Math on the Web: A Status Report

September, 2003
Focus: Interactive Math

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: Interactive Math

by Robert Miner and Paul Topping

A bias toward future trends is hard to avoid in a periodic status report. Reading and writing about how things will be better tomorrow is more appealing than the difficulties of today. Past issues of this report have doubtless been guilty of "future bias" by focusing mostly on emerging technologies and new products. As a counterweight, in this report, we will look at some empirical evidence about the math actually on the web today.

The methodology is simple. We take a look at the first 50 hits in the search engine Google[1] for four topics: finding a common denominator, factoring polynomials, Taylor series, and beam bending. These 200 pages form a cumulative record of evolving web technology and practice. Categorizing the resulting collection according to technology and mode of presentation reveals some interesting results. For example, interactivity is present in around 20% of the pages. This is quite a high figure considering that it is mostly older pages that rank highly in search engines.

Another theme we will examine is the progress of MathML[2], the standard XML encoding for mathematics developed through the World Wide Web Consortium[3] in 1998. XML is a generic set of rules for specifying markup languages. Markup languages defined in terms of XML cover a huge range of applications from web pages to database records. By using a common XML format for diverse kinds of data, managing data using generic tools becomes easier and cheaper. That is the theory anyway, and even if a footnote or two might be required to qualify that statement in practice, there is no real question that the movement toward XML today is broad and deep.

Consequently, trend watchers have long predicted that MathML will become increasingly important as XML does. This is already happening in a number of concrete ways which we identify below. However, the trend toward MathML in the future must be tempered by the fact that it does not yet appear in the 200 Google pages in our study. We will explore some of the reasons why MathML technology doesn't appear to have percolated down to the authors in the trenches yet, and what software vendors can do to help end-users take advantage of the benefits of MathML as part of the larger XML world.

MathML Consolidates Gains as XML Data Format

Archival Data vs. Content Delivery

MathML, as an encoding of math notation, is seldom useful in isolation. MathML is almost always used in combination with other markup languages, such as HTML. MathML only describes the math within the document, while some other markup language is necessary to describe the text and high-level document structure, such as section headings and so on.

When MathML was originally developed at the World Wide Web Consortium (W3C), the expectation was that it would mostly be used together with HTML, the markup language for web pages. In fact, MathML was developed as only one of a collection of standards meant to create a rich, interactive medium for content delivery which in past issues of this report we have called the MathML+HTML Platform. The MathML+HTML platform includes style languages (CSS[4] and XSL[5]), a programming model for interactivity (DOM[6] and JavaScript), and a variety of standards and conventions detailing how to use multiple XML languages together (namespaces, schemas, XHTML[7] modules, etc.).

The effectiveness of the web as a means of delivering content depends not only on standards, but also on browsers. Though the standards of the MathML+HTML platform are quite stable and mature, implementation in browsers has lagged behind. There has been significant progress, especially in the last year. However, interoperability problems remain an issue, especially in contexts such as education where many older browsers are still in use. Inertia is also a major culprit. As the

wave front of web innovation has slowed and broadened, the introduction of new browser technology has become proportionally more difficult. While new sites can take advantage of new technologies relatively easily, retrofitting a pre-existing site is often not a priority unless there are other reasons justifying a re-design of the site as well. Consequently, adoption of MathML as a means of content delivery is proceeding slowly.

In sharp contrast, MathML as an archival data format is enjoying great success. Unlike the web, which arose from nothing in the mid-90's, publishers of technical documents have long been dealing with older formats and workflows for math and other technical data. In large part, XML was developed specifically to address problems evident in these older formats. However, because content providers had workable solutions in place and a level of quality they needed to maintain, their migration to XML workflows has been slower and more deliberate than the unconstrained rush to the web of the mid-90's.

