The Planetary System: Web 3.0 Active Documents for STEM

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 4 (2011) 598–607

Procedia Computer Science

The Planetary System: Web 3.0 & Active Documents for STEM

Michael Kohlhase[a], Joseph Corneli[b], Catalin David[a], Deyan Ginev[a], Constantin Jucovschi[a], Andrea Kohlhase[a], Christoph Lange[a], Bogdan Matican[a], Stefan Mirea[a], Vyacheslav Zholudev[a]

[a] *Computer Science, Jacobs University Bremen, Germany*
[b] *Knowledge Media Institute, The Open University, UK*

Abstract

In this paper we present the Active Documents Paradigm (semantically annotated documents associated with a content commons that holds the corresponding background ontologies) and the Planetary system (as an active document player). We show that the current Planetary system gives a solid foundation and can be extended modularly to address most of the criteria of the Executable Papers Challenge.

Keywords: active documents, executable paper challenge, content commons, semantic annotation, ontologies

1. Introduction

The Executable Papers Challenge (EPChallenge) calls for a framework that allows papers to be *executable*, give access to *provenance* and *copyright/licensing* information, and support various forms of *validation*. We contend that these challenges can be realized based on *semantically annotated documents* together with semantic background ontologies, which we call the **content commons**. This information can then be used by user-visible, interactive services like program execution, computation, visualization, navigation, information aggregation and information retrieval. Finally a document player application can embed these services to make documents executable. We call this framework the **Active Documents Paradigm** (ADP), since documents can also actively adapt to user preferences and environment rather than only executing services upon user request. In this paper, we present the ADP with a focus on the Planetary system as a document player on three distinct annotation levels, "presentation structure", "semantic", and "formal".

The Planetary system (see [1] for an introduction) is a Web 3.0 system[1] for semantically annotated document collections in Science, Technology, Engineering and Mathematics (STEM). In our approach, *documents published in the* Planetary *system become flexible, adaptive interfaces to a content commons* of domain objects, context, and their relations. The system achieves this by providing embedded user assistance through an extended set of user interactions with technical documents, based on an extensible set of client- and server-side services that draw on explicit (and thus machine-understandable) representations in the content commons.

2. The Planetary System: Current State

Development on the Planetary system began in summer 2010 based on an established set of knowledge management components; we have tested it on a variety of paradigmatic document collections, which we will describe

[1] We adopt the nomenclature where Web 3.0 stands for extension of the Social Web with Semantic Web/Linked Open Data technologies.

Selection and/or peer-review under responsibility of Prof. Mitsuhisa Sato and Prof. Satoshi Matsuoka
doi:10.1016/j.procs.2011.04.063

