Report on the MathML conference

University of Illinois, October 2000

Pam Bishop,
Maths, Stats & OR Network

P.Bishop@bham.ac.uk

During 1999 CTI Mathematics helped to organise two workshops on the theme "Putting Mathematics on the Web", and the report of these workshops [1] contain many references to the tools available for the creation and rendering of MathML. Clearly there have been a lot of changes since then, and I attended the international MathML conference in order to find out the current state of play. The conference was sponsored by Wolfram, IBM, Netscape, Maple and the World Wide Web Consortium (W3C), and was attended by almost 200 people with a wide range of interests. The organisers intend to publish presentation materials on the conference web site. [2]

1. MathML tutorial

The conference was preceded by a tutorial session which set MathML in context. MathML is just one of the Web technologies being developed by W3C into standard "vocabularies" of XML (extensible markup language) which will replace HTML for Web documents. Other examples include CSS (cascading style sheets) and SVG (scalable vector graphics).

MathML is in fact the first XML vocabulary to be agreed by W3C. It consists of elements like <math>, <mtext> etc. which can be qualified with attributes. Presentation markup has about 30 elements and 50 attributes, and enables correct display of a mathematical expression. Content markup has about 100 elements and 12 attributes, and encodes the mathematical meaning of an expression. These elements can be combined together, and with interface elements that embed graphics, sound or video. One extension is the use of MathML with voice software to read mathematics aloud for visually impaired people.

We learned a few presentation and content markup tags and attributes, but it became clear that writing MathML is not something to be done by hand - it is much more likely to be generated from other mathematical software. There are also programs which can validate MathML documents, whether generated by hand or machine.

XML vocabularies can be embedded in HTML by using an XML namespace which shows where the relevant standard can be found. For example, a browser would need to recognise the following in order to render MathML:

```
math xmlns = "http://www.w3.org/1998/Math/MathML"
```

The Amaya and Mozilla browsers are already set up to allow this. Other browsers would need helpers in the form of applets or plug-ins. Further details follow, but during the tutorial we sampled some of the possibilities, especially IBM techexplorer. This is a plug-in which renders TeX or MathML not only in standard browsers but also in programmes like PowerPoint.

Computer algebra packages are introducing MathML in various ways: REDUCE has a calculator, Maple recognises content markup, Mathematica can accept presentation and content markup. In version 4.1 of Mathematica it is possible to save the contents of a NoteBook as a techexplorer document which will display in a browser and allow for interactive mathematics, ie the browser will act as a front end for Mathematica.

There is an online version of this report at:

http://ltsn.mathstore.ac.uk/reports/mathml.htm

2. Recognising the need for collaboration

In a plenary presentation, Bob Sutor of IBM quoted MathML as an example of how collaboration on standards can help communications, and showed how this is being extended to the business world. The "Universal Description, Discovery, and Integration" project [3] led by Ariba, IBM, and Microsoft aims to create a huge directory in which businesses identify themselves, describe what they do, and outline how others can conduct electronic transactions with them. Officials at the three companies believe that by making such information widely available in a standard format, they can reduce the confusion that has seriously held back e-commerce between businesses.

3. MathML in Internet Explorer

Internet Explorer [4] has been developed using a component model known as behaviours, ie there is a core functionality based on standard HTML, with the possibility of adding combinations of XML components. The components are developed by specialists, and can either be installed as plug-ins on an individual client machine, or referenced dynamically from a trusted source. Two MathML plug-ins are available: IBM techexplorer and the coming MathPlayer from Design Science. For dynamic referencing a special declaration is needed in the head of an HTML document, using the <object> tag and MathML namespace. Behaviours may not be available on Mac platforms at present. Dynamic referencing requires that security is set to accept ActiveX controls, which may not be the case in a student lab.

4. MathML in Mozilla

Mozilla [5] allows full integration of MathML, which can be mixed with other XML vocabularies, including HTML which is XML compliant (XHTML). This means that learning and teaching documents containing mathematical expressions mixed in with ordinary text, graphics etc. are possible. It remains to be announced when these aspects of Mozilla will be incorporated into Netscape. MathML is not enabled in the new Netscape 6, although this is the first to be powered by the Gecko engine. Gecko Technologies are being adopted by IBM, Intel, Sun etc because they consistently support a number of W3C standards across platforms and devices.