Consequently, MathML as a data format is on the right side of the XML technology adoption curve. MathML, as a relatively mature standard with no real competitors, is the obvious choice for math in XML. Because it is the obvious choice, that is where software development and workflow integration resources are going. This in turn makes MathML even more attractive, and so on, thereby effectively "locking in" MathML as the solution of choice for math in XML. Significant new MathML adoption activity in the last six months on the part of organizations such as Elsevier, Wiley, the American Institute of Physics and Marcel Dekker suggests MathML lock-in is underway, just as trend watchers have long predicted. Another indicator of MathML's success in this arena is its inclusion in new document formats[8,9] for the archival holdings of the National Library of Medicine and PubMed Central.[10]

Cost savings are one obvious motivation behind the move by content producers to XML and, by extension, MathML. The ability to store and process math markup inline with the rest of a document simplifies maintenance and increases reuse, thereby reducing costs. The alternatives of storing math as images or in formats such as TEX which require external processing are widely regarded as error prone, inflexible, and difficult to maintain in the context of a publishing workflow. Beyond cost savings, however, MathML is an information-rich encoding which creates many opportunities to add value. MathML enforces a structured approach to encoding mathematics notation, and can contain semantic hints about the meaning of a formula as well. As a result, MathML holds promise for facilitating a variety of "smart" services for mathematics. Examples include making math accessible to the visually impaired, making math available for calculation, better searching of mathematics, and of course, interactivity. There are a variety of research and development projects underway in all of these areas, and as more content producers shift to using MathML, it will be interesting to see what value-added services emerge as content producers seek to differentiate their products.

The MathML Software Landscape

The adoption of MathML by large content producers is fueling the development of high-end, MathML-aware software. However, progress in software aimed at supporting individual authors is more mixed. In particular, MathML support in browsers has suffered some recent setbacks, as well as some substantial advances.

On the positive side, browser support for MathML in Internet Explorer[11] under Microsoft Windows has become much more widely available. Design Science's MathPlayer[12] extension, which adds MathML support to IE, has over 100,000 down-loads. In another significant development, Microsoft has licensed MathPlayer for use with MSN. MathPlayer will be distributed with the MSN client software, bringing MathML browser support to many millions of desktops over the coming months. MSN subscribers will use MathPlayer to view content in their Math Homework Help feature, part of an MSN offering for students and parents.

MathML support also continues to improve in recent releases of the Mozilla/Netscape[13, 14] browser. However, in May, AOL Time Warner and Microsoft announced they had reached an agreement which will make Internet Explorer the default browser for AOL users. While it is not yet clear what impact this will have on future development of Netscape or Mozilla, its open source twin, it seems unlikely to be good. At the same time, Apple announced a switch to the Safari[15] browser as the default browser for OS X. Safari at present offers no math support, though MathML ranks high in informal polls of requested features. It is to be hoped that the Safari team listens to its users. With the Netscape/Mozilla browser under siege, and a Microsoft announcement that it will no longer be developing Internet Explorer for the Mac, Apple users' prospects for robust math support are once again lagging behind other platforms.

Turning from browsers to authoring tools, there are several significant new developments. Design Science introduced MathFlow;[16] a suite of MathML tools for use with Arbortext's Epic Editor.[17] Epic is a high-end XML editor, and MathFlow aims at supporting content producers using MathML within XML workflows. One can expect to see further activity in the area of MathML+XML editing tool integration in coming months. A more end-user oriented tool is SciWriter,[18] recently introduced by Soft4Science. SciWriter is a dedicated XHTML+MathML editor. New releases of MathType[19] and WebEQ[20] from Design Science, and Scientific Word[21] from MacKichan Software also add new features for MathML authoring.

Two areas that have not seen much progress are TEX translation software, and support for MathML in page layout programs such as Quark XPress[22] and Adobe InDesign.[23] As we will see in our analysis of math currently on the web in the next section, TEX translation software remains an important means of putting math on the web for a significant class of academic authors. The continued lack of good TEX-to-MathML conversion software is therefore an obstacle to MathML adoption by academic authors which should be removed as soon as possible. Similarly, the adoption of MathML by publishers is increasing pressure for MathML support in page layout software. While several groups are exploring solutions, this remains a significant gap in the MathML software landscape.

