# Esprit Thematic Network — OpenMath Notes of Berlin (ZIB) meeting 6/7.8.2001

These notes drafted by James Davenport
Present: Stephen Buswell (Stilo), Olga Caprotti (RISC),
David Carlisle (NAG), Arjeh Cohen (RIACA),
Stephane Dalmas (INRIA), James Davenport (Bath),
Mike Dewar (NAG), Winfried Neun (ZIB),
H.-G. Stork (CEC Luxembourg), Andreas Strotmann
(Tallahassee/Köln), Victoria Wicks (Springer)

September 10, 2001

## 1 Introduction

### 1.1 Administration

WN gave some administrative details.

MCD reported apologies from St. Andrews (whom he would represent) and DFKI (Saarbrücken). It was reported that Explo-it's representative was unable to attend. SD would represent the University of Nice as well as INRIA. He proposed that JHD be elected as Chair of the Network, which was agreed *nem. con.*

He then reminded the meeting of the objectives of the network:

- evolution of MathML;

- applications of OpenMath in Education;

- maintenance of OpenMath.

## 2 Electronic Books — AMC

AMC spoke to IDA II — the continuation of the Interactive Document on Algebra project. Springer had published a commercial version of the result of that project. He is aiming for a new version at September 2002. This will mostly be the same content in new technology.

- MathBOOK is an XML DTD under construction, based partly on OMDoc and on DocBOOK. This is intended for the construction and editing of interactive mathematical books. It is based OpenMath.

- Java technology throughout. XSLT style sheets[1] translate into LaTeX, HTML, XML (using Mozilla/MathML) and JSP.

- JSP (Java Server Pages) will replace the Java plug-in. It is necessary to keep information on the page in use. JSP has four native scopes: session, page, application and request; but he expects to create new scopes, corresponding to OMDoc's ranges.

- Phrasebooks. RIACA uses their own "generic" phrasebooks, supporting dynamic loading of CDs. They currently interface with GAP and Mathematica. They intend to extend to CoCoa (not hard) and Maple (harder, since this only supports content MathML currently[2]).

- CDs. They intend to produce:

    - programming CD, which should include constructs such as `block` and `conditional_block`;

    - polynomial CDs — they are writing their own;

    - Group Theory CDs;

    - IDA-specific CDs.

    DPC pointed out that it would be necessary to run a server program in the local machine, which might conflict with local security standards.

- There was a lengthy debate on the technology, with no very firm conclusions. MCD pointed out that some of the delays were caused by licencing the version of Mathematica, which led to a lively debate. SB noted that Scientific Workplace came with a Maple engine, but if one wanted a Mathematica engine, one had to buy Mathematica separately from Wolfram. This was both more expensive and far less convenient. MCD pointed out that there was $W^3C$ interest in the subject of digital rights.

## 3 Sorted Generalised Quantifiers — AS

How do we represent sorted quantifiers in OpenMath? Kohlhase had said "let's discuss sorted quantifiers". AS gave the following examples:

- $\forall x \in \mathbf{R} : x^2 \geq 0$;

- $\{(x, y, z) \in \mathbf{N}^3 | x^n + y^n = z^n\}$ — note that there are naturally two arguments here: $\mathbf{N}^3$ and the predicate $x^n + y^n = z^n$, and so is hard to express in the current OpenMath;

- $f : \mathbf{R} \times T \to \mathbf{R}$.

He required:

- clean scoping rules for OMBind objects;

---

[1] Thanks to David Carlisle.

[2] MCD pointed out that Maple 7 now supported loading libraries into the kernel, which might help. AMC pointed out that Mathematica's `fullform` construct was very helpful, and it was a pity that Maple did not have the equivalent.

- a Categorical Type logic for all OpenMath objects;

- explicit distinction between type and range — not generally made in mathematics.

On the first point, in $\texttt{OMBIND}(q, v, e)$, the OpenMath standard should state explicitly that the scope of $v$ is $e$ (and the consequence that $v$ is not bound in $v$ or $q$). Also, the currying rules in the OpenMath standard (pp. 11–12) are not helpful. If there are qualifiers on the quantifier, the scope rules arising from currying are bizarre. He would like to see the currying rules removed.