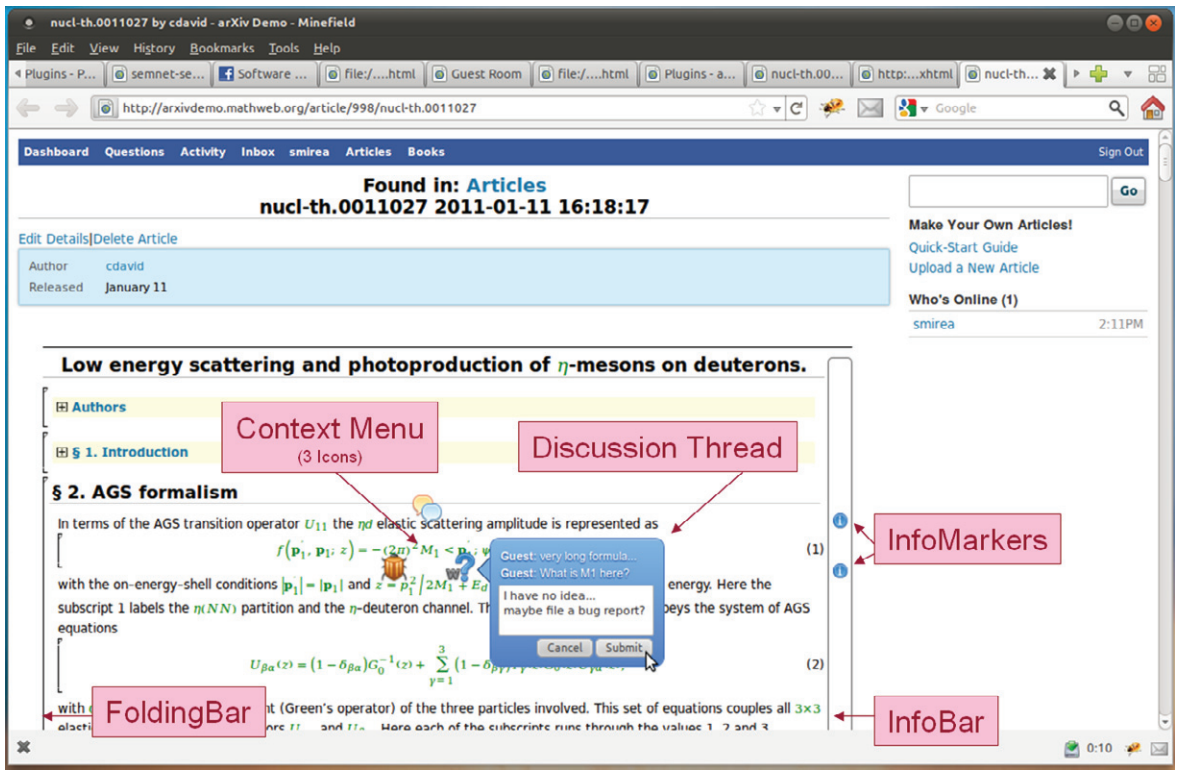


Figure 1: Interacting with an arXiv Article via `FoldingBar`, `InfoBar`, and Discussion Thread

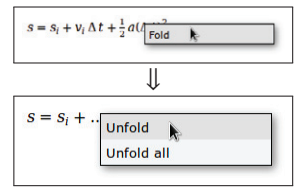
now. We group these by the level of semantic markup involved: generally, more semantic markup translates into more active documents, but requires a higher investment of the authors (but see Section 3.1). Note that PLANETARY gracefully adapts to the level of markup available: even though it was designed for the "semantic level" (see Section 2.2), it can also function well with a level of markup that can be extracted from pre-existing document formats, e.g. LaTeX (we call this the "presentation structure level"). The study of different levels of semantic saturation of individual documents is complemented by looking at libraries and collections of interlinked documents; Section 2.4 covers this aspect.

2.1. The Presentation Structure Level: PLANETARY *arXiv and PlanetMath*

The importance of the presentation structure level is that PLANETARY can turn legacy documents into active documents by transforming them into XHTML+MathML+SVG-encoded documents with semantic annotations in RDFa. We have transformed over half a million articles from the Cornell ePrint arXiv to XHTML+MathML with LaTeXML [2], preserving properties like document and formula structures and embedded them into an instance of the PLANETARY system [3].

The document structure can then be exploited for a `FoldingBar` service (see on the left in Figure 1) and for localizing discussions about document content to document structures and subformulae – e.g. for questions/answers, or reviewers' comments. In the situation in Figure 1 we have clicked on formula (1), which pops up the `IconMenu` with three options: reporting errors in the content (bug icon), asking/answering a question (question mark icon), and accessing the discussion threads of this element (balloons icon). Here, a click on the question mark icon allowed us to pose a question and hope for an anser by other users in the forum. Figure 1 also shows the PLANETARY `InfoBar` with information markers on the right, which indicate the availability and state of the discussion threads pertaining to information objects in the line they are horizontally aligned with. Clicking them will highlight all items that have discussions. Localized discussions have proven a very valuable tool for community-based validation of papers, especially if they are coupled with a discussion subscription/trackback system for readers and personal notification system for authors.

The formula structure allows us to provide semantic services like subformula folding. The image on the right shows a formula from [3] where we right-clicked on the formula $s = s_i + v_i\Delta t + \frac{1}{2}a(\Delta t)^2$ for the law of uniformly accelerated motion to get the PLANETARY context menu. Folding a subformula (i.e., replacing it by an ellipsis marker yielding $s = s_i + \ldots$) yields the lower (simplified) formula, which in turn can be unfolded again whenever wanted.



2.2. The Semantic Level: Semiformal Digital Libraries

We can considerably improve the user experience by extending the depth of semantic annotations. For this we employ OMDoc [4] (Open Mathematical Documents), an XML-based content-oriented, semiformal representation format for scientific and technical documents. It builds on the OpenMath/MathML3 semantic representation format for mathematical formulae [5, 6]. OMDoc extends OpenMath with an infrastructure for context and domain models from Formal Methods, as well as a generic document infrastructure. At the semantic level PLANETARY is based on STEX documents [7], which can be transformed to OMDoc and via a user-adaptive and context-based presentation process further to XHTML+MathML+SVG+RDFa. The generated OMDoc documents are committed to an instance of the versioned XML database TNTBase [8] that indexes them by semantic functional criteria, and can then perform server-side semantic services via user-defined XQuery queries.