Focus on Interactive Math

Debate on Technology in Education

Interactive, multimedia documents are frequently touted as one of the great appeals of the web, and that is unquestionably true at some level. Of course there are many cases where interactivity is not appropriate. Few people want to study their bank statements amidst a multimedia swirl of sound, animation and imagery. However, the debate over when interactivity and multimedia are effective for instructional purposes is not always so clear cut. This is particularly true in the area of online learning.

The debate runs that while engaging students' interest is good, replacing substance with glitz is bad. Finding the proper balance is difficult, though one suspects that using technology to enhance learning is probably no more difficult to do well than it is to use any other educational methodology well. Further, only some topics are well-suited to the use of web technology, and even in those cases, using technology to good effect is apt to require a good deal of creativity, energy and persistence on both the part of the instructor and the students.

Nonetheless, using the web in math and science education has many proponents. In part, this is also a form of future bias. A certain fraction of teachers and students always will be caught up in using the latest technology simply because it is new and exciting. In a kind of placebo effect, teaching and learning benefits simply because learning benefits whenever teachers and students are energized for any reason. Because of the need to constantly re-energize teachers and students, there is a long and venerable tradition of new initiatives in teaching methodology, and that is a good thing.

The appeal of the new, however, is too short lived and too tied to individual personalities to have lasting impact. The long term impact of the web on

math and science education will depend largely on the extent to which it can move out of the province of enthusiasts and into the mainstream. An obvious point of comparison is use of graphing calculators, which has become entrenched over the last decade. The argument for graphing calculators originally ran much the same as that for the web today. They were a means of engaging students by presenting material in graphical and computational ways. While graphing calculators also had an early cadre of enthusiasts, a key factor in their success was that they could be used by average teachers in mainstream settings. The learning curve was not too steep; textbooks could be written that incorporated calculator investigations within the capabilities of average students. And as calculators only cost a bit more than a textbook, and any classroom with an overhead projector could become a calculator lab, the financial burden of incorporating calculators into the curriculum was bearable.

The possibilities that the web offers obviously far exceed the capabilities of a graphing calculator. However, the challenges of making web-based instruction work for the mainstream are also proportionally greater. Problems of hardware and software compatibility, and network and computer access remain significant for web-based education on a large scale. Because of the huge range of possibilities, integration of web-based materials into mainstream curricula largely remains ad hoc and proprietary. Obtaining the necessary software tools and the technical skills to use them also remains difficult for both students and instructors.

Though these challenges are substantial, the appeal of the web is sufficiently great that they have not deterred people from taking them on. One approach that has gained momentum over the last several years is the widespread use of learning management systems (LMS) such as WebCT,[24] Blackboard,[25] and eCollege.[26] Such systems provide a generic framework for web-based courses, and even some functionality aimed specifically at math and science, such as math-enabled message boards. However, LMS's are generally agnostic when it comes to the actual course materials instructors manage with them. Typically, an instructor uses the LMS to create a course web site, and then uploads instructional materials to that site. If the 200 documents we survey in the next section are any indication, there is little convergence in approaches used for math content on the web.

The Google 200

To find out how math on the web, particularly interactive math, is really being used, we examine the first 50 hits in the search engine Google for each of the following topics:

- finding a common denominator
- factoring polynomials
- Taylor series
- beam bending

These topics were selected with several objectives in mind. The level of the topics ranges from basic to advanced. They also span several disciplines, with beam bending being a standard topic in engineering, and Taylor series figuring prominently in a number of disciplines. Finally, an attempt was made to choose topics where interactivity might reasonably be employed to good effect.

The resulting collection of 200 documents is scarcely a scientific sample of math on the web, but it is informative nonetheless. Many of the pages appear to be quite old, but no attempt was made to quantify this impression. In most cases, duplicate hits to the same site have been counted as a single item. Pages that are too broken or unfinished were eliminated from consideration.

The general profile of the collection is as follows:

completely off topic	38 - coincidence, advertisements, software manuals, etc.
newsgroup threads	12 - mostly homework help
professional or research related	19 - lesson plans, research articles, etc.
educational	81 - 23 commercial, 58 academic
other	50 - broken, unfinished, duplicate, etc.