One solution would be to attach scoping information to binding constructs via signatures. He suggests a Lambek-style categorical system, where $\bullet$ represents application and $/$ represents abstraction:

$$bind(b, x, y, e) : tb \bullet (tb/tx/ty).$$

In general

- quantifiers are $forall : (X \to T) \to T$;

- generalised quantifiers (ECC) are $the : (X \to T) \to X$;

- 

He suggests that a binder is *any* OpenMath object expecting a unary function as its first argument. "Exists uniquely" can therefore be defined as a compound expression:

$$\exists! \equiv \lambda P.(\exists x.P(x)) \wedge ((\forall y, z.(P(y) \wedge P(z)) \Rightarrow y = z.$$

## 3.1 Categorical Typed Logic — d'après Lambek

AS gave a brief summary[3].

- Abstraction: $A/B$ returns something of type $A$, given something of type $B$, corresponding to STS's $mapsto(B, A)$.

- Application $A \bullet B$ applies $A$ to an object of type $B$.

- Note that $(A/B) \bullet B$ is of type $A$.

- $bind(a, v, e)$ is of type $A \bullet (E/V)$. This is not quite the same as ECC or STS. Binders themselves are therefore generally of type $Y/(E/V)$.

- Attribution can then be $attr(A = B, E)$ having type $(A \bullet B) \bullet E$. This is very different from STS and ECC.

He emphasised that this was a framework for specific type systems, and claimed that it:

- factored out the structural semantics;

- supported "radical lexicalism": the type logic on atoms (e.g. compatibility) is orthogonal to this theory.

---

[3]See Moortgat in the *Handbook of Logic and Language* (ed. van Bentham, ter Meulen).

DPC said that AS was essentially saying that $\lambda$ was the only binder: a point of view that he had (unsuccessfully) argued. OC said that there was currently no formal semantics of `OMBIND`. AS felt that this was a weakness of OpenMath. There was a problem of recursion: a CD cannot define the semantics of CDs — the standard may need to use some meta-language to express these semantics.

# 4 Recent Development in MathML and W$^3$C — DPC

DPC introduced his paper as describing the developments since the end of the OpenMath Research Project. His presentation was based on DocBOOK.

## 4.1 MathML Developments

MathML was released as a W$^3$C *recommendation* in February 2001. It included: the following items.

- New content elements.

- `csymbol`, which is very relevant as an analogue to `OMS`;

- new presentation elements (e.g. labelled equations) — probably not very relevant to OpenMath.

- Unicode 3.1 and 3.2 character sets — Unicode 3.1[4] happened about March 2001, and 3.2 is still being voted on. Much software does not support the "above $2^{16}$" characters. The MathML DTD supports a many-one mapping from MathML names to Unicode names, but SB pointed out that this was an issue in the world of schemas.

The MathML group has been re-chartered for May 2001–2003, but the scope is *not* to produce MathML3, rather to maintain and clarify MathML2.

The Document Object Module (DOM) is important to allow different tools (e.g. Techexplorer) to access relevant parts of an in-core document. The political impetus here is from SVG[5], which has similar requirements on browsers and the DOM.

There is a W$^3$C requirement for bi-directional text, but Patrick Ion was only able to find one example of bi-directional mathematics in the whole of his (mathematical Reviews) database. The MathML WG has managed to argue that this requirement was not necessary for managed, and changing $3 - 2$ to $2 - 3$ might well be unhelpful.

The MathML Working Group is explicitly chartered to co-operate with OpenMath, whose contributions to MathML2 were "measurable".

---

[4]This takes Unicode about the $2^{16}$ characters boundary. In particular `0x1D4xx` and `0x1D5xx` are extra mathematical symbols, e.g. blackboard bold (except for **C** and a few others, which were already in `0x029xx`), fraktur (actually 3.2), script etc.

[5]Currently a W$^3$C Candidate Recommendation.

## 4.2   Other XML Developments

XHTML Modularisation became a Recommendation just after MathML, allowing extensions, such as MathML to HTML.

