Towards A Community of Practice Toolkit Based On Semantically Marked Up Artifacts

Christine Müller and Michael Kohlhase

Computer Science, Jacobs University Bremen
{c.mueller,m.kohlhase}@jacobs-university.de

Abstract. Almost all aspects of scientific research and communication are now supported by software systems. Even though most of these systems allow the user to specify interaction preferences or even employ user modeling techniques, every system is an island with this respect. In particular, different systems cannot share user models or predict in the absence of prior interactions. We use ideas from the theory of Communities of Practice to arrive at declarative models that predict prior knowledge, preferences, and learning paths of scientists. Our models build on collections of semantically marked up artifacts (CoPfolios) that inscribe scientific practice. Our vision is to provide a toolkit that encapsulates CoPfolios allowing scientific applications to share user data.

1 Communities of Practice in Science

In the late 80s [LW91] coined the term *Communities of Practice (CoP)* to express the need for a new theory of learning. Nowadays, the concept is a well-known and widely accepted theory, which has a great impact on various disciplines: Meant to be useful for the debate on education, the concept has been applied to domains such as government, science, education, as well as industry and is of interest to both, researchers and practitioners.

[KW05] analyzed CoPs in the domain of science. We agree that the CoP theory is suited to describe scientific communities. In particular, the concept allows us to better understand the collaborative and emergent nature of science. Moreover, as CoPs act as "platforms for building a reputation" [Wen05], they support an important concern in science, i.e. to increase one's reputation and impact on the community.

We apply the theory of CoPs to the *Science, Technology, Engineering, and Mathematics (STEM)* disciplines. We view *STEMicians* as mathematical practitioners, who understand mathematics as the *language of science* and as the *basis for several disciplines*. The STEM community is very heterogeneous: Although outsiders may get the impression that mathematical practitioners form a homogeneous, unified community and share the same practices all over the world, they actually form various sub-communities that differ in their *preferred notations* [Caj93,SW06,Mül08b], *basic assumptions* [Rab08], and *motivating examples* [KK06].

We observe that scientists *primarily interact via their artifacts*, including documents in a more traditional understanding such as conference proceedings, journal publications, and books as well as documents in a wider interpretation such as emails, forum postings, online reviews, and wiki entries. Artifacts also include mathematical concepts and foundations as well as software and libraries. We assume that scientific interactions, and more generally mathematical practices, are *inscribed* into artifacts and aim at extracting the *inscribed scientific practices* to *model* scientific communities and their participation. [KW05] follow a similar approach to adapt the principles for cultivating CoPs by [WMS02] for the *Computer Supported Collaborative Learning (CSCL)* community. Based on a *citation analysis* of the CSCL conference proceedings they model the participation of their scientific community. In contrast, we build on a *semantically marked up corpus* of artifacts, which facilitates to distinguish different types of artifacts and relations on distinct granular levels. Moreover, [KW05] compose design principles by analyzing and *describing the current situation* in CSCL, while we aim at modeling the STEM community in order to *predict* membership, interest, activities, and preferences.

2 Community Oriented Technologies for Science

[Wen01] emphasizes that the success of a CoP primarily depends on social, cultural, and organizational factors and secondly on technological features. Taking this into account, [Wen01] and [WWSR05] provide aspects of CoPs that can be supported by technology and point out implementations that are suited to support these areas. We have applied the discussion to the STEM community, in particular, to analyse tools that support interaction via scientific artifacts. From our survey, we conclude that STEMicians have been using technology for many years. Many specialized tools exist that support various scientific activities, but none of them supports all. We observed that in science an *all-embracing implementation seems impossible* since the requirements of scientists are very *diverse* and even *contrary*. In particular, the choice of tools often depends on the scientist's basic assumptions and foundations, which in turn depend on his "personal preferences and the character of the current problem" [Rab08]. Since scientists use various tools to accomplish different tasks, their repertoire of artifacts is scattered across various system-internal database and so is the repertoire of their CoPs. However, we need to consolidate artifacts and *integrate existing scientific tools* to facilitate scientists to *manage and share their data across systems* and, thus, to facilitate collaborations. Since the STEM community is very *document-centered*, we propose an integrative approach that pools scientific tools with various specialized functionality around *common corpora of artifacts* in a *common representation and markup format.* For the course of this paper we focus on systems that build on our XML-based Open Mathematical Document Format (OMDoc)[1] [Koh06].

