

The Planetary Project: Towards eMath3.0

Michael Kohlhase – for the PLANETARY Group

Computer Science, Jacobs University Bremen, Germany
<http://planetary.mathweb.org>

Abstract. The PLANETARY project develops a general framework – the PLANETARY system – for social semantic portals that support users in interacting with STEM (Science/Technology/Engineering/Mathematics) documents. Developed from an initial attempt to replace the aging portal of PlanetMath.org with a mashup of existing MKM technologies, the PLANETARY system is now in a state, where it can serve as a basis for various eMath3.0 portals, ranging from eLearning systems over scientific archives to semantic help systems.

The PLANETARY project aims at developing a general framework – the PLANETARY system – for social semantic portals that support users in interacting with STEM documents. It is carried by enthusiasts from Jacobs University and The Open University.

Main Concepts and Project Genesis

Work on the PLANETARY system was triggered in August 2010 by the realization that the KWARC group at Jacobs University had developed semantic counterparts of much of the components underlying the PlanetMath portal [Pla]. PlanetMath.org is an online community that creates and manages an encyclopedia of mathematical concepts; hundreds of regular contributors have published about 8500 encyclopedia entries called articles. PlanetMath was founded in 2000; even before Wikipedia, and is thus one of the first Web2.0 systems. The Noösphere system [Noo] underlying the portal – essentially a \LaTeX -based Wiki implemented in Perl – is showing its age and becoming hard to manage. We felt that extending Planetmath to an eMath3.0 system – a social semantic web platform for Mathematics – via MKM technologies, might breathe additional life into the PlanetMath community and at the same time serve as a showcase of MKM technologies into the mathematics community.

The pre-existing MKM components that can be combined to form a semantic counterpart of Noösphere are (see also Figure 1):

1. TNTBase [ZK09; Tnt] for web-enabled, versioned storage
2. The LaTeXXML daemon [GSK11; Mil] for transforming \TeX / \LaTeX document fragments to HTML¹

¹ We use HTML5 as it integrates HTML for document layout with MathML for formula presentation, SVG for diagrams, and RDFa for document-embedded metadata and is supported by the major browsers.

3. s \TeX [Koh08; sTeX], a semantic variant of \LaTeX that can be transformed to OMDoc [Koh06] and further to semantically annotated HTML5 [KMR08]
4. JOBAD [GLR09; JOBAD], a JavaScript API embedding semantic services into Web documents

The only missing piece was a front-end that integrated them, added user and permissions management, and added discussion fora (an essential feature of PlanetMath). We found this component in the open source Vanilla Forums system [Van], that could easily be extended by wrapping the MKM components into Vanilla plugins. Another plugin that had to be added to the mix for PlanetMath feature parity was a system for metadata management and visualization: PLANETARY exports metadata to an RDF triple store (here an instance of the Openlink Virtuoso system [Olv]) and integrates custom SPARQL queries into the user interface e.g. to allow access to PlanetMath articles via the MSC2010 classification [Msc].

Already the proof-of-concept implementation in Fall 2010 made it clear that this combination of MKM technologies could be much more useful than for just re-implementing PlanetMath.

Framework for Semantic Publishing & Active Documents

The PLANETARY system has been generalized into a comprehensive eMath3.0 framework for semantic publishing and knowledge management, which has been instantiated prototypically in a variety of settings to validate the framework and support communities. The portals realized in the PLANETARY sys-

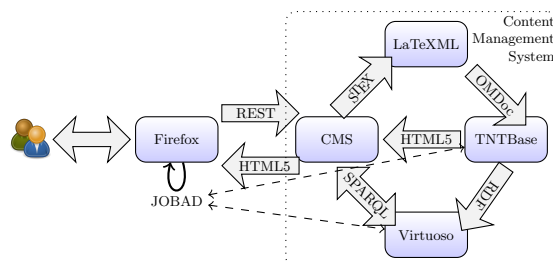


Fig. 1. Architecture of the PLANETARY system

tem range from eLearning systems over scientific archives to semantic help systems. All share the common basic architecture (see Figure 1), which integrates the components discussed above as web services into a central *Container Management System* (CMS) that mediates all user interaction. Note that in this MKM-centric architecture, we greatly extend the role of the *content management subsystem* (denoted by the dotted box in Figure 1). The CMS (initially Vanilla Forums, later Drupal) supplies management and interaction at the “container level”, i.e., without ever looking into the documents it manages (hence the somewhat non-standard name). The management of *structured document content* is split between TNTBase and the RDF triple store in PLANETARY, since they can perform semantic services.

Note that the level of semantic interaction afforded by the PLANETARY system depends on the depth of semantic annotations in the documents, and thus on different instances of the PLANETARY system: They range from simple folding

and localized commenting services in a front-end system for the Cornell ePrint arXiv to a front-end system which features in-place type reconstruction and elision of arguments and brackets for the fully formal LATIN atlas [Cod+11]. In June 2011, the PLANETARY system became one of the finalist systems in the Elsevier Executable Papers Challenge [Koh+11]. But the development push to reach this milestone also revealed crucial shortcomings of the CMS at the heart of the PLANETARY system, and the system was ported to Drupal whose container model and editing facilities are more suitable. Unfortunately, work on the port, on improving the subsystems, and data conversion issues have delayed any deployment of production systems based on PLANETARY (e.g. PlanetMath).

Current and Future Work

Currently, the work in the PLANETARY project focuses on four PLANETARY-based systems:

- finishing a production-ready PLANETARY instance of PlanetMath,
- developing a Web2.0 frontend with lightweight semantic features for `http://arxiv.org`, an archive of over 700 000 scientific documents. Particular care will be placed on extracting functional semantics from give L^AT_EX documents and using this in formula search,
- re-establishing the separate compilation and linking functionality for modular semantic publishing (see [Dav+11b]) in the eLearning3.0 System Pan-taRhei used in teaching CS courses at Jacobs University, and
- integrating PLANETARY as a knowledge provider in semantic allies; see [Dav+12].