XML Blueberry (which would have been called XML 1.1) will change the set of well-formed documents, due the change in definition of "character". The change is politically necessary, but will be painful in practice. There will also be a change in the definition of white space: adding EBCDIC newline — probably a bad idea.

Xpath2 is a joint development of the XSLT Working Group and a new XML Query Working Group.

Xlink is a Recommendation (finally) to generalise the HTML hypertext link feature. There are very few implementations as yet.

XML Schemas are "probably a good idea". This is now a Recommendation. See OC's talk later.

The "Semantic Web" activity is currently very high-profile in $W^3C$. An important element is the Resource Description Framework, which is a form of metadata. Tools for it are beginning to appear, and MathML and OpenMath should take advantage of these. SB pointed out that RDF is a framework, and there are layers above it beginning to appear with semantic content. We could therefore consider RDF-based formulations of CDs.

SOAP is a lightweight protocol for XML-based remote procedure call.

## 4.3   Tools

DPC demonstrated IE5.5 (his private version) rendering HTML with embedded MathML, using CSS font handling. He uses XSLT to cope with the fact that IE5.5 does not handle Unicode $\geq 2^{16}$. The demonstration file also renders in Amaya (which ignores the XSL) and a MathML-aware Mozilla (if one turns off the Mozilla XSLT engine). The current $\beta$ of IE6 has the plane-1 support turned off[6], but DPC hopes that this decision will be reversed. This problem cannot be circumvented by XSLT, since the document containing plane-1 is rejected before XSLT is called.

AS asked if RDF supported variable binding. The answer was "not directly".

# 5   Content Dictionaries — JHD

See separate presentation/paper.

One topic was Eindhoven's own polynomial CD: maintaining variable name (perhaps misunderstanding of student looking at old polynomial CD not current Bath polynomial CDs). On the topic of the `polyslp` CD, AMC commented that `monte_carlo_ eq` was probably more general than polynomials.

AS wondered whether we should use the `arith1 times` or need new symbols in the dimensions CD: note that 0*1m is 0 of dimension "length". DPC pointed out that MathML are interested in adding units/dimensions in any future mml, and we should coordinate.

In the context of groebner, AMC said that Maple and GAP would have is... operators. AS pointed out that one of first OM applications for sets of

---

[6]In the document parser, but the XSLT stylesheet is parsed by a fully-conformant parser!

inequalities used formula with "and", and the `groebnerBasis` constructor is similar. MCD added that this was similar to Matrix: a constructor symbol rather than attribute.

There was a debate on special functions, especially "to curry or not to curry". The general feeling was to curry where appropriate, e.g. Bessel.

JHD had stressed the distinction between an "algorithms" CD rather than "programing", which was supported by AS. OC asked if this meant we should have a Recursion theory CD. AMC said that their current programming CD is modest trial, aiming to do what Fateman fights against: reusable code. Involves casting returned values to expected OM "type".

# 6  Schemata — OC

A DTD can be viewed as a mechanism for constraining the *use* of markup. The XML Schema Working Group is looking at schemata as means of constraining the doucument's

- structure;

- content;

- semantics.

SB said that there were three reasons driving schemata:

- DTDs are not written in XML, whereas schemata are;

- in application-application exchange (e-commerce) we need data types, and this is impossible with DTDs;

- politically, Microsoft is opposed to DTDs.

There is a Proposed Recommendation of March 2001, divided into Primer, Data Types and Structures. OC contrasted

**DTD** `<!ELEMENT OMI (#PCDATA)>`

**Schema** `<xsd;element name="OMI" type="xsd:integer">`

DPC stated that Microsoft's XML-parser version 4 implemented Schema. SB mentioned Xerxes. AS asked whether the XSD built-in types were compatible with OpenMath: OC said that `string` was, but that `integer` probably did not cover the whole range of `OMI`s. In the case of the syntax for names in `OMS`s, the XSD restriction mechanism can be used to quote the regular expression in the OpenMath standard, which cannot be expressed in the current OpenMath DTD. DPC stated that, in IE, one can right-click to get a document validated against a DTD or Schema. OC pointed out that schemata are closely coupled to name spaces. She posed as an open question the relationship between the nesting of these name spaces and the structure of CDs. This mechanism (if there were an OpenMath schema) would allow OpenMath to be included in other schemata, e.g. physics and chemistry.

