OpenMath Meeting 26–28 September 2001
RISC, Hagenberg, Austria

Minutes by James Davenport & David Carlisle

January 23, 2002

1 Joint OpenMath/MKM day 26.9.2001

1.1 The NIST Digital Library of Special Functions

This talk was given by Daniel Lozier of NIST. He emphasised that this project was *not* about taking the 36-year old[1] information in the book and placing it on the web. The traditional Abramowitz and Stegun is but one of many sources going into the new information source. The publication date is expected to be in 2003. He outlined the subject matter as being:

- special functions of applied mathematics;
- validated technical data;
- to meet proven needs in physics and other sciences (i.e. not things only useful within mathematics);
- chapters on individual functions;
- methodology chapters (numerical methods, analytical methods, asymptotic methods and computer algebra);
- indexes and search engine.

The format will be a book *and* a search engine. The current web site is `http://dlmf.nist.gov`.

A typical chapter's contents would be

- mathematical notation and properties — the original Abramowitz and Stegun had a powerful normative influence on notation;
- metadata;
- graphs and visualisations;
- sample applications;

[1] Or even more: much of the information was collated by Abramowitz before his untimely death in 1958.

- computational methods;
- pointers to software (commercial and non-commercial);
- references.

The style will be terse "handbook" style, aimed at experienced scientists, rather than a traditional mathematical or pedagogical style. The book will be about 1000 pages, roughly the same size as the original, but containing twice as much mathematics, since the tables of the original will not be necessary.

There is a special LaTeX class for the DLMF, with macros, equation breaking, double (largely for the printed form[2]) and single (mostly on-screen) column style, and optional editing forms. This is accompanied by a special version of `latex2html`.

The metadata has two purposes.

1. Author's notes to help the reader: proof hints, additional references, notational reminders and acknowledgements.
2. Indexing metadata to construct both paper and Web indices, and to drive the search engine. One index will cross-reference the original Abramowitz and Stegun equation numbers to the new format.

The metadata is generally tied to the subsection, rather than the individual equation.

The speaker gave a guided tour of the Airy functions chapter[3] in various formats. One new section is that on error bounds, for both real and complex values, for the various asymptotic expansions presented. Recent work on exponentially-improved asymptotic expansions is also included. Under graphs and visualisations, there are line plots, contour plots, diagrams, and (colour) surfaces. It is unlikely that there will be many animations. However, the colour surfaces, which are largely for complex functions, are full VRML surfaces, and can therefore be rotated, intersected with various coordinate planes, etc. A very graphic presentation of the branch cut of a Hankel function was given.

The goal of the search engine is to be able to search in equations, which is not supported by current technology. There is a partial solution based on metadata and a very extensive thesaurus, together with a "pidgin math" parser. It is possible to search for `Gamma(1/3)`, `Ai^2+Bi^2`, `b_$` (a subscripted b, with the subscript unspecified) and so on. The engine is parameterised by a "search depth", which will, for example, find $Ai(z)^2 + Bi(z)^2$ as an answer to the second query.

Dr. Lozier also spoke to the presentation that Bruce Miller would have given. He pointed out that the distinction between parameters and arguments is somewhat artificial, but only somewhat. The markup is `\BesselJ{\nu}` or `\BesselJ{\nu}(z)`. There is much more known about hypergeometrics than in the days of Abramowitz and Stegun, where there are essentially three kinds of arguments:

[2]There is also a printed form with one column of text and one column of metadata, largely for use by authors and validators.

[3]This was the sample chapter produced in 1998.

```
\hyperpFq{p}{q}@{a_1,\ldots}{b_1,\ldots}{z}
```

might be used.

1.2 Certified and Portable Mathematical Documents

Martijn Oostdijk presented joint work with Olga Caprotti and Hermann Geuvers arising out of the "Algebra Interactive!" project. He pointed out that, within a mathematical document, there are many links between definitions, theorems, proofs etc., as well as within proofs, thus making these ideal for hyperlinks. In COQ's CIC, both objects and proofs inhabit the same universe, thus meeting de Bruijn's criterion and allowing a small type-checker. He noted that it was necessary to produce a "COQtree" in Java, and that the HEML project had presented a neat way of doing this via XML. Given a type-annotated COQtree, we can produce an OMdoc tree, which can be converted to an XML document, which an XSLT stylesheet can convert to the HTML which is actually displayed.