[1] Please note that we are not restricted to the OMDoc format but emphasize on the OMDoc functionality. This approach can be applied to other powerful markup formats, such as sTEX [Koh05b] or CNXML [HG07].

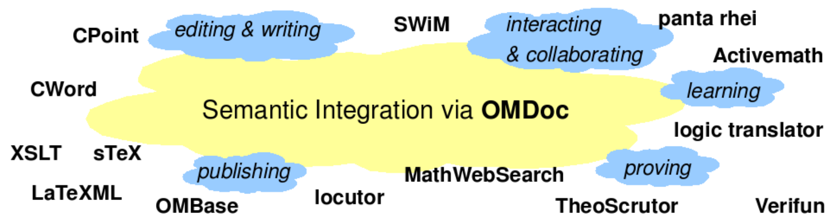

Fig. 1. The OMDoc universe: Integrating scientific tools via a common representation format.

We view OMDoc as a semantic integration platform, which offers means to annotate the *structural semantics* of artifacts, that is to "identify the structure, the meaning of text fragments, and their relations to other knowledge" [Koh06]. This markup facilitates the automatic processing of scientific documents, while allowing authors to choose the level of formality of their documents [Mül08c].

Figure 1 illustrates scientific tools that have already been integrated via the OMDoc format and which support various scientific activities[2]. In particular, the OMDoc universe comprises the intelligent version database [OMB08], the management of chance system [loc07], the semantic wiki [Lan08], the discussion platform [pan08], the course management system [Act07], the automated prover [Ver08], the semantic search engine [Mat07], the theory-browser [NK07], the logic translator by [Rab08], and the invasive OMDoc editors in Microsoft PowerPoint and Word [Koh05a]. However, none of these systems provide an integration of user data and preference settings.

In the following sections, we build on the notion of *portfolios*, which integrate artifacts from several systems wrt. to a single scientist, in particular, his user data and preference settings. Based on these *single-owned portfolios*, we propose **CoPfolios**, which include *artifact collections of groups* and facilitate the portability of common preference settings across various systems (cf. Sect. 3). Moreover, portfolios and CoPfolios provide portable and (partially) public community and user views that help to contextualize the adaptation of information. In contrast to many other adaptive systems, we focus on semantically enriched artifacts. We believe that *semantic technologies* support the *reification* and *extraction* of *scientific practice* and discuss novel services based on the extracted information (cf. Sect. 4 and 5).

3 Towards Portable User and CoP Models

3.1 From Portfolios to CoPfolios

The term "*portfolio*" (lat. portare "carry" and folium "folio") refers to a collection of objects of a specific type that document an individual's activities and

[2] The figure is patterned after an overview in [WWSR05] on general CoP oriented technologies. Please note that we do not cover all relevant scientific tools in this paper, but focus on the OMDoc universe.

preferences [Wik08]. They include *artifacts* the individual created, as well as *self-evaluation*, *reflection*, *best practice*, and *methods*. Portfolios consists of private as well as public parts that allow others to get an impression of one's practices. Portfolios have been used in several domains, in particular, in *industry* (finance or project portfolios), *education* (learner portfolios), and *human resource management* (career portfolios) to support *life-long learning* and to *track personal or community developments*. We view portfolios as platform-independent collections of *interrelated semi-structured* artifacts an individual creates, shares, discusses, or uses as well as *metadata* on these artifacts that express the individual's self-reflexion, interest, and views.

Building on the portfolio idea of *single-owned* collections of artifacts, we propose CoPfolios, which are artifact collections shared by a CoP of scientists and are interpreted as *representations of CoPs*. Note, as portfolios are special cases of CoPfolios for *singleton CoPs*, we only refer to CoPfolios in the further course of this paper.