5. MathML in Amaya

Amaya [6] is simultaneously a browser and an editor which supports MathML and other Web technologies:

- XHTML (XML compliant HTML)
- CSS (cascading style sheets)
- SVG (scalable vector graphics)
- RDF (resource description framework, a metadata language for annotations)
- HTTP (for publishing web pages remotely)
- XLINK

Mathematical documents normally contain a mixture of formatted text and equations, and Amaya deals with this by combining different XML vocabularies in a single structure, represented by a tree. As an editor it has a good user interface, and provides several views of an expression, all of which can be edited:

- formatted view (what you see in the browser)
- XML source code (presented as XHTML and MathML)
- structured view (known as a Document Object Model tree)

Changes in the structured view are reflected in the content window in real time, and mathematical expressions can appear within a diagram. Style attributes can be edited locally or globally, eg all <msup> elements can be constrained to be coloured blue. Hypertext links can be added to formulas in the same way as text or images.

6. MathML helpers

IBM techexplorer 3 works interactively within many Windows applications - Netscape, IE, Microsoft Office applications, and future MathSoft products. A Mac version is also available. As a browser plug-in it allows live interactive maths on an HTML or XML page, and interoperability with the Mathematica 4.1 kernel via the MathLink DLL which ships with this product. It can also be used in audioformatting, or speech augmented by non-speech sound cues - pauses, intonation (down for a subscript, up for a superscript), text in brackets spoken more softly - using the IBM ***ViaVoice*** add-in.

Mathcad 2001 with techexplorer gives a translation including content MathML. HTML files can be edited with a Mathcad application to produce interactive pages. Some examples can be found at the StudyWorks online site [7].

MathPage will be included in the next version of Design Science's MathType. It is a plug-in for Microsoft Word that modifies the "save as Web page" feature to work on all browsers (or can be optimised for IE). The output using Javascript and gifs will be competitive with PDF files for printing. Alternatively it will produce MathML compatible with Mozilla or MathPlayer.

MathPlayer is a plug-in specially tuned for IE5.5 which renders TeX and MathML. It installs automatically like any ActiveX control, and will shortly be available from Design Science.

Leibniz [8] is a freely available easy-to-use equation editor suitable for upper school students and undergraduates, and can be used as a front end for the Mathematica engine. From mathematical expressions it will generate HTML with either gif images or MathML content markup.

Handwriting interfaces - two were presented which take input from a pen/tablet and convert to MathML, LaTeX or Braille. The one developed in Japan can read and edit LaTeX or MathML files. It links to Mathematica, and can be used as a tutorial for writing mathematics correctly.

7. Two-way communication and collaboration

Web forms looking like exercise sheets with blanks that can take a student's maths input for server processing would be widely welcomed. If students could create MathML versions of their input, this would be easy, but not many have that facility at present.

The Math Forum takes a low tech approach for its Ask Dr Math pages, which are aimed at students, newcomers, scientists/engineers and maths educators. It uses flat HTML which is easy to read using the Lynx text-only, or on browser palmtops, cellphones, and other low end access to the Web, and guidance is available for entry of mathematical expressions [9]. The result is easily scanned by search engines.

Clifford Johnston's project at West Chester University [10] has developed Perl scripts which allow students to enter expressions in a format similar to that of the TI-85 graphical calculator, convert them to MathML and submit them to a server. This allows mathematical discussion between student and tutor or between students, and the ability to create online tests. It is currently restricted to elementary areas of maths (algebra, calculus, statistics) but the output could be MathML-generated audio.

Other technologies that can be used for communication of this type are

- PDF
- TeX entry (with a preview facility)
- Javascript
- WebEQ hypernews equation editor [11]
- techexplorer editor with the <embed> tag

There are plans for an equation editor applet based on the Maple user interface, including a syntax checker, which will produce content MathML.

8. Use of MathML in courseware - three case studies

Stratum Technical Services [12] have transformed an existing maths courseware book into a Web based course, which assumes access to computer algebra. This uses content MathML coordinated with presentation markup, and new definitions to extend functionality beyond the MathML 2.0 specification.