They had improved Coscoy's rendering of COQ into "natural language", by sometimes created sub-documents to reduce the amount of nesting that would otherwise be involved. He then gave a demonstration of this.

1.3 Mathematical Knowledge Representation

James Davenport presented this thesis, that knowledge representation was a vital pre-requisite for mathematical knowledge management.

Michael Kohlhase pointed out that the algebraic specification community, unfortunately not well represented at MKM, had answers to many of the questions. JHD agreed partially, but said that they did not have all the answers.

1.4 Meta stylesheets for the conversion of mathematical documents into multiple forms

Bill Naylor presented joint work with Stephen Watt on this subject. The problem being addressed is that of converting (extended) MathML to other forms, caused by the many↔many correspondence between presentation and semantics. One example is the many forms of the binomial coefficient, where $\binom{n}{m}$ can also be represented as ${}_mC^n$, or, worse, because totally ambiguously, C_m^n or C_n^m. They therefore claimed that one could store, either in the content dictionaries or in a parallel system analogous to the `.sts` files, notational information.

They therefore proposed a `<Notation>` element. This could have (several) `<version>` elements, one for each possible presentation of a semantic concept, which could, say, be `<math>` (presentation MathML), `<tex>` (for LaTeX) or `<image>` (an URI to a `.gif` file). There would also be matching semantic templates, as in

```
<semantic_template> <OMOBJ> <OMA>
  <OMS cd="combinat4" name="choose"/>
  <OMV name="n" id="argChoose1"/>
```

```
  <OMV name="m" id="argChoose2"/>
</OMA> </OMOBJ> </semantic_template>
```

where the `id` tags are used as cross-references in the `<version>` elements to the various arguments. There could be more than one `semantic_template` elements, since, say, integration would need different forms depending on whether it was $\int_a^b f(x)\,\mathrm{d}x$ or $\int_{x\in S} f(x)\,\mathrm{d}x$. It is also necessary to have template functions, e.g. to compute n in the notation $\frac{\partial^n}{\partial x^2 \partial y^3 \dots}$.

In this system, content dictionaries, or there associated files, could be processed by a "meta" style sheet that would produce various combined stylesheets for converting to various formats, such as straight presentation MathML, or presentation MathML with `csymbol` OpenMath references, or ….

They suggested that, say, `<xmml:choose style="2"/>` could be used to choose one of the `<version>` elements from the `<Notation>` elements mentioned above. If this wasn't given, then information could come from the defaults in the "meta" style sheet, or in the content dictionaries themselves.

He concluded by saying that the OpenMath Content Dictionary concept could provide a useful carrier for notational information. DPC pointed out that it might be necessary to have the full XSL functionality (as in `xpath`) to select sub-sub-arguments etc. His example was that of $\int \lambda x.\sin x$, which should be rendered as $\int \sin x\,\mathrm{d}x$. Sacerdoti Coen was worried about the performance implications of a large XSL stylesheet.

AMC asked whether this wasn't simply a phrase-book. MK thought that this approach was more flexible than a simple phrase-book approach.

1.5 Mathematical Software: the Next Generation

Mike Dewar presented a joint paper with David Carlisle on this topic. He began by summarising the history of NAG and mathematical software. The "temporary" Fortran library written to tide it over until the rise of Algol was still going thirty years and twenty versions later.

He noted that the scientific software market was very conservative. It was also very heterogeneous: research users; production users and education/training users. Take-up of web-based services would be very different in different segments.

He summarised the general wish for "plug and play" mathematics via the web. He then listed various problems.

- Semantics of objects and software.
- Handling errors and exceptions in a distributed environment.
- Tracing and debugging in a distributed environment.
- Embedding existing software in an environment it was not designed for.
- Scalability (partly of problem size, but also of the software life cycle).
- Reliability and reproducibility.

- Commercial and licensing issues (not today's issue, but a very real one).

The outline strategy was to develop a framework for embedding software components automatically, based on abstract specification: software "glue"; documentation; testing and verification material. There was also a need for resource discovery, based on standards such as RDF and WSDL.

He said that JHD had already spoken about OpenMath Content Dictionaries, but pointed out that OpenMath provided lightweight mechanisms for semantics.

