

# Hermes

## An Effective Converter from TeX into MathML

Romeo Anghelache & Bernard Schutz  
Albert Einstein Institute  
Potsdam, Germany

[romeo@psyx.org](mailto:romeo@psyx.org)

[Bernard.Schutz@aei.mpg.de](mailto:Bernard.Schutz@aei.mpg.de)

<http://relativity.livingreviews.org/Software/>

<http://www.aei.mpg.de>



# Authoring Mathematics for the Web

- Mathematical display on the Web is still rudimentary.
- Most journals with e-portals offer full text only as a download.
- MathML becoming a standard, but not easy to write.
- Physicists and mathematicians write comfortably in TeX. A tool to put this on the web would not only help journal publishing. Research departments and institutes could build better websites, scientific collaborations could exchange information internally more effectively, and mathematical searching could be supported.
- EU-sponsored **MoWGLI** project is developing mathematical databases, theorem tools, journals with common search interfaces.
- *Living Reviews in Relativity* joined MoWGLI to develop a MathML authoring tool. **Hermes** is one of the deliverables.
- **Hermes** is a tool that allows authors to create MathML by writing in LaTeX.



# Living Reviews Family

- LR Relativity first published in 1998 by AEI.
- Open-access review journal, now very widely used.
- Authors keep articles up to date, revisions 1-3 yr.
- Authors write in LaTeX with LR style file, already provide plentiful keyword information for articles and cited references.
- All articles displayed on web in full-text HTML:
  - mathematics included as gif images (not scalable, browser displays inconsistent);
  - HTML enhanced to offer many reader-friendly navigation aids ([example](#));
  - downloads offered in PDF, PS, HTML.
- Conversion TeX  $\rightarrow$  HTML by customizable open-domain article builder (<http://www.zim.mpg.de/projects/toolkit/index.html>).
- LR family growing: Solar Physics is now on web, European Political Science expected soon.



# Living Reviews and MoWGLI

- MoWGLI: LR develops MathML-based XHTML publishing/authoring tools. MoWGLI partners plan a unified search engine that spans all partner sites.
- With MathML, the HTML presentation scales properly and is (almost) browser-independent.
- More interestingly, equations can include metadata to assist *searching* and other bibliographic purposes.
- LRR Editorial Board is committed to enhancing searches:
  - develop a dictionary for relativity mathematics;
  - index equations with metadata to help relatively un-intelligent searches;
  - implement user searches for mathematics;
  - within MoWGLI, implement automated search interface.
- MathML may also permit us to offer *cut-and-paste* of content-MathML from the journal to formula-based software, such as algebraic mathematical programs or to spreadsheets.



# Hermes

## Author: Romeo Anghelache

- There have been various attacks on LaTeX  $\rightarrow$  MathML conversion. Our approach is *not* to try to solve all problems with LaTeX. Simply map presentation(TeX)  $\rightarrow$  presentation(MathML).
- **Hermes** starts from DVI file:
  - no need to re-invent TeX compiler;
  - DVI contains formatting, font information;
  - DVI nearly independent of TeX/LaTeX varieties, packages;
  - must interpret/ignore `\special` constructs for, say, dvips.
- **Hermes** uses `\special` commands itself to preserve mathematical meaning and insert metadata into DVI file.
- Author does this using LaTeX macros provided in a style file. Most are transparent to author: renaming of standard LaTeX. So *Hermes permits authors to use LaTeX to author MathML.*
- **Hermes** is open-source, protected by the GNU GPL license.



# How Hermes works: 4 steps

1. Seed the LaTeX source with `\special` commands
  - Authors use LaTeX macros, never deal with “raw” `\special` command
2. Parse the DVI file, recover the `\special` commands
  - Does not interfere with TeX, leaves rendering untouched
3. Output XML
  - Text mapped to Unicode
  - Math expressions/symbols mapped to MathML
4. Create Web presentation using style sheet
  - XSLT → XHTML+MathML



# 1. Seed the LaTeX source with `\specials`

- In author's article:

```
\documentclass{article}  
\input ltxdef  
\begin{document}
```

...