A collaborative WebCT project in Georgia is aimed at extending access to non traditional student audiences. The courses are to be available at home, via a modem, and no pictures or PDF files are allowed. MathML is used with the WebEQ applet. It was found that WebCT worked better with Internet Explorer than with Netscape or Amaya. The lecturers have asked for TI-86 formulation for student input of maths.

MERLOT (Multimedia Educational Resource for Learning and Online Teaching) is a searchable collection of high quality interactive online learning materials, assignments and reviews designed for staff and students in higher education [13]. Created in 1997 by the California State University Centre for Distributed Learning, information hosted in MERLOT is free to use for educational purposes.

9. Computing with MathML

The use of MathML in encoding proofs and large formal mathematical documents has been tested on the standard library of COQ, a proof assistant developed at INRIA in France as part of the Hypertextual Electronic Library of Mathematics project [14].

Numerical Computations: using content MathML to mark up algorithms can enable the generation of code for different computer languages. For example, an algorithm for solving ODEs has been given MathML markup using the Java-based software tool Visual Maths Editor, and this in turn can generate source code in C++,

Fortran or Matlab. In the latter case the resulting code can be exported to Matlab or used as a plug-in in other applications.

Mathematica is already being used as a computation engine for other programmes via MathLink. Input can be submitted in Mathematica code or in MathML and the results returned in a number of forms. Content markup can be pasted into a NoteBook and will appear as Mathematica expressions which can be edited or evaluated. Presentation markup is also recognised providing it is well-structured, as would be the case if created by a software tool. Otherwise a heuristic algorithm is used to guess the correct structure. Conversion of TeX is also possible providing that some context is given.

Mathematical computation over the Web is being addressed by the IAMC project at Kent State University [15]. It includes a MathML extension to represent curves and surfaces.

Project Links [16] has developed a set of Javascript functions which can be included in Web documents containing MathML or TeX, and which create extra HTML code according to whether a plug-in such as techexplorer is installed on a client machine. The function arguments are a TeX or MathML string plus width/height information. The function interrogates the browser. If a plug-in is installed it writes out an <embed> tag, otherwise an <applet> tag, and it specifies the width and height of the rendering to the browser to match the font size of the rest of the page.

10. MathML and TeX

There were several interesting contributions regarding MathML and TeX, which are important for the conversion of legacy mathematical documents.

Stilo have released the ***MathWriter*** package [17], which will convert TeX to MathML. The ***Omega*** System [18] extends TeX for printing all of the world's languages and for generating XML+MathML documents. It can be used to generate MathML from TeX or LaTeX. Other tools include

- TEX4HT, which produces hypertext from TeX
- XMLTEX, which takes an XML document and produces a DVI file for printing or screen display.
- PassiveTeX, which processes an XML document to TeX or, if using PDFTeX, to high quality PDF files

A project at the University of Illinois has converted the SGML versions of 50,000 full text scientific journal articles to XML using Javascript. The mathematics articles do not approach the quality of printed mathematics, so these are now being converted to MathML, and the project will compare its first approach to the use of plug-in or browsers with native MathML capability.

11. Summary - the state of the art - and references

Although the standard browsers (Netscape and IE) do not have native MathML capability, ongoing development in this area is rapid. PDF is still the best way to display static documents on the web, since the format is portable and documents consist of a single file. Dynamic mathematical documents can be published for standard browsers using HTML with embedded MathML and a plug-in, or for Amaya and Mozilla using XHTML and presentation MathML.

12. Summary - the state of the art - and references

[1] http://www.bham.ac.uk/ctimath/workshops/mathml.htm
[2] http://www.mathmlconference.com
[3] http://www.uddi.org
[4] http://msdn.microsoft.com/ie
[5] http://www.mozilla.org
[6] http://www.w3.org./Amaya
[7] http://www.studyworksonline.com
[8] http://www.leibnizsoftware.com
[9] http://mathforum.com/algpow/writing_answers.html
[10] http://math.wcupa.edu/~johnston
[11] http://www.mathtype.com/webmath/tech/dynamic.stm
[12] http://www.stratumtek.com
[13] http://www.merlot.org
[14] http://www.cs.unibo.it/~asperti/HELM/home.html
[15] http://www.mcs.kent.edu/home.html
[16] http://links.math.rpi.edu
[17] http://www.stilo.com/mathwritertechoverview.htm
[18] http://omega-system.sourceforge.net