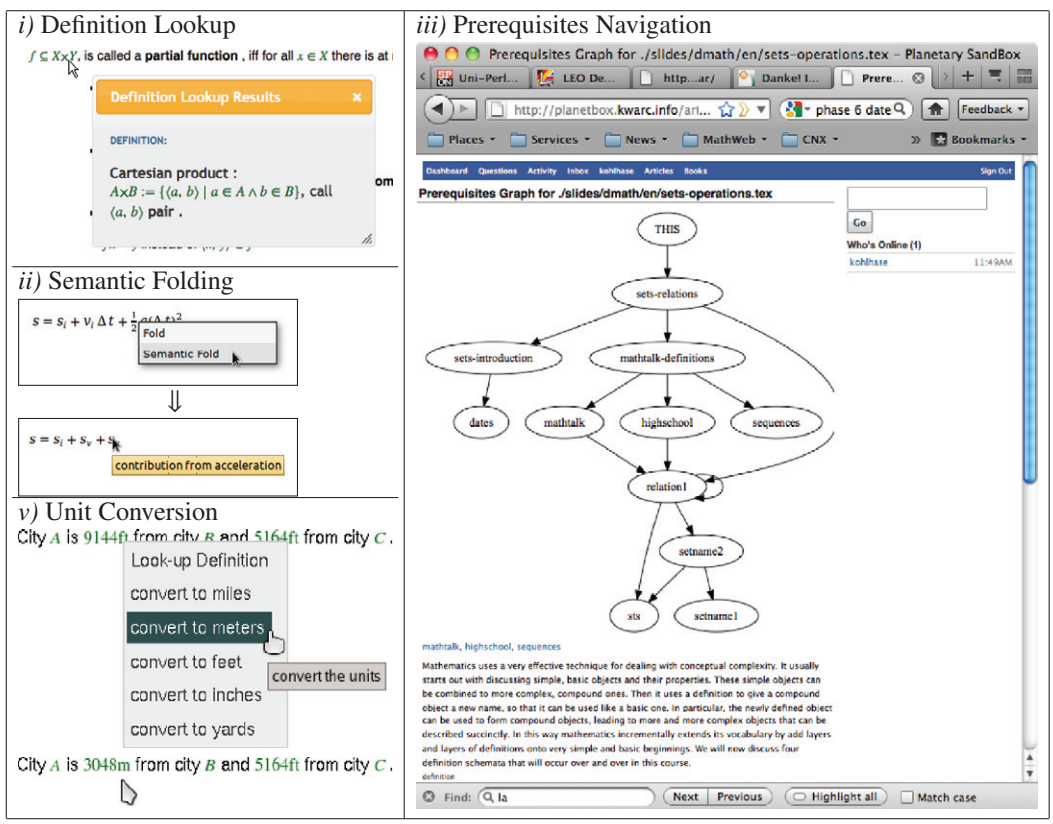
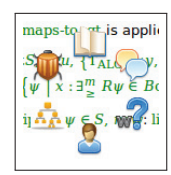


Figure 2: User Services at the Semantic Level

TNTBase thus becomes a source of the user-adaptive, custom-generated documents forming PLANETARY's content commons. Many semantic services can directly be derived from this setup. At the semantic level, the `IconMenu` we know from Figure 1 can be extended, depending on the services available for the semantic item in focus. The book icon triggers definition lookup (see item *i)* below) and the graph and portrait icons prerequisites navigation (see *iii)*). Figure 2 shows results of these and other services.



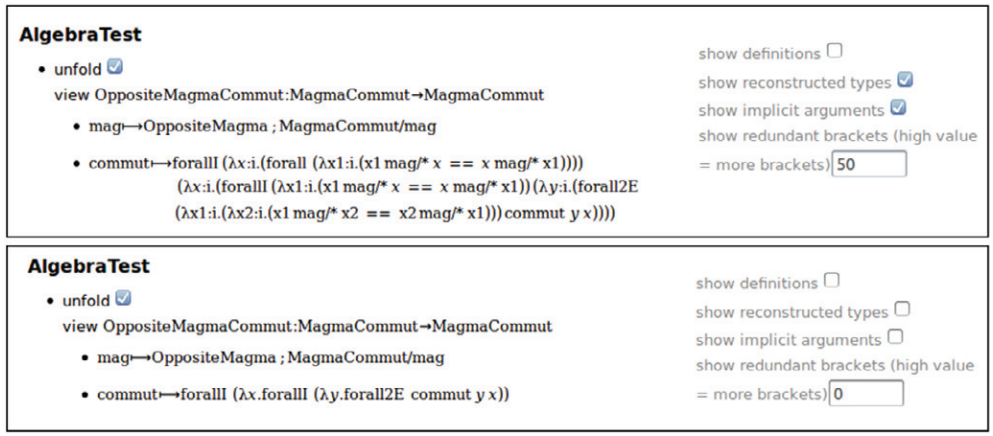


Figure 3: Formal Representations Adapted to Distinct User Settings (Customized via the `DocDash` Widget on the Right)

i) Definition Lookup: All technical terms and symbols in formulae presented in PLANETARY are linked to their semantic counterparts in the content commons, which in turn are linked to their definitions [9].