He pointed out that the NAG Fortran library had 1200 user-callable routines, documented in 100Mb of PDF with 96,000 mathematical expressions. This was, unfortunately, unstructured, presentation-oriented and Fortran-specific, with limited verification and validation against the software. Nevertheless, he claimed that this was a significant repository of mathematical knowledge.

Many component frameworks are based on the Interface Description language (IDL). This is not powerful for mathematics. NAG therefore has an extension, which specifies for parameters: intent, purpose (argument, control, workspace, array dimensions), concrete versus abstract[4] type, defaults, constraints and specifications. This was semi-automatically derived from the existing documentation, and is represented in XML.

He then spoke about NAG's ideas on mathematical services. We need to be able to describe these, and support automatic service discovery. This needs

- problem analysis;
- qualitative decisions;
- explanation and justification mechanisms;
- mathematical and non-mathematical criteria;
- refinement of decisions in the light of experience and/or failure.

His example was the computation of $\int_0^1 \sin x \, \mathrm{d}x$, which could be done by `D01AJF`, or via the removable singularity routine `D01AHF`, or the sine-multiple routine `D01ANF`, or conceivably the sine-integral routine `S13ADF`. How is this to be elicited?

1.6 Likely Opportunities in Framework 6

Herr Hans-Georg Stork[5] spoke to this, using a mathematical analogy.

Definition The biggest pillar of FP6 will be "integrating European research". The other two are "Strengthening the ERA" and "Supporting the ERA" — ERA = "European research Area". Within this biggest pillar, there are various "priority thematic areas", and "anticipating Science and Technology needs". One priority thematic area is "Information Society Technologies".

[4] such as "upper triangular matrix".

[5] European Commission offices, Luxembourg. `Hans-Georg.Stork@cec.eu.int`

Two of the major instruments of FP6 will be "Networks of Excellence" (anticipated to have more funding than current networks, say 8–15 MEuro) and "Integrated Projects", which certainly have to include technical development, and might be 20–100 MEuro. There will also be room for the equivalent of the current IST projects.

Theorem The probability that MKM will be included in FP6 looks very high.

Proof This is divided into two cases.

(MK)M Management of Mathematical Knowledge. Past projects have been OpenMath and EULER. The OpenMath Thematic Network is alive.

M(KM) Mathematics of Knowledge Management. Past projects include IBROW, which used ontologies and ontology-based reasoning. Projects such as Ontologging, OntoKnowledge XML-Knowledge management etc. are alive. In the intersection, there are projects such as Calculemus and Types.

He noticed the amount of information available on Cordis under the keywords "mathematics" and "knowledge".

There is the current call under "Semantic Web Technologies". One line is "creating a useful formal framework", which is very relevant to MCD's talk.

The FP6 IST draft includes a box called "knowledge technologies", which he expects to be quite substantial.

Corollary He expects that both M(KM) and (MK)M will be funded. He recommended `http://www.cordis.lu/rtd2002` for the state of the discussion on FP6, and `http://www.cordis.lu/ist/ka3` for the current FP call.

Andrzej Trybulec asked about the status of Poland. It was hoped that Poland would be fully included in FP6. Michiel Hazewinkel said that there would be an MKM project-planning meeting in Amsterdam in November, and invited HGS to it.

1.7 Modelling for Understanding of Scientific Knowledge

Saverio Solerno gave this talk, which is situated in the framework of e-learning and intelligent tutoring systems. The aim was to set forward a representation method for a domain (e.g. calculus) in which points such as inductive or interdisciplinary ones can be considered as well as the hierarchical deductive points. At this stage, they are thinking of a fixed domain of knowledge.

It is important that the system understands the user's strengths and weaknesses, and can have a model of his misconceptions. The mathematical model used is that of a multigraph, whose nodes are the atomic concepts of the domain.

1.8 Contribution of Ontology Engineering to MKM

Francky Trichet gave this talk. The GINA project deals with sketches and sentences expressed in natural language (constraints etc.). The system includes geometric knowledge to interact with the user and to handle queries. This knowledge of projective geometry must be formally represented in the system. The knowledge acquisition process can be viewed as a corpus, being transformed by conceptualization to a conceptual model, which is transformed by ontologization to an ontology, which is then given a formal representation. He saw a semi-formal ontology as having two components: a formal part with a clear and consensual semantics, and an informal part which did not have a consensual semantics. In this case, the corpus was Hilbert's "Grundlagen der Geometrie", where the knowledge is already conceptualized: concepts (point, straight line, plane etc.), relationships (membership etc.) and axioms.