The incidence of off-topic pages was highest in the "beam bending" and "finding a common denominator" searches. For beam bending, this is most due to coincidental hits on beam bending equipment and engineering company web sites. In the case of common denominators, most of the irrelevant hits were coincidental, where the phrase was being used as a figure of speech. Similarly, the density of research-related pages was predictably highest in the Taylor series and beam bending searching, while lesson plans and other pages aimed at teaching professionals were denser in the other searches.

Diverse Formats

Of the Google 200, 81 pages are instructional materials. The following table summarizes the wide range of technologies employed for the math in these pages. HTML pages in the table are described as "plain," "fancy" and "latex2html." The categories are purely subjective. Plain pages look as though they were probably created by hand-editing HTML code, while fancy pages exhibit more graphic design and were likely created using dedicated web authoring tools. Pages generated by the converter program latex2html[27] are fairly plain, but they have a unique look and feel, and they are the only group of documents produced with a single tool large enough to be worth noting. The rows of the table shaded blue are formats for static math. Those shaded gray are approaches for interactive math.

Technology Used	Common Denominator	Factoring Polynomials	Taylor Series	Beam Bending	Total
HTML+Images	3 plain, 5 fancy	3 plain, 2 fancy, 1 latex2html	4 plain, 4 fancy, 7 latex2html	1 plain, 1 fancy, 1 latex2html	34 (42%)
PDF	4	#81 was scanned handwriting!	6	7	18 (22%)
HTML only	4 fancy	3 plain, 1 fancy	1 plain, 1 fancy	2 plain	10 (12%)
PowerPoint	3	#101 was HTML output from PowerPoint	1	1	5 (6%)
Word document	3			2	5 (6%)
Applets		1	2	2	5 (6%)
webMathematica		2	1	1	4 (5%)
Maple worksheets		2	1		3 (4%)
JavaScript	2				2 (2%)
CGI		2		#62 is a former project of the authors	2 (2%)
Flash		2			2 (2%)
GIF Animation	1				1 (1%)
Proprietary	1 plug in				1 (1%)

Commercial Sites

Nearly a quarter of the 81 instructional pages were on apparently commercial or non-academic, non-profit sites. Such sites were more prevalent at lower levels, often aimed at parents looking for math help for their children, or parents home schooling children. The 12 newsgroup threads in the collection should probably be counted along with the commercial content pages, since the majority of the newsgroup threads were moderated homework-help message boards on commercial or non-profit sites.

Some notable sites from the collection are:

- ExploreMath[28] is an innovative new site offering a large archive of Flash "gizmos", lesson plans, and hosted course pages.
- eFunda[29] is a engineering resource site making extensive use of webMathematica.[30]
- The Math Forum-Ask Dr. Math[31] is in a class by itself in the category of moderated homework help sites. Having practically invented the genre, the depth of their moderated forums is prodigious.

Interactive Math

While about 20% of the 81 instructional pages had some form of interactive math on them, proprietary commercial technology did not figure prominently. With the exception of some Maple worksheets, some webMathematica sites, and a custom plug-in, all interactivity was simple JavaScript, or homegrown, custom applets or CGI scripts. At the same time, there were also a number of abandoned pages with custom applets or scripts now broken and in disrepair. Such pages, often the work of students, frequently have the air of having been a great learning experience for the author, which were abandoned as soon as the project was finished.

There were no pages using LiveMath,[32] Maplet.NET,[33] MathWright,[34] or Techexplorer[35] to mention a few of the better known commercial vendors of interactive math solutions. There was a single page using WebEQ, but it was broken. There are, of course, many pages on the web using all of these technologies, but they didn't make the Google 200. Freely available toolkits for interactive math faired no better. Two excellent examples of this genre are Configurable Java applets[36] at Hobart and William Smith Colleges, and Manipula Math.[37] This reinforces the impression that historically interactive math pages have often been more for the author than the reader.