ii) Semantic Folding: If any explanations of the meanings of subformulae have been added as annotations, folding can use these instead of "...". In Figure 2/*ii)* the motion law above is semantically folded to $s_0 + s_v + s_a$ and the abbreviations s_* are explained via flyover help.

iii) Prerequisites Navigation: As the content commons has an inherent notion of semantic dependency, we can use that to show prerequisites leading to a concept. Currently PLANETARY supports two ways of dealing with prerequisites: *i)* a *concept graph view*, where the required concepts can be navigated on demand by clicking on concept nodes, and *ii) guided tours*, where the necessary content is generated in a coherent narrative [10].

iv) Executable Formulae: As the formulae in the documents are generated from OpenMath objects, we can export them to a computer algebra system like Mathematica (or in our case the open-source system GAP [11], which has been made available as an OpenMath-based Web service [12]) for evaluation, graphing, or experimentation.

v) Unit Conversion: In scientific papers, formulae often contain expressions for measurable quantities; these can be automatically converted to other unit systems [9].

Note that the underlying OMDoc format and the services based on it address a peculiarity of documents in the STEM disciplines. Here, documents are *dynamic* in the sense that they declare new concepts, definitions, model assumptions, terminology, notations, etc., as they go along, or else they explicitly (or implicitly) import such items from other documents. As a consequence, all knowledge items in STEM documents have a non-trivial context of declarations. This must be managed explicitly in our representations of these documents, in the content commons, and within user interaction in order for semantic services to be effective. This effect is especially pronounced in mathematical sciences (including Computer Science, Physics, etc.), and somewhat less so in Chemistry and the Life Sciences, where a global context in the form of external terminology and notation databases often suffices.

2.3. The Formal Level: Specification and Verification

Finally, we can use PLANETARY as a frontend system for completely formal content. OMDoc is also a foundation-agnostic integration format for mathematical knowledge that can express web ontologies [13, 14], program specifications and verifications [15], and even representations of logics in logical frameworks like LF [16]. In formal systems, documents are dominated by complex formulae, and users need support in navigating, abstracting, and evaluating them in order to cope with this complexity. Figure 3 shows how fully formalized formulae can be adapted to user preferences, about, e.g., the level of brackets and the availability of inferred arguments or definitions. Further semantic services at the fully formalized level include access to automated theorem provers via the HETS system [17] and argument reconstruction via the Twelf system [18].

It is important to note that programs (and program fragments) are also knowledge items at the fully formal level, as all their semantics can be recovered by parsing. Program fragments are not yet supported by PLANETARY (but see section 3.4).

2.4. The Network Layer: Library Integration of Document Metadata

We have concerned ourselves with individual active documents at various levels of semantic saturation so far. The PLANETARY system adds infrastructure for document collections and libraries as well (see Figure 4 and [19] for details). The library organizational level is realized as individual PLANETARY installations and provides the base URI of the system and its content. Examples include our PLANETARY-based scientific encyclopedia portal [20] or a course system. In these examples, the collection level corresponds to individual disciplinary encyclopedias such as PlanetMath [21] or individual courses in the course system. The contribution of the library and collection levels consists mostly in an inter-document, or network, layer of semantics via the aggregation of document metadata annotations. We represent metadata and cross-links between documents and their fragments as RDF graphs, reusing existing metadata vocabularies/ontologies[2] wherever appropriate. PLANETARY makes metadata and links editable to authors via LaTeX convenience macros, whereas some administrative metadata (e.g. the date of last modification) are recorded automatically. For services that interact with an active document, we employ *semantics-preserving transformations* that embed the metadata and links as user-invisible but machine-readable RDFa annotations.

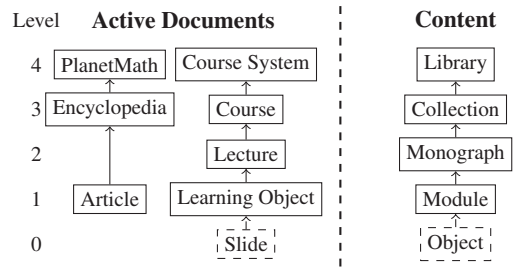


Figure 4: Levels of Organisation of Content

We keep metadata separate from the analysis of document semantics above to emphasize its independence of the document content, going one meta level up. This is also the justification behind its primary role as a connector (or least common denominator) between collections or entire libraries of documents, as opposed to the intra-document aspects outlined previously.[3] To enable this functionality, we add an RDF triple store as a system component and integrate it via its SPARQL endpoint API.