This is to be represented in the Conceptual Graphs model. This has two levels of representation: terminological (concept types and relation types) and assertional (representation of facts with conceptual graphs etc.). There is a hierarchy of concepts: straight line $\in$ flat curve $\in$ affine curve $\in$ set of points etc. and a hierarchy of relationships. There is a projection from a graph G into a graph H if G is more general than H. A rule R is applicable to a graph G if there is a projection from the hypothesis of R to the graph G, and then the conclusion of R can be added to G. Constraints are a pair of graphs, which can be negative (if A is present, then B must be absent) or positive (if A is present, then B must be present). In this case, there are:

- 5 negative constraints, such as the incompatibility of $\in$ and $\notin$;
- 17 rules representing axioms;
- 10 rules representing implicit knowledge;
- 2 definitions of relation types;
- 1 definition of concept types.

Such a conceptual model can be used in knowledge management and in automatic theorem proving.

It was asked whether a description logic language could not be used instead. Buchberger asked if set theory and first-order logic were not sufficient for this.

1.9 Panel discussion

The panel was chaired by Michiel Hazewinkel, and included all the participants. Buchberger reminded the meeting that the idea for this MKM workshop came from a workshop where he was invited to Amsterdam by MH. Various questions had been circulated, and some were addressed by the panel.

1. Monolithic or modular? Farmer said that the system he envisaged in his talk was not necessarily monolithic, but that the ultimate system would have to address the whole

of mathematics. MK had not envisaged Farmer's system as monolithic, rather as the specification of a protocol. He would like to see a distributed system. Baumgartner said that his system need not be web-based, and was not necessarily intended to model mathematics from its foundations. Ion said that one could not have a monolithic system for mathematics: there is just too much of mathematics. Buchberger said that it was much too early to conceive of a monolithic system: many different approaches needed to be explored. Even the question of which logic to use was not obvious. Here he saw the distinction between the formal system and the foundational system. It was not always necessary to go back to first principles: in his project it would be unrealistic to reduce the whole of Hilbert space theory to set theory. It was stated that there were various needs for knowledge management: experimentation, formal verification etc., and this found wide-spread approval. Constantini (Rome) said that it was necessary to have some framework for communication. Trybulec said that, for every Mizar article, there was an (information-theoretic) ancestor. Though there were over 7,000 articles, the maximum height of this tree was 12. Farmer summarised by saying that people should be able to use multiple logics, but there needs to be some way of communicating between them. MH said that one could waste an enormous amount of time making sure that all the systems could communicate with each other, and we had to do something about "approximate communication".

2. "Is an MKM feasible? The MKM conversion factor is likely to be 4 or 5 (analogous to the de Bruijn factor)." Ion said that the answer was to start, in order to improve the technology. MK said that Psyc project, in A.I., was a pretty miserable failure, since (a) the methods were not ready, and (b) the attempt was monolithic. He was more optimistic about the current activity. Borwein asked what the boundaries of mathematics were defined to be. MH said that this was a difficult question. Borwein said that we needed to formulate some realistic sub-goals. MH said that we ought to manage most that part of mathematics which was most used. Special functions probably fell into this category. MK said that we should not aim at being complete. AC said that we should try, at least, to standardise the notations. HGS asked precisely what Knowledge Management was, and why was Mathematical Knowledge Management special. MH said that several of the talks had addressed this, partly giving the reason that mathematics was well-structured. Ion said that the world assumes that mathematics is well-defined, so this was a good test case. There was a call for stating precisely what the issues of MKM were, at the mathematical, logical and communication levels.

6. "Formal proofs are fragile — Caldwell." A slight change to a theorem prover can mean that previous proofs no longer work. JHD suggested that it was important to store some of the intermediate stages, e.g. the output of the tacticals. Buchberger said that some informal, human, proofs were also fragile. MK thought that fragility was largely a technology problem, and a failure of knowledge management. Borwein commented that the fragility of human proofs depended on the size of the user community. AMC commented that there was, in fact, no proof of the classification of finite simple

groups, but everyone believes it. Hardin said we had to distinguish between truth and confidence.