3.2 Maintenance of CoPfolios

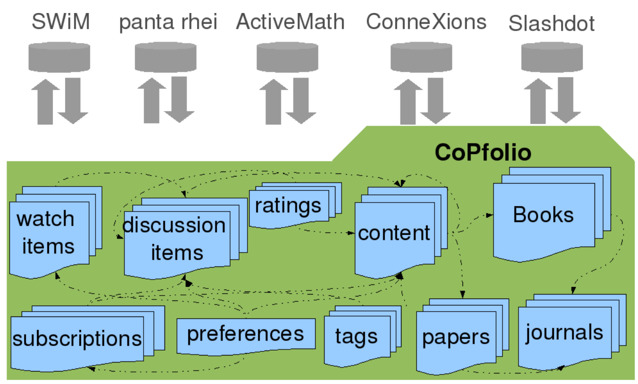

Fig. 2. Potential types of artifacts for CoPfolios. Several systems from the OMDoc universe, such as SWiM, *panta rhei*, and ActiveMath, as well as non-OMDoc systems, such as [CNX08] or [sla08], can initialize these types.

We propose `CoPit`[3], a community of practice toolkit, for managing CoPfolios. As illustrated in Fig. 2, CoPfolios include several *interrelated* types of artifacts such as *papers*, *preprints*, *discussion items*, *preference settings*, *ratings*, *tags*, or *subscriptions*. Moreover, they include profile data, such as email, fullname, address, and messenger IDs, as well as preference settings, e.g. with respect to general subscription preferences or notation systems. These types of artifacts are initialized by system-specific data provided by various systems, whereas the CoPfolios are maintained by `CoPit`. Each system provides an export to and import from `CoPit`. For more information on `CoPit` see [Mül08a].

[3] Please note that the following illustrations are visionary and experimental. We are in the progress of implementing a proof-of-concept prototype, with which we will evaluate and refine our approach.

4 Services based on CoPfolios

The immediate benefit of CoPfolios for individual scientists is the consolidation of their user data across systems. Accordingly, the immediate advantage of CoPfolios for CoPs of scientists is the platform-independent management of a CoP's repertoire. The machine-processable collection of semantically marked up artifacts also facilitate *CoP services*, i.e. services based on the sharing of information with CoP members and the access to other scientists' repertoires.

4.1 Views and Lenses

The preference settings in CoPfolios can be interpreted by systems to provide *views* or *lenses* on artifact collections, particularly, wrt. their *presentation*, *structure*, and *selection* [Mül08d]. For example, a group of scientists may prefer a particular *notation system*. Their *notation preferences* (or notation lens) are stored in the group's CoPfolio and can be reused by other scientists to view artifacts through the *CoP's lens*. Vice versa, instead of learning the CoP's notation system, a scientist could also use his *own lens* to make sense of the repertoire in the CoPfolio.

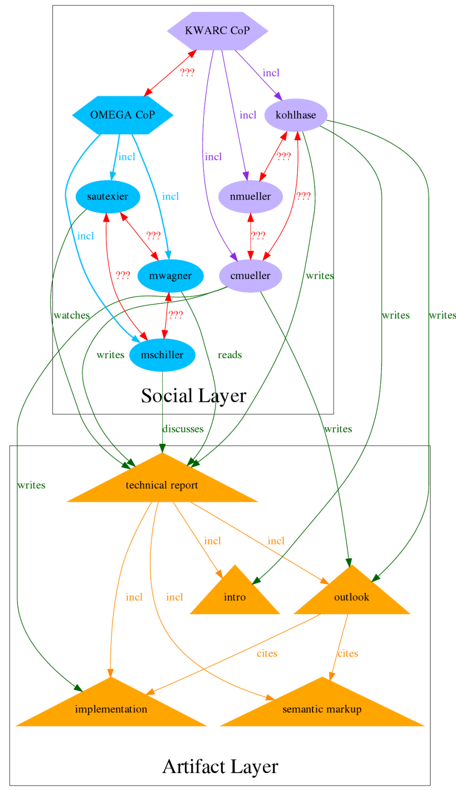

Fig. 3. Artifact & Social layer

4.2 Visualizing and Browsing

We can visualize the semantic interrelations within and across CoPfolios to e.g. facilitate the search for related work and potential collaborations across systems. Based on an interactive visualization, we can facilitate the semantic browsing along these relations extending existing approaches such as citation-based browsings.