The Absence of MathML

MathML makes no appearance at all in the Google 200. However, there are several factors one needs to take into account when interpreting this statistic. First, pages using MathML are likely too new to rank highly at Google. Rankings at Google depend on other sites referencing a page, as well as the content of the page itself. Secondly, much of the MathML-based content of which we are aware is in professionally developed content to which access is controlled. For example, MSN Math Homework Help uses MathML, but it is only available to premium subscribers and, therefore, does not usually register in search engines.

Another likely place one might find MathML is in course web sites in university learning management systems. However, pages within LMS's are access controlled and, therefore, they do not typically register with search engines either. As many of the web-savvy authors that a few years ago were developing the pages that currently appear in the Google 200 are likely the same people who now use LMS software, this is a potentially significant factor.

The Need for Authoring Tools

Over half the pages in the Google 200 use HTML with bitmapped equation images. More than half of those have been prepared with the assistance of mainstream HTML editing software. However, in very few pages did the mathematics match the production values of the rest of the document. In most cases, the math was badly aligned and often poorly typeset. Clearly, there is a need for better math support in mainstream HTML editors, such as those that made many of the pages in the collection.

In the area of interactivity, most of the pages in the

collection were developed either by professionals or by individual enthusiasts. The techniques employed are ad hoc and idiosyncratic. Consequently, the material in the collection doesn't point a clear direction for authoring tools. However, one thing is clear. Unless authoring non-programmers are comfortable with it, significant use of interactivity will largely remain restricted to professionally created sites.

The recent 3.5 release of the WebEQ Developers Suite takes a tentative step in the direction of reducing the need for programming in authoring interactive math. The new release includes a Solutions Library consisting of high-level JavaScript libraries and HTML templates. The JavaScript libraries provide authors with three categories of objects that can be used in a page:

- *Equations*, which give authors an easy way to insert equations into pages using JavaScript. Equation objects handle the low-level details of displaying MathML equations across browser platforms. Equation objects also provide an easy way to manipulate equations in response to user actions from JavaScript.

- *Controls*, which insert applets into the page for displaying, editing, graphing, evaluating and comparing MathML equations. The Control objects provide a simple way to incorporate the "math controls" into a page and manipulate them via JavaScript.

- *Logic Modules*, which implement a number of standard interactive math tasks in a configurable way: quizzes, step-by-step exposition, and animations.

In addition to the JavaScript libraries, the Solutions Library contains several dozen HTML templates and sample pages. The templates illustrate a variety of interactive math activities, and can be easily adapted to new subject matter with minimal JavaScript programming. While the Developers Suite is still a collection of tools for programmers, Design Science has announced plans for an authoring tool for non-programmers, building on the Solutions Library and integrating with the Dreamweaver HTML editor.[38]

News Round-up

This section spotlights important developments that have been announced since the most recent edition of the Status Report was published in January, 2003.[39] The list may not be complete, and the authors apologize in advance for any omissions.

- **Microsoft and Design Science announce MathPlayer for MSN.** Microsoft and Design Science announced a licensing arrangement that will provide MathPlayer™ to users of MSN's Math Homework Help content. Math Homework Help gives students step-by-step guidance with problems in commonly used textbooks, and MathPlayer is software that enables high-quality display, print and interaction of mathematics within Microsoft's Internet Explorer for Windows web browser.

- **Microsoft and AOL Time Warner strike a deal.** CNN/Money reports that Microsoft agreed to pay AOL Time Warner $750M to settle an anti-trust lawsuit filed by AOL on behalf of its Netscape subsidiary. The companies also agreed to a seven-year licensing deal that allows AOL to use Microsoft's Internet Explorer web browsing technology.

- **National Library of Medicine announces archival formats using MathML.** The National Library of Medicine (NLM) announces the creation and free availability of a standard model for archiving and exchanging electronically journal articles.

- **WebEQ 3.5 Developers Suite released.** Design Science announced the release of WebEQ™ 3.5, a developer's toolkit for building web pages which include interactive math. The new version includes new web-based controls for graphing and evaluating equations, and a Solutions Library to reduce the development time in creating interactive pages.