7. "Most mathematical knowledge is in the heads of mathematicians, and is not written down — Farmer". Hardin asked if it was possible to make this knowledge explicit. In France, there was a problem here in the area of nuclear systems. JHD drew attention to Traverso's submission to this conference, which drew attention to the amount of undocumented knowledge in computational algebraic geometry. MH asked if this was unique to mathematics. HGS said that one cannot learn mathematics from books alone, which is a converse of this challenge, and solving this problem would be solving a hard A.I. problem. Borwein said that recoiling from the excesses of Bourbaki had made certain things more explicit than before.

8. "Who profits from MKM?" Farmer said that most users of mathematics need a very small amount of mathematics. MK thought that much of what we were doing here carried over to Physics, or much of hard science or engineering. This was where to find rich and receptive customers.

MH concluded by reminding all that there would be a meting in Amsterdam in the middle of November. What tasks should be given to the group there? MK thought that a convincing answer to point 8 was the key. AMC said that the setup of OpenMath was flexible enough to be used as a basis for communication. JHD said that he thought OpenMath was indeed powerful and flexible enough for the communication of *mathematical objects*, and the challenge was to convert that to the communication of *mathematics*. HGS wanted an (optimistic) proof of the existence of an infrastructure within the mathematical community to perform MKM. HGS said that answering question 8 should also link to other areas, such as the Ontoweb Thematic Network.

2 OpenMath Workshop 27.9.2001

Approximately 20 people were present.

2.1 MathML tools

David Carlisle spoke to this topic. He said that MathML had now been around for a while, so that it was appropriate to ask what tools had been developed for it. The state of browsers was summarised as follows.

- Mozilla+MathML — with a presentation MathML compile-time option (provided that the font support is there: apparently a problem on the Macintosh).

- Mozilla+XSL+MathML — will let one add content MathML. Still needs the compile-time option. The XSL implementation is currently somewhat suspect in Mozilla.

  He noted that the MathML, which was not suspect, was not in the Netscape version of Mozilla, whereas the suspect XSL support was.

- Amaya supports presentation MathML.
- Internet Explorer: Microsoft has said that they will not support native MathML. Currently the mechanisms to download a plug-in/applet into IE 5.x are somewhat clunky, which is a serious problem for the take-up of MathML. IE 6's support of XSL is currently broken, which is a problem.

  If XSL, DOM support and Javascript work, it is possible to simulate MathML support in IE: this should mean that documents do not need to state which plug-in they require, and the style sheet can work out what is available, and use DPC's simulation otherwise.

He discussed the state of Techexplorer and Webeq. Design Science (who make Word's equation editor and MathType) were to have demonstrated at Linz: DPC demonstrated this, and showed that one could cut from Mathtype and paste as MathML.

There is a problem with type-setting quality. There are three TEX packages that read MathML. It is also possible to use XSL (say) to convert MathML into TEX, though these are not quite as good as they could be, typically due to the amount of hand-tuning in TEX input. Wolfram Research would claim that Mathematica is the way of typesetting MathML, and this is viable.

tex4ht and omega are DVI-based tools for converting TEX into MathML: since they are DVI-based, specialised mark-up (e.g. that used in DLMF) will have been lost.

In questions, PL asked for a MathML resources list, to contain the current state of this information. The new W3C staff member for MathML (Max Froumentin) is hoping to work on this. MCD commenting that Maple 7 had support for downloading relevant applets. PL asked for support in converting presentation (LATEX or presentation MathML) into content MathML. Several people commented that you needed to know the context. WAN mentioned the work at Western Ontario, which should be appearing on their Web site.

2.2 OMDoc in use

Baumgartner spoke to this work, joint with Antje Blohm and Margret Gross-Hardt. He was speaking from the customer's point of view. The In2Math project is about electronic teaching material, where the student has access to domain tools, e.g. a computer algebra system. One application at Koblenz is in teaching logic. He emphasised that students might wish to use several logic tools, so portability of formulae was important. There was also a requirement for different styles of input, e.g. clausal logic, predicate logic, as well as the STRIPS planning language.

He complained about the readability of OpenMath, and the difficulty of tool writing. OC pointed out that arity checking, etc., was possible, but not yet implemented in the Java library. This became a debate on the rôle of phrasebooks.

After talking to Kohlhase, he thought that they could use the OMDoc "presentation" element to control associativity etc. Hu summarised by saying that there were three options.

XML intuitive, but home-made.

OpenMath standard, but little structure and tool writing was difficult.

OMDoc A practical meta-stylesheet approach, but still little structure.