Fig. 3 illustrates semantic interrelations on the *artifact layer*, i.e. relations such as *includes*, *cites*, and *refutes*, and the *social layer*, i.e. relations such as *knows*, *trusts*, or *collaborates with*. The relations between social layer and artifact layer, such as *writes*, *reads*, *implements*, *watches*, or *subscribes to*, allow the *bidirectional propagation of dependencies* between the layers, that is *bottom up* and *top down*: By analysing the content, structure, presentation, and metadata of an artifact, we can identify similarities that eventually propagate to the social layer: For example, the artifacts interrelation can be used to *construct social networks* or to *predict similarities between users*. In contrast, a top down approach allows to use the information on users and their social relation to *define* the adaptation, selection, and structuring of artifacts [Mül08c].

4.3 CoP Spawning and Browsing

We propose the interpretation of semantically marked up artifacts to compute the *differences* between user (or rather between the user's artifacts and their interrelations) to eventually build *parametrized clusters* of similar users, henceforth referred to as *virtual CoPs*. The parameters define different *dimensions* for the clustering, e.g. the common *basic assumptions* or *background*, the common *choice of examples*, or the common *notation preferences*. We do not claim that the computed clusters (or virtual CoPs), are CoPs wrt. to [LW91] as they only consider selected dimensions. However, they provide initial means for other users to cope with information without prior interactions.

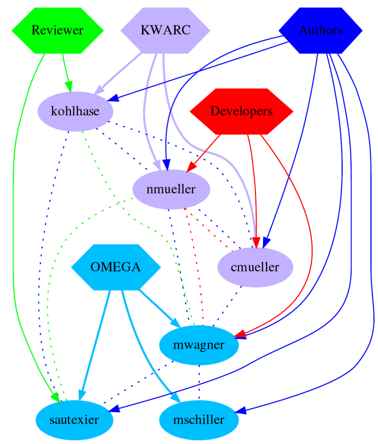

Fig. 4. Virtual CoPs

For example, based on the interrelated repertoire of the KWARC and ΩMEGA group in Fig. 4, subcommunities within and across both CoPs can be identified. Potentially, these *virtual CoPs* help to identify *similarities* and relations among the members of both CoPs, on which basis new CoPs (or collaborations) can be established. For example in Fig. 5, the virtual *Developer CoP*, with the members nmueller, mwagner, and cmueller, as well as the CoP's *implementation repertoire*, including mmlkit [MMK08], the `locutor` system, and PlatΩ [pla07], are displayed. The scientists nmueller and cmueller relate via their collaborative development of `mmlkit`. mwagner implements PlatΩ and watches the development of the `locutor` systems, which eventually allows to infer his relation to nmueller or the relation between his implementation and the `locutor` system.

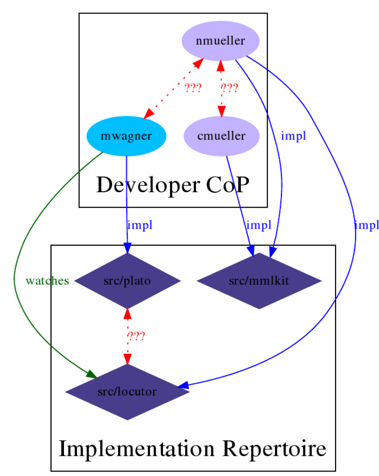

Fig. 5. Developer CoP and its repertoire.