The utility of such annotations comes from (i) their reintegration into access methods or identifiers to collection and library subsets, and from (ii) the way PLANETARY makes the *vocabularies* of these annotations customizable and extensible within the system. Regarding (i), a generic mechanism of a "view" on a collection of documents can be instantiated to both simple and complex realizations, such as viewing the encyclopedia of PlanetMath alphabetically, by the Math Subject Classification (MSC [23]) categories of its articles or by posing the query "top rated new articles on Abelian groups, which were revised in the last three months (i.e. since I last read about such groups)". By making such views accessible by their own URLs, PLANETARY leverages the power of the resource locators in order to point to dynamic pages, which resolve an encoded metadata query. This new use of URLs generalizes the classic idea of what a resource is, in order to provide a more expressive way to communicate for users, for example in discussions about refactoring content, citing prior work, or pinpointing the next steps of a future effort. Regarding (ii), PLANETARY treats the annotation vocabularies used for metadata and links – their informal description, their formalization, and their documentation – exactly like any other of its STEM documents. In previous work (cf. [13, 24]), we have demonstrated that the OMDoc and STEX languages subsume common semantic web ontology languages, such as OWL, and go beyond them in their support for logical heterogeneity (cf. Section 2.3), integrated documentation up to literate programming (cf. Section 3.4), and modularity and reuse. By making the vocabularies editable inside PLANETARY, we enable the community to customize them. This is particularly powerful when a vocabulary immediately defines the functionality of a *service*, as is the case with our semantic discussion and refereeing facility explained in Section 3.3. Users (with sufficient permissions) can thus enhance the functionality of the service by introducing new vocabulary, they can refine the functionality by extending the formalization of existing vocabulary, and they can enable better user assistance by elaborating the documentation of those vocabulary terms that system functionality is based on. In that sense, *the ontology is the API* of the PLANETARY system.

We thus have the potential to envision metadata-enhanced social-semantic features such as subscribing to custom news feeds, social bookmarking, interoperability with external libraries and more.

[2] Here, we use the terms "vocabularies" and "ontologies" synonymously. Ontologies are the usual means of formalizing the semantics of vocabularies on the Semantic Web.

[3] Note that this separation is not total as there are interactions between the metadata level and the document semantics level. For instance, following a practice from the PlanetMath encyclopedia we use the NNexus tool [22] to semi-automatically discover concepts within documents (a document feature) and add links to their defining encyclopedia articles (a network feature).

2.5. The Active Documents Paradigm and the EPChallenge

We see the main impact of the Active Documents Paradigm and the Planetary system to the EPChallenge concentrated on the semantic level described in Section 2.2. At the presentation structure level the embeddable semantic services are largely limited by the semantic information that can be recovered from the LaTeX sources[4]. In the future some semantic annotations can be automatically discovered by heuristic or linguistic analysis methods [25], enabling services from the semantic level to be deployed on legacy documents. These analysis methods are subject to active research in our group, but beyond the scope of this paper.

Importantly, the Planetary system supports the incremental semantic upgrade of documents: any semantic annotation is directly made available to the semantic services, making documents more interactive, and thus contributes to the user experience of the enhanced active documents. This constitutes an opportunity for crowd-sourcing the semantic annotation of documents: Communities may propose annotations to documents which the document owner can then accept or reject. We can also think of alternative methods for quality control of semantic annotation, e.g. reputation/karma-based ones. A prerequisite for this will be tools that allow other users to express the semantic annotations, collect them, and incorporate them into the Planetary content modules as if they were author-supplied.

3. The Planetary EPChallenge Version

In this section we sketch how the Planetary system can be extended to answer most of the EPChallenge criteria before ICCS in June 2011. Details can be found on [26].

3.1. Semantic Editing Support

One of our priorities is to make the authoring process of semantic documents/discussions (to reach the semantic level) as easy as possible. We do that by providing a web-based authoring environment which lifts in-browser editing of sTeX documents to the level of popular LaTeX editors. Moreover, the environment is aware of sTeX semantics and can highlight fragments according to their functions (semantic syntax highlighting) as well as suggest autocompletions based on the context at request location. Crucially, the editor supports the creation of semantic links between documents, which is very tedious without this extension. For example, we implemented an instant search feature which helps the author in finding semantic concepts defined in other documents and subsequently reuse them without leaving the editing environment. Also, hovering over a concept shows its definition (if available), that is a definition lookup in the *source* document.

