebEQ Author includes a Solutions Library of templates for interactive mathematical web pages such as online quizzes, interactive graphing and plotting, and online tutorials. The templates utilize both dynamic web program techniques, and MathML-aware applets to provide graphing, evaluation, and equation editing capabilities within a web page.

WebEQ Author is slated for release in 2003. The JavaScript APIs and MathML-aware applets that go into the templates will also be included in version 3.5 of the WebEQ Developers Suite, which will also be released in early 2003.

Filling in Workflow Gaps

The MathFlow/Epic and WebEQ Author/Dreamweaver combinations are significant because taken together with MathType/Word, they provide end-to-end MathML workflow solutions where none have previously existed. However, they will not remain alone for long. Other vendors are also moving to fill in remaining gaps in XML + MathML workflow tools.

Plans have been announced to develop a version of MathFlow for Corel's XMetal[30] editor. XMetal has an import from MS Word feature and supports word-processor-like editing of XML documents. It also has basic printing functionality, though typically in workflows XMetal is used in conjunction with other composition engines such as XyVision XPP[31] which has also recently added MathML support.

While workflow tools such as MathFlow and Epic compose to PDF, many book and magazine publishers use QuarkXPress for composition. There are several mature math plug-ins for Quark of which Powermath is perhaps the most well-known. None of them currently have MathML support. However, a conversion tool, MathMonarch[32] from Westwords Publishing, can help bridge this gap.

MathMonarch 5.0, currently in beta testing, can do two-way translation between MathType Equation Format, MathML, LaTeX and WWDoc, the math markup language used by Powermath. MathMonarch launches from the toolbar in MS Word, and displays a control panel where the user specifies the input and output formats. Equations are converted to the desired format in place in the Word document. Other tools must be used to process the non-math portions of the Word document into other formats such as Quark's format or XML.

At present, there are a number of third-party software packages that address the problem of converting MS Word documents into XML. At least one, eXtyles from Inera, Inc,[33] uses Design Science's MathType technology to convert Word equations to MathML. However, the upcoming release of Microsoft Office 11 will have a large impact in this area, since support for XML is a major new feature in Office 11. Consequently, it seems likely that support for MathML in Word to XML conversion will remain somewhat ad hoc in nature until the dust from the Office 11 release settles.

News Round-up

This section spotlights important developments that have been announced since the most recent edition of the Status Report[34] was published in September 2002. The list may not be complete, and the authors apologize in advance for any omissions.

- The MathML Handbook is published. Charles River Media has published a book by Pavi Sandhu on MathML. The book provides a primer of MathML concepts, discusses techniques for working with MathML, and provides reference material.

- MathFlow for Arbortext Announced.[25] Design Science announced its MathFlow for Arbortext product at XML 2002.

- WebEQ Developers Suite 3.5 beta released. Beta testing for WebEQ Developers Suite version 3.5 began in December 2002. New features include a MathML-aware graphing applet, MathML evaluation capabilities, and templates and JavaScript libraries for creating dynamic math web pages.

- MathPlayer is Seybold Hot Pick.[15] MathPlayer, Design Science's high-performance MathML rendering behavior for Internet Explorer was chosen as a Hot Pick at the Seybold 2002 conference in San Francisco.

- XSLT MathML Library Version 2.0.[17] An open source project to develop XSL stylesheets to convert from MathML to $\LaTeX$ was launched at Source Forge.

- New version of MathML Test Suite released. The official MathML Test Suite has be expanded and updated.

- OMDoc mode for Emacs released. A beta version of an extension to the popular Emacs editor was released as part of the CCAPS project at Carnegie Mellon University. When completed, the OMDoc "mode" for Emacs will contain support for editing MathML.

References

1. Netscape 7.0, http://channels.netscape.com/ns/browsers
2. Mozilla 1.1, http://www.mozilla.org/
3. MathPlayer, http://www.dessci.com/en/products/mathplayer/
4. Microsoft Internet Explorer, http://www.microsoft.com/windows/ie/default.asp
5. Adobe's Portable Document Format (PDF), http://www.adobe.com/acrofamily/main.html
6. MathML, http://www.w3.org/Math
7. World Wide Web Consortium (W3C), http://www.w3.org/
8. Hypertext Markup Language (HTML), http://www.w3.org/MarkUp/Overview.html
9. Extensible Hypertext Markup Language (XHTML), http://www.w3.org/MarkUp/Overview.html
10. Resource Definition Framework (RDF), http://www.w3.org/RDF
11. Scalable Vector Graphics (SVG), http://www.w3.org/Graphics/SVG
12. Extensible Stylesheet Language (XSL), http://www.w3.org/Style/XSL
13. Cascading Style Sheets (CSS), http://www.w3.org/Style/CSS
14. Document Object Model (DOM), http://www.w3.org/DOM
15. Seybold 2002 Hot Pick, http://www.seyboldreports.com/Specials/HotPicksSSF2002/index.html
16. Formula Finder, http://f2.moris.org/
17. XSLT MathML Library Version 2.0, http://xsltml.sourceforge.net/
18. MathML Workflows in STM Publishing, http://www.dessci.com/en/reference/white_papers/mathml_workflows.htm
19. Wolfram Research (Mathematica), http://www.wolfram.com/
20. Waterloo Maple (Maple), http://www.maplesoft.com/
21. VoiceXML, http://www.voicexml.org/
22. MathML Adds Value to STM Publishing, http://www.dessci.com/en/reference/white_papers/mathml_adds_value.htm
23. NewsML, http://www.newsml.org/
24. Publishing Requirements for Industry Standard Metadata (PRISM), http://www.prismstandard.org/
25. Press Release, http://www.dessci.com/en/company/press/releases/dec02.htm
26. MathType, http://www.dessci.com/en/products/mathtype/
27. WebEQ, http://www.dessci.com/en/products/webeq/
28. Epic, http://www.arbortext.com/html/epic_editor_overview.html
29. Dreamweaver, http://www.macromedia.com/software/dreamweaver/
30. XMetal, http://www.softquad.com/products/xmetal
31. XyVision XPP, http://www.xyvision.com/xpp.asp
32. MathMonarch 5, http://www.monarchsuite.com/newmath.html
33. Inera's eXtyles, http://www.inera.com
34. Math on the Web Status Report (all editions), http://www.dessci.com/en/reference/webmath/status/

Design Science
How Science Communicates™