5 `CoPit` Case Study

As `CoPit` supports the sharing of user-specific data across systems and among other users, we need to take the authentication and rights management into account. We will use the *decentralised identity service* [Ope08] via URL-based identities to uniquely identify scientists across systems. Systems can adapt their authentication process to these open identities or rely on their own user management.

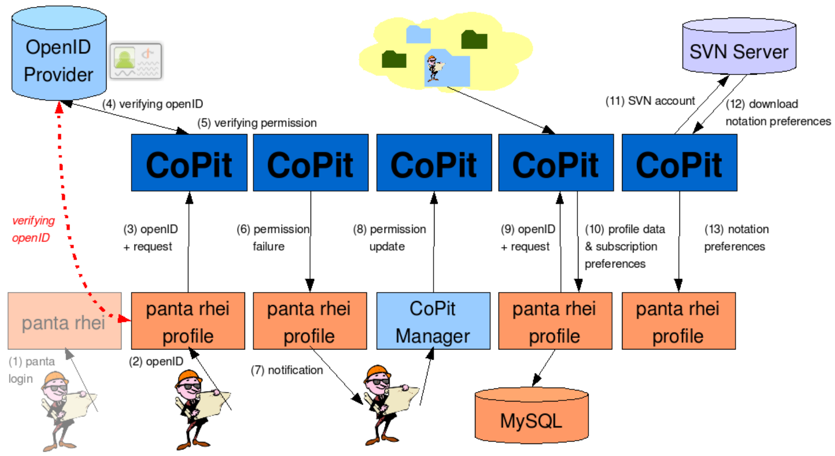

Fig. 6. Authentication and Rights Management based on OpenIDs

Figure 6 illustrates a potential OpenID-based scenario. The user *kohlhase* logs into the panta rhei system (for the first time) using his panta rhei account (1)[4]. He access his profile page in the system, provides his openID, e.g. `http://copit.kwarc.info/kohlhase`, and assigns panta rhei to display his profile data from his `CoPit` CoPfolio (2). panta rhei prompts `CoPit`, provides the openID, and request the profile data of kohlhase. `CoPit` prompts an OpenID Provider, e.g. `http://copit.kwarc.info/`, to verify the openID (4). It then verifies the access rights for the panta rhei system on the kohlhase CoPfolio (5). Since panta rhei has no access, the request fails (6). Panta rhei prompts the user to grant the required permissions (7). kohlhase uses the `CoPit` manager to grant panta rhei read access on his profile and preference data (8). panta rhei re-initializes its request (9). This time the system has access to the respective parts in the kohlhase CoPfolio, receives profile data, and initializes its internal profile page (10). Since panta rhei provides email notifications, the system can interpret the subscription preferences in the user's CoPfolio (10). Being a notation aware system, panta rhei can also interpret the user's notation preferences. However, the notation preferences are stored in the kohlhase SVN repository. Consequently, panta rhei only receives pointers and needs to prompt the SVN server for the actual files. However, the system is not permitted to access the respective SVN directories yet. It thus needs to notify the user to modify the respective SVN permissions. Alternatively, `CoPit` provides an abstraction to the storage of the notation preference files in SVN `CoPit` stores the SVN account of kohlhase, which is now used to access SVN (11), to download the requested files (12), and to finally pass them to panta rhei (13).

[4] Alternatively, panta rhei could directly be based on openID. Instead of verifying the system-internal user account the system would prompt an OpenID provider to verify the openID

6 Related Work

Our approach of pointer-based CoPfolios is analogous to the user modeling [Mel01] in the ActiveMath [Act07] system, an eLearning system based on an OMDoc corpus of course material, which provides the generation of user-specific courses and exercises. The generation is based on *learner models* consisting of concepts, technically pointers to the OMDoc repertoire, as well as competencies of the users. The model focuses on the representation of a learner's knowledge and preferences wrt. layout and notation preferences. CoP models or the portability of models across systems is not considered.