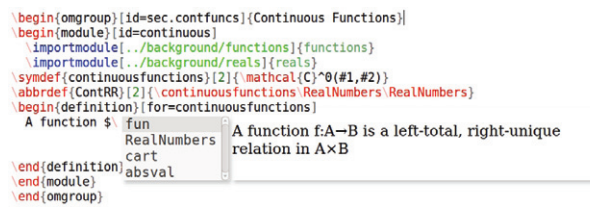


Figure 5: Autocompletion with Definition Suggestion

3.2. Integrated Semantic/Formula Search

Currently, the Planetary system only supports standard text-based search of the LaTeX/sTeX sources of articles and forum posts. We plan to integrate our currently independent semantic formula search engine MathWebSearch [27] for semantic content, using semantic copy-and-paste to alleviate/circumvent the remaining "math input problem". With such a formula search engine, a user can search mathematical content by its structure, e.g. the recollection that the energy of a signal $f(t)$ is given by the square of the signal integrated over time which can be expressed by a schema $\int_{\boxed{l}}^{\boxed{h}} \boxed{\mathrm{f}}(x)^2 dx$, where the boxed variables are search variables. This would find e.g. Parseval's Theorem $\frac{1}{T}\int_0^T s^2(t)dt = \sum_{k=-\infty}^{\infty} \|c_k\|^2$ (the search also took into account renaming the bound variable) which allows to compute the energy. Note that the definition lookup feature shown above comes in very handy here to look up the meaning of c_k: they are Fourier coefficients of s. The main problem to be solved here is to automatically infer the context of semantic elements in the documents. Formula search at the presentation structure level can be realized via heuristic transformations from presentation to content markup.

[4]Note that even though we have mainly studied documents that are automatically generated from LaTeX sources, initial experiments show that the Planetary system is equally applicable to other sources. In particular, the XML+MathML sources of Proceria Computer Science provided by Elsevier to the EPChallenge finalists could be integrated into the system after an afternoon's work.

3.3. Semantically Supported Discussions and Refereeing

We are semantically enhancing discussions with our previously developed argumentation ontology (which includes extensions to STEM-specific argumentative primitives) [28, 14]: Users can indicate categories for forum posts and fine-grained relations between forum posts and documents, or between posts and other posts. In particular, the argumentation ontology defines what types of *problems* apply to what types of STEM knowledge items (e.g. a declaration of a mathematical symbol can be ill-formed, but not invalid, as a proof could be), and what types of *solutions* apply to these problems (e.g. renaming a definition that a reader does not understand is less likely to solve the problem than providing an example for it).

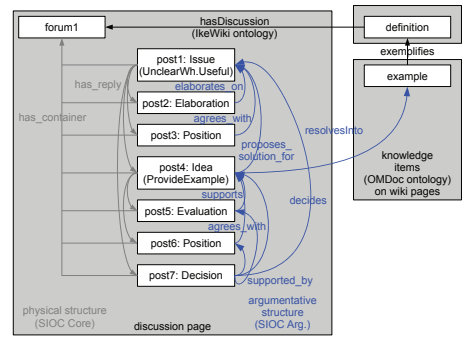


Figure 6: RDF Graph of a Discussion Thread about a Mathematical Definition (using our STEM-specific argumentation ontology)

The discussion facility is an application of our "ontology as API" approach introduced in Section 2.4, in that the discussion forum user interface uses such information from the ontology to guide users who want to write or reply to a post, and additionally reuses the documentation of the argumentation ontology as context-sensitive online help for the discussion forum user interface. As the argumentation ontology is editable inside PLANETARY, any community running a PLANETARY site can adapt it to their preferences, e.g. by defining new types of problems and solutions from their particular STEM subdomain, or refining the existing ones.

Adding versioned references [29] (facilitated by the TNTBase storage) allows incremental document/content commons development and commenting workflows to be managed in distributed teams. Building on the PLANETARY system's user management, we can extend this into a localized refereeing feature.

3.4. Program Fragments and Semantic Literate Programming

OMDoc and STEX have been used to deal with semiformal [15] and formal software engineering, but support for programming in the PLANETARY system is still lacking. The main problem to solve is that of modeling the scope of program fragments relative to background theory, e.g. in Computer Science papers. When that is solved, we will have a basis for "semantic literate programming", which can be supported by semantic services. To this end, we are currently working on integrating the notion of "biform theories" [30] into OMDoc for a more principled support of program fragments at the formal and semantic levels. Given this, we can make program fragments in a programming language $\mathcal{L}$ executable via an "universal OpenMath $\mathcal{L}$-Machine", a web service which accepts an $\mathcal{L}$-fragment f and arguments $a_1, \ldots, a_n$ encoded as OpenMath objects (i.e. in the form in which they occur in the documents that appear in the content commons), executes $f(a_1, \ldots, a_n)$, and returns the result as an OpenMath object. With this web service, program fragments will be executable in analogy to formulae (see the services discussed in Section 2.2).

3.5. Management of Change

Keeping document collections consistent in the face of changes induced by development is a grand challenge in the STEM disciplines. Even though published documents are considered immutable, changes do occur especially in the development of STEM knowledge. It is a long-standing assumption that the semantic annotations in documents that drive the machine-supported interaction can support automated change impact analyses at the same time. Generally, any change in a semantic object (say a definition) induces an impact on all objects that depend on it. In [31] we have taken first steps of integrating a change impact analysis into the editing and management workflow of the PLANETARY system by establishing a flexible interface that allows to deal with impacts. One limitation of the current integration that we want to alleviate in the near future is that our integration currently assumes a single-user mode of operation. The main conceptual problem to be solved here is how to deal with propagating changes by "other authors". For that we plan to build in the notion of versioned links proposed in [29] in addition to our TNTBase versioned storage.