SB suggested a fourth possibility.

binary Dissociate semantics from presentation, by writing a (XSL) style-sheet to convert OpenMath to presentation MathML.

DC felt that it would be easy to map from any of these formats to any other, so that the problems were fundamentally the same. PL pointed out that there was also a problem of user input in this application.

2.3 Electronic Books and OpenMath

Reinaldo Barriero (Eindhoven) spoke to this title. He said that MathBook, an XML application, should be capable to produce Web applications (HTML pages, JSP pages, XML documents) and also LaTeX, and hence DVI and PDF. They had been using OMDoc, but found this too rich. He defined a JSP page as a text-based document specifying how to process a *request* to generate a *response*. This is claimed to be "write once, run anywhere", to separate the rôles of developers and authors, to encapsulate functionality (JavaBeans and tag libraries). He illustrated this with a sample mathematical page.

We continued to describe the IDA tag library. This allows actions such as interaction with back ends (currently GAP and Mathematica), transformations (OpenMath $\rightarrow$ MathML), casting OpenMath objects (sets into lists, lists of lists (as returned from Mathematica) into matrices), parsing into OpenMath etc., flow control, and working with scopes (in the sense of scopes across pages and within sections of the book) and variables. He illustrated this with a use of a programming CD to send a Fibonacci program to Mathematica. He claimed that this was much shorter than the JavaBeans approach from the author's point of view.

He summarised future work as being:

- more phrasebook work;
- new CDs;
- new tags;
- efficiently producing good presentation MathML from OpenMath — essentially a MathML phrasebook[6];
- more work on MathBook and its relationship to OMDoc.

PL asked whether JSP could be used to generate LaTeX: the answer was affirmative. OC asked where this information was available: it is currently in private CDs. DC commented that JSP itself was not XML, and asked what problems this caused. SB remarked that,

[6] which might also mean a way of giving presentation hints in OpenMath.

unlike what had been shown, the Java code could be external to the page, and stored in the bean, and there is an XML encoding for enough JSP to support this. AMC commented that there was an "algorithm" CD, and this meant that the same page could work with GAP as well as Mathematica.

2.4 Special Functions in OpenMath

JHD spoke to this issue. He said that one unsolved problem was functions defined by analytic continuations. Farmer said that there were three kinds of definition.

1. An object defined by a formula, as for arctan in the `transc1` CD.
2. An object defined uniquely by a set of properties, which might incorporate JHD's `odesolution` and similar concepts.
3. An object specified by a set of conditions, but not necessarily uniquely.

The second class might subsume JHD's various worries about specification of special functions. JHD said that he would like to work a few examples.

2.5 Changes to the OpenMath standard

DPC spoke to this issue. he said that there were various levels of changes.

- Textual corrections, e.g. `http:www` → `http://www`. There are other problems of these natures: would fixing this result in a change of version number? The formal standard probably ought not to contain the change log and marginal notes it currently contains. So what is the formal reference version: LaTeX, PDF or what? He is now of the opinion that there should be a master XML source as the normative version. Again, what implications for the versioning? AMC commented that there had been discussions about version numbering. MCD thought that these were about CD versioning, rather than versions of the standards. JHD thought that the change of normative language was a significant change, e.g. to 1.1, and there seemed to be some consensus here. We would have to continue to distribute the PDF version.
- MathML-related changes. Some might be errata (e.g. JHD's comments on differentiation) and others might be implied by changes in MathML. In theory, MathML might require changes in the standard, though this should be avoided as far as possible.
- AS's concerns in Berlin about binding symbols. The standard specifies that binding multiple variables is equivalent to multiple bindings, which can mean confusing variable capture.
  - Leave it as is.
  - Specify that B is a symbol.
  - Specify that B has no free variables.
  - Specify that the free variables of B exclude v_i.

  - Specify that this equivalence is only true after α-conversion.

  JHD commented that there were implications for his `ODEsolution` syntax. Some people thought that the comment on equivalence should be deleted.

- The formal definition of an `OMF` refers to the IEEE standard, which is not very available. The view was that there was not much we could do about this, as it was the official reference. More seriously, the IEEE standard is parameterised, and AMC had queried whether we had tied them all down, and specified byte and bit order completely. DPC thought that we had, but this issue does require checking. A worked example would be helpful.