The emerging standard FOAF [FOA08] (an acronym for *friend-of-a-friend*) is an machine-readable ontology describing persons, their activities, and relations to other persons and objects. Users can use the RDF [LS99] extension to describe themselves and publish their user data. Several systems provide imports and exports of these FOAF files to initialize their user models. However, the FOAF format is restricted to individuals and does not support to represent all relevant user data such as specialized preference settings.

7 Conclusion & Outlook

An important aspect of the Knowledge Society is the challenge of digital knowledge management (KM). We believe that KM requires the digitalization of knowledge and its social context to provide sophisticated (automated) interactions. As interaction is practice, we make use of the theory of communities of practice to reify practices and represent interaction.

In our previous work we have discuss the reification of several practices as preference settings [KMR08,KMM07a,KMM07b]. In this paper, we focused on a natural place to collect, store, and manage these preferences to facilitate the sharing of user and CoP data. We made use of the notion of *portfolios* to consolidate artifacts and context (represented as preference settings) of scientists across systems (of the OMDoc universe) and extended portfolios towards collections of CoP artifacts, i.e. CoPfolios. These are maintained by `CoPit`, which facilitates the sharing of user and CoP data across systems. To motivate our approach, we presented several CoP services and a case study, which we want to implement in the *panta rhei* system. We plan to integrate functionality of the *locutor* system to extract semantic dependencies for the browsing of artifacts as well as computation of *virtual CoPs*. For the implementation of views and lenses, we will integrate the `mmlkit` [MMK08] system, which facilitates the adaptation of artifacts wrt. to different notation preference.

Acknowledgments We would like to thank the KWARC group for their valuable feedback and discussion on our work. Special thanks go to Normen Müller, Christoph Lange, and Andrea Kohlhase. This work was supported by JEM-Thematic-Network ECP-038208.

References

Act07. ActiveMath, seen February 2007. web page at `http://www.activemath.org/`.

Caj93. Florian Cajori. *A History of Mathematical Notations*. Courier Dover Publications, 1993. Originally published in 1929.

CNX08. Connexions. web page at `http://cnx.org`, seen June 2008.

FOA08. Friend of a Friend (FOAF) project. `http://www.foaf-project.org/`, seen June 2008.

HG07. Brent Hendricks and Adan Galvan. The Connexions Markup Language (CNXML). `http://cnx.org/aboutus/technology/cnxml/`, 2007. Seen June 2007.

KK06. Andrea Kohlhase and Michael Kohlhase. Communities of Practice in MKM: An Extensional Model. In Jon Borwein and William M. Farmer, editors, *Mathematical Knowledge Management, MKM'06*, number 4108 in LNAI, pages 179–193. Springer Verlag, 2006.

KMM07a. Michael Kohlhase, Achim Mahnke, and Christine Müller. Managing Variants in Document Content and Narrative Structures. pages 324–229, 2007.

KMM07b. Michael Kohlhase, Christine Müller, and Normen Müller. Documents with flexible notation contexts as interfaces to mathematical knowledge. In Paul Libbrecht, editor, *Mathematical User Interfaces Workshop 2007*, 2007.

KMR08. Michael Kohlhase, Christine Müller, and Florian Rabe. Notations for Living Mathematical Documents. In *Mathematical Knowledge Management, MKM'08*, LNAI. Springer Verlag, 2008. in press.

Koh05a. Andrea Kohlhase. Cpoint, 2005. http://kwarc.info/projects/CPoint/.

Koh05b. Michael Kohlhase. Semantic markup for TEX/LATEX. Manuscript, available at `http://kwarc.info/software/stex`, 2005.

Koh06. Michael Kohlhase. OMDoc – *An open markup format for mathematical documents [Version 1.2]*. Number 4180 in LNAI. Springer Verlag, 2006.

KW05. Andrea Kienle and Martin Wessner. Principles for Cultivating Scientific Communities of Practice. In Peter van den Besselaar, Giorgio de Michelis, Jenny Preece, and Carla Simone, editors, *Communities and Technologies*, pages 283–299. Springer Verlag, 2005.