3.6. Integration of Scientific Data and Provenance

So far we have been mainly concerned with infrastructures for abstract and theoretical sciences. Papers in the natural sciences need an infrastructure for dealing with *experimental and measurement data*. We have already dealt with foundational aspects of experiments in [32]. For dealing with data aspects at the semantic level, we plan to build on the SACHS system [33, 34], our semantic user assistance system for spreadsheets, which understands spreadsheets as active documents with respect to a background ontology in the content commons. We plan to integrate the "semantic illustration" technology [35] originally developed for SACHS into PLANETARY, directly implementing its user assistance [36] into PLANETARY functionality, and generalizing the data handling and provenance functionality for scientific data. We conjecture that this will yield an integrative enabling technology for transparent, semantic embedding of data. When we additionally combine this with the executable programs feature from 3.4, we can document the computation steps that aggregated the data reported in papers from their experimental foundations and open them to experimentation for the user.

3.7. PLANETARY *Pit Stop*

Finally, users need to manage interface complexity (not all services are interesting to all users) and want to contribute their own services. We will build an "app store" for semantic services that hides the complexities of installation, shadowing, and removal.

4. Conclusion

We have presented the PLANETARY system as an example for the Active Document Paradigm (ADP) – semantically annotated documents connected to a content commons accessed through an adaptive document player – and as an answer to the challenge of creating "executable papers". We have shown the initial feasibility of the concept in a variety of publicly available case studies [3, 21, 37, 38][5]. Details on the system are available in the PLANETARY developer forum [39]. In particular, see the EPCsystem page [26] for more information and [40] for a sandbox installation of the system.

The PLANETARY system addresses all of the challenge categories. Concretely:

Executability is achieved by an extensible and configurable collection of semantic services that can be embedded into documents. In particular, code can be made executable via external computation machines, and computational experiments can be repeated and varied (via the same mechanism). Where the functionality of a service depends on ontologies, the user community can customize the service by customizing the ontology inside PLANETARY.

Short and long-term compatibility is guaranteed by usage of open standards in representation formats and protocols (XML, RDF/RDFa, OpenMath, MathML, XQuery, SPARQL, XHTML, SVG) supporting a web service framework, and hence operating-system-independent. Of course, PLANETARY can export monographs, collections and even entire libraries both as PDF (*inactive* documents) or in the EPUB eBook format.

Validation and in particular human refereeing and scientific validation can be facilitated via an in-text discussion feature. Moreover, documents can be automatically validated via semantic services, e.g. automated SI-dimensionality checking.

Copyright/licensing is represented by fine-grained RDFa-based metadata annotations in STEX and OMDoc, which are maintained over the presentation process. So they can be used for filtering or attribution either on the backend storage level (TNTBase, RDF triple store) or in the frontend PLANETARY system. Together with the user management and permission system, PLANETARY can be extended to enforce compliance. In fact, as documents are assembled for the user at view-time they can be adapted to the license status of the user (e.g. it is possible to make a document license conforming by leaving out examples that are not licensed to her specific institution).

Systems As the PLANETARY system is entirely based on web standards and communicates via RESTful interfaces, it is simple to wrap external systems into web services, if we can equip them with OMDoc, OpenMath, or RDF interfaces. In fact, that is exactly what runs the unit conversion service in Section 2.2/*v)*.

[5] Note that all of these are research systems under constant development, so your mileage may vary.

Size Even though individual human-written documents are modest in size, journals and encyclopedias can get big – consider e.g. the Wikipedia or the arXiv. The PLANETARY system has been tested on the latter; the underlying data stores scale sufficiently for large document collections. Furthermore, the modular and semantic document formats accommodate the integration of external data stores via 'special' links, which the PLANETARY player can interpret on view, hence keeping the storage minimal and the experience optimal. To the best of our knowledge, the semantically transparent integration of data into a document player application is a new feature of the PLANETARY system.

Provenance comes in various aspects. Data provenance can be specified by the techniques for semantic integration of data fields. For instance we can specify units (as OpenMath objects) and computations to obtain the displayed data from raw data, etc. As all of these are content representations in the documents or the content commons, they can be handled with semantic services. The system state provenance (i.e., what actions of the user led to the current state of interaction in the PLANETARY system), can be handled by recording system data ("who did what when") in the metadata store of PLANETARY. This can be opened to querying the system ontologies we have developed for semantically transparent system self-documentation (see Section 2.4).