- MK would like to extend OpenMath with a genuine Record construct, so that a Group could be defined as a record of (carrier,operation). He had a worked suggestion: `<OMR>` to start a record, individual children being `OMAVP` pairs, and `<OMSEL>` as a selector. DPC pointed out that this could also be achieved by adding a `record` CD, which would not require a language change and not break all existing software. SB pointed out that, for fixed records, nothing would be needed. The feeling of the meeting (explicitly checked by JHD) was that the most that should be done was a `record` CD.

- The OpenMath standard essentially predates XML namespaces. At the very least, we should automatically add the OpenMath name space properly, but this change should be made as soon as possible. PL added that OMDoc had already done this. Could the XML name space be used to make OpenMath less verbose, so that, with the appropriate declarations, `<alg1:times/>` could replace `<OMS name="times" cd="alg1"/>`. Alternatively, we could use `<OMS name="alg1:times"/>`. JHD argued against the suggestion of `<OMS name="times" cd="mathml:alg1"/>`, and DPC agreed. Using name spaces for CDs would mean that globally-unique names for CDs were no longer needed. DPC was worried that we didn't have a plausible mechanism for allocating CD names.

In answer to a question from JHD, DPC thought that there was no longer an 8+3 restriction on CD names. He was asked to check this.

3 OpenMath Workshop 28.9.2001

3.1 The M@inline Project

Peter Sander et al.

Initially Peter Sander gave an overview of the project.

The Group had been running for 18 months, Multimedia Applications Involving Non Linear Information for Networked Education.

Affiliated with I3S (CNRS) & ESSI (IT & engineering school, Univ. Nice).

Previous work on JOME in OpenMath project, Industrial partner loses rights if does nothing in 6 months so soon have full rights to JOME and make public CVS.

Courseware: developing introductory course on Java. Used powerpoint c 800 ppoint slides. need something better. Aim to develop platform with database of 'bits' of course. Front end extracts XML to be delivered as lecture notes, slides, web course notes, etc. 2 student projects (XCC, XIND) implement prototypes:

XIND is an SVG diagram creator (e.g. commutative diagrams).

Infrastructure project: lack of information systems at the school. student registrations on one machine, mark information in excel on instructor's machine, course notes elsewhere. so developing in house project to develop coordinating software, not competing with commercial blackboard/webct

Developing content: something better than latex/latex2html is required to produce more interactive course.

Current projects: OM (RIP), OM TM, Trial solution (database of "slices" of latex courses), EML (education modeling language), TICE (technology and infomatics applied to education) something better than French translations of (e.g. MIT) courses, need to develop native French/European content, E-MIAGE.

Peter mentioned close collaboration with the CAFE group at INRIA, Mark Gaëtano then spoke for the INRIA group.

CAFE: Computer Algebra Functional Equations: mainly algorithms for differential equations. Started 1998 after SAFIR. affiliated with INRIA and I3S.

The group develops interfaces to Computer Algebra systems, especially for ODE. More generally develop software tools for mathematics.

Manual Bronstein, developing algorithms implemented in Aldor (and some maple). what to do with this code. Aldor is good but not widely available or used. Aim to provide implementations as web services so end users don't need Aldor locally. Notes that these are very specialised systems (not like maple etc).

Hope to use OM/MathML and the formula database still being developed, hope to release as a web service. A prototype database was developed on OM project, but next release more usable. hope to implement a reasonable part of A&S.

Projects: OM (RIP), OM TM, Cathode.

Finally Séphane L. spoke on SVG rendering of mathematics.

Previously worked (MSc) on emath (mathematical editor) at INRIA: common interface to Computer Algebra (maple, Mathematica, etc), then for PhD on OFR: Optical formula recognition: recognise printed/handwritten formula. (96-2000) working on interface to Computer Algebra, sending OM to Mathematica. Joined m@inline in 2000.

Courseware.

static display: XML/XHTML/SVG (MathML2SVG) (SVG, scalable vector graphics, is a W3C Recommendation.)

dynamic display : live documents (XIND, Xind is not Dia)

Demonstrated rendering of Content MathML by the mathml2svg tool using Adobe SVG plugin in the browser. Demonstrated how this is movable/zoomable/scalable.

FIXIDEA (son of JOME) editing structured mathematical documents (including charts graphs etc) Java using svg canvas for display.

Future work:

Enhance MMLC -¿ OM. Add support for new operators.

Develop online course with mathml2svg and/or OM2svg, FIXIDEA, OM Broker.