She said that the major open question was how much of the OpenMath standard could be encapsulated in an OpenMath schema, and how much would

have to remain as text. The automatic DTD→Schema tools cannot add the extra information that is not in the DTD, but only in the OpenMath Standard. Kohlhase has a hand-crafted schema for OpenMath as used in OMDoc. SB said that schemata could be used both for encoding OpenMath objects and for OpenMath CDs — the latter being the more difficult. AMC asked if RDF and Schemata were linked: SB said that RDF was an XML Schema for encoding ontologies, but other schemata tended to build on RDF.

# 7  Online Mathematical Services — MCD

MCD had two motivations:

- to move away from the technology to look at the business case;

- to start Semantic Web discussions.

He claimed that there were three markets:

- Research & Development: modelling and experimentation — here the needs are unpredictable;

- production systems — here a specific mathematical process is run routinely, which is a component market;

- education — teaching assistants (e.g. Maple with worksheets) or true interactive books.

Increasingly, NAG is selling into the second market as well as the first. NAG does not see the third as its core market (though it sells components to Maple, which does sell into this market). Existing market demand needs

- embeddable components — Excel addins (sold directly, and via the Web), and components embedded into Maple and PeopleSoft — this is a high-cost long-latency activity in terms of doing the embedding;

- easy installation[7] and maintenance[8], which tends to lead to the delivery of a packaged system that does not take account of specifics, e.g. Maple has the sequential NAG library, even if you are running on an SMP and have the NAG SMP libraries;

- "Just in Time" purchasing, especially in the non-University sector.

Therefore the commercial considerations behind moving to mathematical services are (from NAG's point of view):

- improved time-to-market;

- a reduction in the viable product size;

- reduction of product costs and sales overhead;

- quantitative marketing data — hard to get for sales of complete libraries;

---

[7]Nevertheless, some corporate customers do want licence management.
[8]Quite difficult to get right: Web-enabled patches versus 'point releases'.

- better access to today's primary computing environments, e.g. Excel, Maple, Matlab[9];

- better product updates;

- products versus services — the continuing relationship invoked by services is beneficial to both sides;

- "web culture" — if one asks, say, Google, for 'optimisation' one gets to NAG's home page, but then the potential customer has to move to telephone contact, with all the problems of time zones, physical delivery etc.

From the users' point of view:

- easy access to state of the art functionality, from different environments and locations;

- advice and guidance (and explanations);

- reproducability and predictability;

- reliability.

One important point is scalability: the services market deals with occasional access via the Web, first-time access and comparative shopping, local plug-and-play. The product market delas with large problems and physical infrastructure difficulties.

He therefore proposed a project, **M**athematics **O**n the **NET**, based around:

- Services

  - user interfaces
  - mathematical servers
  - databases

- Brokers —

  - register local services;
  - route requests for services;
  - handle logging, payment, access.

- Distributed Object Manager —

  - objects can be encoded in application-specific or neutral ways;
  - constant/mutable etc.

He summarised the main points as being:

- protocols and ontologies (hopefully using generic standards with application-specific content);

---

[9]The Linux version of a Matlab toolkit is a commercial success.

8

- security and privacy — AMC commented that this was probably not mathematics-specific. MCD agreed at one level, but pointed out that the *data* often need to be encrypted, not just the e-commerce charging data as is often the case today;

- infrastructure (brokers and object managers);

- tools for building new services;

- examples.

# 8  CDs — JHD

JHD spoke to his plans and questions.

## 8.1  Yesterday's Decisions

- Augment `list2` with (at least) nil, reverse, append.

* Agreed.

- Attributes to approx: abserr, relerr and $O$.

  The question was raised of attributes or new CD of symbols? There were worries about round tripping with MathML. general feeling: attributes but in a new CD not in core.

* Agreed for abserr and relerr; $O$ referred to the asymptotics CD (in draft).

- More work on the units CDs, especially FMPs.
  (get a draft to MathML for MathML3)

- Special functions:

  – curried where sensible;
  – `<OMS name="J" cd="Bessel"/>`.