Lan08. Christoph Lange. SWiM: A semantic wiki for mathematical knowledge management. web page at `http://kwarc.info/projects/swim/`, seen February 2008.

loc07. *locutor*: An Ontology-Driven Management of Change, seen June 2007. system homepage at `http://www.kwarc.info/projects/locutor/`.

LS99. Ora Lassila and Ralph R. Swick. Resource Description Framework (RDF) Model and Syntax Specification. W3C recommendation, World Wide Web Consortium (W3C), 1999. `http://www.w3.org/TR/1999/REC-rdf-syntax`.

LW91. Jean Lave and Etienne Wenger. *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives S.)*. Cambridge University Press, 1991.

LWA08. *Wissens- und Erfahrungsmanagement LWA (Lernen, Wissensentdeckung und Adaptivität) Conference Proceedings*, 2008.

Mat07. Math Web Search. web page at `http://kwarc.info/projects/mws/`, seen June 2007.

Mel01. Erica Melis. User model description. DFKI Report, DFKI, 2001.

MMK08. Normen Müller, Christine Müller, and Michael Kohlhase. The math markup language toolkit (mmlkit). `http://kwarc.info/projects/mmlkit`, seen June 2008.

Mül08a. Christine Müller, editor. *2nd Workshop on Scientific Communities of Practice (SCooP-2008)*, 2008.

Mül08b. Christine Müller. A Survey on Mathematical Notations, 2008. `http://kwarc.info/publications/papers/kw1_notationSurvey.pdf`.

Mül08c. Christine Müller. Towards CoPing with Information Overload. [LWA08]. in submission.

Mül08d. Christine Müller. Towards the Adaptation of Scientific Course Material powered by Community of Practice. [LWA08]. in submission.

NK07. Immanuel Normann and Michael Kohlhase. Extended formula normalization for ϵ-retrieval and sharing of mathematical knowledge. In Manuel Kauers, Manfred Kerber, Robert Miner, and Wolfgang Windsteiger, editors, *Towards Mechanized Mathematical Assistants. MKM/Calculemus 2007*, number 4573 in LNAI, pages 266–279. Springer Verlag, 2007.

OMB08. OMBase Project, seen April 2008. available at `http://kwarc.info/projects/ombase/`.

Ope08. OpenID: Shared Identity Service, seen June 2008.

pan08. The panta rhei Project. `http://kwarc.info/projects/panta-rhei/`, 2008.

pla07. PLATO: Interactive Mathematical Authoring, seen August 2007. system homepage at `http://www.ags.uni-sb.de/plato/bin/view.pl`.

Rab08. F. Rabe. *Representing Logics and Logic Translations*. PhD thesis, Jacobs University Bremen, 2008. To appear.

sla08. slashdot. web page at `http://slashdot.org/`, seen July 2008.

SW06. Elena Smirnova and Stephen M. Watt. Notation Selection in Mathematical Computing Environments. In *Proceddings Transgressive Computing 2006: A conference in honor of Jean Della Dora (TC 2006)*, pages 339–355, Granada, Spain, 2006.

Ver08. VeriFun: A verifier for functional programs, seen February 2008. system homepage at `http://www.verifun.de/`.

Wen01. Etienne Wenger. Supporting Communities of Practice: A Survey of Community-oriented Technologies, 2001. Report to the Council of CIOs of the US Federal Government; published at `http://www.ewenger.com/tech/`.

Wen05. Etienne Wenger. Communities of Practice in 21st-century organization, 2005.

Wik08. Wikipedia: Portfolio (aus Wikipedia, der freien Enzyklopädie). `http://de.wikipedia.org/w/index.php?title=Portfolio&oldid=47657869`, 2008.

WMS02. E. Wenger, R.A. McDermott, and W. Snyder. *Cultivating of Communities of Practice*. Harvard Business School Press, 2002.

WWSR05. Etienne Wenger, Nancy White, John D. Smith, and Kim Rowe. Technology for Communities, 2005.