3.2 In Search Of The Semantic Spider

Stephen Buswell spoke on the relationships between OpenMath and the Semantic Web.

What is the Semantic Web: Collections of ideas and technologies, web services, 2-way transactions, user profiles

What is OM: 2.5 XML languages (CD OM Object, OMDOC. Mathbook,...) and mathematical resources (CD library, phrasebooks, OM aware applications, OM stylesheets).

Semantic Web has origins in classical metadata:

Genealogy, real estate, pornography (MD for parent/child protection).

librarians (card indexes)

screen scraping

search engines. yahoo (human, top down) google (automatic, bottom up)

e.g. "Apple sacks fuller": Austrian fruit farmers...

first steps: PICS:
grading by 3rd party proprietary tag values.
HTML meta element data.
RDF
Dublin Core (set of standardised metadata labels)
RSS (RDF Site Summary)

Metadata in RDF, modelled (equivalently) as a labelled directed graph. or Triple (Property Resource Value).

RDF has classes, subclasses, transitive relations. Vocabularies constrained by RDFS (RDF Schema restrict values of properties)

value-set for terms: data dictionary, term inter-relationship: ontology.

example (namespace soup, daml, RDF, RDFS namespace prefixes all intermingled)

RDF processor should be able to skip over non RDF namespaced elements.

RDF is ontology-neutral.

Structure of RDF XML Tree is not the structure of the information described.

Ontologies: next steps.

- OIL
- DAML
- Frameworks for more complicated ontologies

  Domain-neutral (cf KR languages)
  Content not process (cf KQML)
  type checking for consistency tests
  higher ontologies for domain specific reasoning.

A web services model
UDDI WSDL (and w3c activity XML Description language) SOAP
User Agent Model - CC/PP

CCPP Composite Capabilities/Preferences Profile.

expressed in RDF. Driven by mobile phones, allows to describe limited capabilities of client.

CommonKDS Transaction Model, two way information flow, negotiation to agree on mutual requirements.

OM-SW comparisons:

- CD — micro-ontology
- CD Group — domain specific ontology
- openmath.org — application-specific UDDI server

Next steps:

- restructure CDs in RDF, easier to do FMPs that are properties of two symbols.
- use URI as unique OM CD symbolname identifier
- Build mathematical services description language over e.g. WSDL
- develop transaction models
- develop agent models over e.g. CCPP
- openmath.org application specific UDDI
- visualisation and navigation

3.3 Semantically Encoded Mathematics On The Web

Paul Libbrecht spoke on the ActiveMath learning environment.

Current solutions:

- GIFs: latex2html
- Mathml
- HTML and Unicode symbols
- PDF from TEX
- applets (e.g. jdvi applet from Berlin)

Changing standards buggy browsers, unpredictable installations.

Using semantic encoding:

- goal for author: write what you mean

requirements

- standard declaration of symbols
- separation of presentation and content
- organisation of content units in a logical way
- possibility of explanations on formulas in presented content
- convertibility to multiple targets. (latex html, maple, mupad, ...)

Using OMDoc, an XML language based on OM, Has items such as definition proof example assertion. Allows dual encoding: machine and human. May be thought of as an ontology that is extensible (symbols have definitions). OMDoc group has developed several XSL stylesheets converting to html/latex.

Adaptive learning:

- add a user model
- add a presentation planner
- add pedagological rules
- get a content presentation that is adapted to the user
- experiment with pedagological theories

Example: one student just sees example questions, another sees full training material, depending on their "profile".

Active Math has been developed at DFKI and University of Saarlandes. It will be licensed open source. A stable version is expected to be released next year.

Producing Content for OMDoc, involves two types of author: choosing mathematical systems, and developing interfaces (developer), writing content (author).

Writing omdocs is not easy, especially OM part.

The QMath processor is an editing tool that takes a latex-like syntax to OMDoc. Also hope to develop a swing based visual editor.

Future work includes:

- enhanced presentation planner
- more configurable' visual authoring
- integration into uni-0nline (for example)
- gadgets: drag-and-drop from content, slide generation, copyright display etc.
- content being developed within BMBF project. conversion of Analysis Individuell, a statistics course, formal methods course.

In conclusion, semantic encoding allows authors to forget the dirty details of the browser, and allows for re-use of content.

ActiveMath provides the shell: authors and developers can now now start writing!