We hope to be ready for production deployments of at least two of these at the time of CICM 2012.

References

- [Cod+11] Mihai Codescu et al. “Project Abstract: Logic Atlas and Integrator (LATIN)”. In: *Intelligent Computer Mathematics*. Ed. by James Davenport et al. LNAI 6824. Springer Verlag, 2011, pp. 289–291. ISBN: 978-3-642-22672-4.
- [Dav+11a] James Davenport et al., eds. *Intelligent Computer Mathematics*. LNAI 6824. Springer Verlag, 2011. ISBN: 978-3-642-22672-4.
- [Dav+11b] Catalin David et al. “A Framework for Modular Semantic Publishing with Separate Compilation and Dynamic Linking”. In: *1st Workshop on Semantic Publication (SePublica)*. (Hersonissos, Crete, Greece, May 30, 2011). Ed. by Alexander García Castro et al. CEUR Workshop Proceedings 721. Aachen, 2011. URL: <https://svn.mathweb.org/repos/planetary/doc/sepublica11/paper.pdf>.
- [Dav+12] Catalin David et al. “SAlly: A Framework for Universal Semantic Allies”. submitted to CICM 2012. 2012. URL: <http://kwarc.info/kohlhase/submit/mkm12-SAlly.pdf>.

- [GLR09] Jana Giceva, Christoph Lange, and Florian Rabe. “Integrating Web Services into Active Mathematical Documents”. In: *MKM/Calculus Proceedings*. Ed. by Jacques Carette et al. LNAI 5625. Springer Verlag, July 2009, pp. 279–293. ISBN: 978-3-642-02613-3. URL: <https://svn.omdoc.org/repos/jomdoc/doc/pubs/mkm09/jobad/jobad-server.pdf>.
- [GSK11] Deyan Ginev, Heinrich Stamerjohanns, and Michael Kohlhase. “The L^AT_EX_{ML} Daemon: Editable Math on the Collaborative Web”. In: *Intelligent Computer Mathematics*. Ed. by James Davenport et al. LNAI 6824. Springer Verlag, 2011, pp. 292–294. ISBN: 978-3-642-22672-4. URL: <https://svn.kwarc.info/repos/arXMLiv/doc/cicm-systems11/paper.pdf>.
- [JOBAD] *JOBAD Framework – JavaScript API for OMDoc-based active documents*. URL: <http://jobad.omdoc.org> (visited on 02/18/2012).
- [KMR08] Michael Kohlhase, Christine Müller, and Florian Rabe. “Notations for Living Mathematical Documents”. In: *Intelligent Computer Mathematics*. 9th International Conference, AISC, 15th Symposium, Calculus, 7th International Conference MKM (Birmingham, UK, July 28–Aug. 1, 2008). Ed. by Serge Autexier et al. LNAI 5144. Springer Verlag, 2008, pp. 504–519. URL: <http://omdoc.org/pubs/mkm08-notations.pdf>.
- [Koh+11] Michael Kohlhase et al. “The Planetary System: Web 3.0 & Active Documents for STEM”. In: *Procedia Computer Science* 4 (2011): *Special issue: Proceedings of the International Conference on Computational Science (ICCS)*. Ed. by Mitsuhiro Sato et al. Finalist at the Executable Papers Challenge, pp. 598–607. DOI: 10.1016/j.procs.2011.04.063. URL: <https://svn.mathweb.org/repos/planetary/doc/epc11/paper.pdf>.
- [Koh06] Michael Kohlhase. *OMDOC – An open markup format for mathematical documents [Version 1.2]*. LNAI 4180. Springer Verlag, Aug. 2006. URL: <http://omdoc.org/pubs/omdoc1.2.pdf>.
- [Koh08] Michael Kohlhase. “Using L^AT_EX as a Semantic Markup Format”. In: *Mathematics in Computer Science* 2.2 (2008), pp. 279–304. URL: <https://svn.kwarc.info/repos/stex/doc/mcs08/stex.pdf>.
- [Mil] Bruce Miller. *LaTeXML: A L^AT_EX to XML Converter*. URL: <http://dlmf.nist.gov/LaTeXML/> (visited on 09/08/2011).
- [Msc] *Mathematics Subject Classification MSC2010*. 2010. URL: <http://msc2010.org> (visited on 11/16/2011).
- [Noo] *Noosphere (diacrisis optional) is the software that underlies the Planet-Math Website*. URL: <http://code.google.com/p/noosphere/> (visited on 09/30/2010).
- [Olv] OpenLink Software. *OpenLink Universal Integration Middleware – Virtuoso Product Family*. URL: <http://virtuoso.openlinksw.com> (visited on 10/22/2009).
- [Pla] *PlanetMath.org – Math for the people, by the people*. URL: <http://planetmath.org> (visited on 09/08/2011).
- [sTeX] *Semantic Markup for L^AT_EX*. Project Homepage. URL: <http://trac.kwarc.info/sTeX/> (visited on 02/22/2011).
- [Tnt] *TNTBase TRAC*. URL: <http://tntbase.org> (visited on 12/16/2009).
- [Van] *Vanilla Forums*. URL: <http://vanillaforums.org> (visited on 09/22/2010).
- [ZK09] Vyacheslav Zholudev and Michael Kohlhase. “TNTBase: a Versioned Storage for XML”. In: *Proceedings of Balisage: The Markup Conference*. Vol. 3. Balisage Series on Markup Technologies. Mulberry Technologies, Inc., 2009. DOI: 10.4242/BalisageVol3.Zholudev01.