  Mike: send Bruce M's + Stephen W's CDs to James. Stephane: send INRIA's CDs to James.

* Agreed — check with NIST.

- Publish a draft of `logic3`.

* Agreed.

## 8.2  Today's Decisions

- Let `polyd` name its variables:

  **Either** an attribute `variable_names`;

  **Or** a second constructor which named the variables.

SB argued one could have string of variable names, white space delimited. AS said that the 'one signature' rule could still be maintained but one could have 2 or a list of variables of length 2. Then there was general agreement on an attribute "variable names", but should names be OM String, or OM Variable? Stephen/Arjeh: if you use OMV should be part of structure, not attribute if you use OMV to allow substitution etc. Should naming be handled by convert operator from the generic `poly` CD? The meeting converged on a general feeling now: always have a list of variables, (i.e. change status quo) and the CoCoa phrasebook would have to work round that change.

* **Change** the `poly_ring_d` constructor to require explicit variable names: cross-check with CoCoa phrasebooks.

- `substitute` command:

  **Either** in `logic3` (one symbol);
  **Or** in a `subst1` CD, as a multiple-in-parallel operator.

  This led to a detailed discussion of the semantics of apply/lambda. OC pointed out the theory of prolog substitution, Horn clauses etc.

* Use lambda-abstraction to express substitution.

- Do we want `is_groebner` as well as (instead of?) `groebner_basis`?

* Agreed[10] (as well as).

## 8.3 Near Future Decisions

- Does Arjeh really need another polynomial CD?
- Abstract algebra: James cooperate with Arjeh.
- Algorithms CD: James cooperate with Arjeh.

When does James visit Arjeh (or v.v.)? It seems likely that MKM it the obvious next meeting point.

## 8.4 The Reviewing Process

MCD raised the question of the reviewing process. AMC said that AS was a referee for the MathML group — the largest group, and guinea-pig for the reviewing process. AS said that the turn-round and the workload (on him and on the authors) had made it difficult. One piece of feedback from this attempt is that mechanical checking had to be done before CDs were sent to the referees.

It was clear that the refereeing process had to be made to work, both by this project and by the OpenMath Society.

**JHD**
**MCD/**
**AMC**

---

[10]In subsequent correspondence, AMC wrote "It was my understanding that `is_groebner` would be recommended to be used 'instead of' `groebner_basis`. Although not relevant for the minutes themselves, I think that,

**a)** IsGroebner will carry the GB in its argument just as `groebner_basis` will,

**b)** there is a danger of the `groebner_basis` and `factored` being too *ad hoc* to be recognized as making a statement on the argument."

# 9 OpenMath Standard/MathML Alignment — DPC

1. There are a few typos.

2. There should be a version without marginalia.

3. Currently only available in PDF: should be an XML version, which presumably be the definitive version.

4. We need a versioning policy.

5. No changes are planned in MathML, but changes might arise, which might in turn affect the MathML CD Group, which is an Appendix to the standard.

6. There have been suggestions on the mailing list about making more use of name spaces. Currently OpenMath is one name space: if each CD were a name space, the element

   ```
   <OMS name="times" cd="alg1"/>
   ```

   could be replaced by

   ```
   <alg1:times/>
   ```

   or even

   ```
   <OMS name="alg1:times"/>
   ```

   or even

   ```
   <OMS cd="mathml:alg1" name="times"/>
   ```

   JHD commented that the last broke the rule that CDgroup names and groupings are irrelevant as far as deciding whether two symbols are the same.

7. Schemas are naturally name-space aware, so the previous issue has to be decided first.

8. AMC pointed out that the question of which floating point number was meant by an OMF was not clear. Others disagreed. **AMC**

# 10 Summary — MCD

1. NAG will work on the web site. **NAG**

2. Next meeting 26–28 September RISC-Linz. **OC**

3. Following meeting Feb 2002 near Cannes. **SD**

4. CD refereeing needs to be clarified. **AMC/ JHD/ MCD**

5. A slightly revised version of the standard is to be produced.

6. Proposals to clarify binding are needed. **DPC AS/all**