- **SciWriter 1.0 released.** Soft4science, a Germany-based software developer, released SciWriter 1.0, an XML based "WYSIWYG style" word processor that integrates writing mathematics and text in the same environment. SciWriter's document format is a subset of XHTML 1.1 and MathML 2.0 Presentation Markup.

- **Integre Technical Publishing acquires Techexplorer.** Integre Technical Publishing has acquired the Techexplorer Hypermedia Browser plug-in and related technologies from IBM Research. Techexplorer is a cross-platform, cross-browser plug-in that supports rendering and dynamic scripting of TEX and MathML markup.

- **MathPlayer available for 3rd party redistribution.** Design Science announced that MathPlayer™ is now available to anyone who wants to distribute MathPlayer on networks, intranets and on physical media, such as CD-ROM.

- **Scientific Word 5 released.** MacKichan Software announced the release of Scientific Word 5.0, a word processor that enables a user to export as MathML+HTML.

- **MathType 5 for Macintosh in beta testing.** Design Science announced the beta release of MathType™ 5 for Macintosh, a native OS X application that also works on OS 9, that matches MathType for Windows feature-for-feature.

- **MathType 5.2 for Windows in beta testing.** Design Science announced the beta release of MathType™ 5.2 for Windows, an upgrade that is fully compatible with Microsoft Office 2003 (Office 11).

References

1. Google, http://www.google.com
2. MathML, http://www.w3.org/Math
3. World Wide Web Consortium (W3C), http://www.w3.org/
4. Cascading Style Sheets (CSS), http://www.w3.org/Style/CSS
5. Extensible Stylesheet Language (XSL), http://www.w3.org/Style/XSL
6. Document Object Model (DOM), http://www.w3.org/DOM
7. Extensible Hypertext Markup Language (XHTML), http://www.w3.org/MarkUp/Overview.html
8. National Library of Medicine Journal Archiving and Interchange DTD, http://dtd.nlm.nih.gov/tag-library/n-i000.html;
9. National Library of Medicine Journal Publishing DTD, http://dtd.nlm.nih.gov/publishing/tag-library/n-i000.html
10. PubMed Central, http://www.pubmedcentral.gov
11. Microsoft Internet Explorer, http://www.microsoft.com/windows/ie/default.asp
12. MathPlayer, http://www.dessci.com/en/products/mathplayer/
13. Netscape 7.0, http://channels.netscape.com/ns/browsers;
14. Mozilla 1.1, http://www.mozilla.org/
15. Safari 1.0, http://www.apple.com/safari/
16. MathFlow 1.0, http://www.dessci.com/en/products/mathflow
17. Arbortext Epic, http://www.arbortext.com/html/epic_editor_overview.html
18. SciWriter 1.0, http://www.soft4science.com/
19. MathType, http://www.dessci.com/en/products/mathtype/
20. WebEQ, http://www.dessci.com/en/products/webeq/
21. Scientific Word, http://www.mackichan.com
22. Quark XPress 6.0, http://www.quark.com
23. Adobe InDesign 2, http://www.adobe.com/products/indesign/main.html
24. WebCT, http://www.webct.com/
25. Blackboard, http://www.blackboard.com
26. eCollege, http://www.eCollege.com/
27. latex2html, http://www.latex2html.org/
28. ExploreMath, http://www.exploremath.com
29. eFunda, http://www.efunda.com/
30. Wolfram Research (Mathematica), http://www.wolfram.com/
31. Ask Dr. Math, http://mathforum.org/dr.math/
32. LiveMath, http://www.livemath.com
33. Maplesoft (Maple.NET), http://www.maplesoft.com
34. MathWright, http://www.mathwright.com
35. Techexplorer, http://www.integretechpub.com/webmath/techexplorer.html
36. Configurable Java applets, http://math.hws.edu/javamath/
37. Manipula Math, http://www.ies.co.jp/math/java/comp/index.html
38. Dreamweaver, http://www.macromedia.com/software/dreamweaver/
39. Math on the Web Status Report (all editions), http://www.dessci.com/en/reference/webmath/status/