- In `ltxdef`:

```
...  
\let\oldcdots=\cdots  
\gdef\cdots{\special{cdots}\oldcdots}
```

...



## 2. Parse the DVI file

- $a+b$  gives, in the dvi file :

```
\imath: ' \fntnum10' 'a' \fntnum7' '+' \fntnum10' 'b'  
':imath'
```

- Has sufficient *presentation* info: Hermes translates it to

```
<mathml><mi>a</mi><mo>+</mo><mi>b</mi></mathml>
```

- We're lucky here: there is *content* information too.

Hermes automatically writes:

```
<mathml>  
  <apply><plus><ci><mi>a</mi></ci><ci><mi>b</mi>  
  </ci></apply>  
</mathml>
```



# 3. Output XML

- Text mapped to Unicode
  - Most TeX font character codes generate corresponding Unicode characters.
  - For accents, Hermes uses the Unicode combining characters mechanism.
- Math mapped to MathML
  - Everything defaults to MathML *presentation*
  - Where the meaning is clear, Hermes wraps it in MathML *content*
- Resulting XML:
  - `<article>`
  - `<par>text<mathml>...</mathml></par>`
  - `</article>`



## 4. XSLT to XHTML+MathML

- The XML document will be the reference version in Living Reviews (i.e. the source for searches, etc). Contains document metadata (subset of Dublin Core).
- For rendering in browsers, this is converted to XHTML using XSL.
- Fine rendering also needs:
  - A unicode font installed in the system
  - Mathematical fonts
  - IE6 needs MathPlayer to render MathML
  - Mozilla renders MathML on its own
  - Example available at Living Reviews



# Content and Hermes: possibilities

- **Manual metadata:** modify LaTeX `\equation` macro to add extra argument where author can insert search metadata.
  - Carried through to the XML via a `\special` in modified macro.
- **Automatic metadata:** suppose we want to index every occurrence of Riemann tensor in all equations.
  - Create LaTeX macro `\riemann` that authors use instead of composing  $R_{abcd}$  by hand.
  - `\riemann` has arguments for LaTeX and MathML presentation (indices, placement) and it also inserts a `\special` containing standard search metadata for this concept.
  - As in the manual case, the `\special` inserts the metadata into the XML.
  - Such macros can be easier to use and produce cleaner LaTeX!



# The Future of Hermes

- Currently in version 0.8.4.
  - Testing by processing back volumes of *Zentralblatt für Mathematik*
- When 1.0 is released:
  - Hermes will be open-source, public domain.
  - Living Reviews will use it to process new and existing articles, offer XHTML/MathML versions alongside standard HTML/gif versions for older browsers.
  - For 2005 (Einstein Year), AEI and Max Planck Institute for History of Science (Berlin) will process and make available about 20 original Einstein papers from *Annalen der Physik*, with content MathML and searchable metadata markup.
- If MoWGLI successor project Authomath is funded:
  - LRR will build relativity mathematics dictionary, create search engine, use Hermes to insert dictionary keywords into equations.
- *Partners welcome: like all open-source projects, the key to the future is building a community of users and developers!*



# Links to further information

- Hermes information/downloads  
<http://relativity.livingreviews.org/Software/>
- Hermes author Romeo Anghelache's email and home page  
[romeo@psyx.org](mailto:romeo@psyx.org), <http://psyx.org/>
- Living Reviews portal  
<http://www.livingreviews.org/>
- Living Reviews in Relativity  
<http://relativity.livingreviews.org/>
- MoWGLI  
<http://www.mowgli.cs.unibo.it/>
- Sample LRR article processed by Hermes  
<http://relativity.livingreviews.org/Info/AboutLR/mowgli/article.xml>

Advertisement: **Berlin 2 Open Access Meeting**, 12-13 May, at CERN. Theme: implementation of Berlin Declaration (2003).

<http://www.zim.mpg.de/openaccess-cern/index.html>