An encouraging aspect of this work is that document authors only need expertise in their own domain. In particular, no system-level programming is necessary for authors: the semantic representation formats involved act as a high-level conceptual interface between content authors and system/service/interface developers. We are convinced that without such a separation of concerns, "the next generation of publishing" will not scale enough to become practical.

References

[1] C. David, D. Ginev, M. Kohlhase, J. Corneli, eMath 3.0: Building blocks for a social and semantic web for online mathematics & ELearning, in: I. Mierlus-Mazilu (Ed.), 1st International Workshop on Mathematics and ICT: Education, Research and Applications, 2010.
URL http://kwarc.info/kohlhase/papers/malog10.pdf

[2] H. Stamerjohanns, M. Kohlhase, D. Ginev, C. David, B. Miller, Transforming large collections of scientific publications to XML, Mathematics in Computer Science 3 (3) (2010) 299–307.
URL http://kwarc.info/kohlhase/papers/mcs10.pdf

[3] arxmliv build system, http://arxivdemo.mathweb.org, seen Sept. 2010.
URL http://arxivdemo.mathweb.org

[4] M. Kohlhase, OMDoc – An open markup format for mathematical documents [Version 1.2], no. 4180 in LNAI, Springer Verlag, 2006.
URL http://omdoc.org/pubs/omdoc1.2.pdf

[5] S. Buswell, O. Caprotti, D. P. Carlisle, M. C. Dewar, M. Gaëtano, M. Kohlhase, The Open Math standard, version 2.0, Tech. rep., The OpenMath Society (2004).
URL http://www.openmath.org/standard/om20

[6] R. Ausbrooks, S. Buswell, D. Carlisle, G. Chavchanidze, S. Dalmas, S. Devitt, A. Diaz, S. Dooley, R. Hunter, P. Ion, M. Kohlhase, A. Lazrek, P. Libbrecht, B. Miller, R. Miner, M. Sargent, B. Smith, N. Soiffer, R. Sutor, S. Watt, Mathematical Markup Language (MathML) version 3.0, W3C Proposed Recommendation of 10. August 2010, World Wide Web Consortium (W3C) (2010).
URL http://www.w3.org/TR/MathML3

[7] M. Kohlhase, Using LaTeX as a semantic markup format, Mathematics in Computer Science 2 (2) (2008) 279–304.
URL https://svn.kwarc.info/repos/stex/doc/mcs08/stex.pdf

[8] V. Zholudev, M. Kohlhase, TNTBase: a versioned storage for XML, in: Proceedings of Balisage: The Markup Conference 2009, Balisage Series on Markup Technologies, Mulberry Technologies, Inc., 2009, available at http://kwarc.info/vzholudev/pubs/balisage.pdf.
URL http://kwarc.info/vzholudev/pubs/balisage.pdf

[9] J. Giceva, C. Lange, F. Rabe, Integrating web services into active mathematical documents, in: Carette et al. [41], pp. 279–293.
URL https://svn.omdoc.org/repos/jomdoc/doc/pubs/mkm09/jobad/jobad-server.pdf

[10] C. Müller, Adaptation of Mathematical Documents, Ph.D. thesis, Jacobs University Bremen (2010).
URL http://kwarc.info/cmueller/papers/thesis.pdf

[11] M. Schönert, et al., GAP – Groups, Algorithms, and Programming, Lehrstuhl D für Mathematik, Rheinisch Westfälische Technische Hochschule, Aachen, Germany (1995).

[12] K. Hammond, P. Horn, A. Konovalov, S. Linton, D. Roozemond, A. A. Zain, P. Trinder, Easy composition of symbolic computation software: A new lingua franca for symbolic computation, in: Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation (ISSAC), ACM Press, 2010, pp. 339–346.

[13] C. Lange, M. Kohlhase, A mathematical approach to ontology authoring and documentation, in: Carette et al. [41], pp. 389–404.
URL http://kwarc.info/kohlhase/papers/mkm09-omdoc4onto.pdf

[14] C. Lange, Enabling collaboration on semiformal mathematical knowledge by semantic web integration, Ph.D. thesis, Jacobs University Bremen, submitted January 31, 2011 (2011).
URL https://svn.kwarc.info/repos/swim/doc/phd/phd.pdf

[15] A. Kohlhase, M. Kohlhase, C. Lange, Dimensions of formality: A case study for MKM in software engineering, in: S. Autexier, J. Calmet, D. Delahaye, P. D. F. Ion, L. Rideau, R. Rioboo, A. P. Sexton (Eds.), Intelligent Computer Mathematics, no. 6167 in LNAI, Springer Verlag, 2010, pp. 355–369, http://arxiv.org/abs/1004.5071.

[16] R. Harper, F. Honsell, G. Plotkin, A framework for defining logics, Journal of the Association for Computing Machinery 40 (1) (1993) 143–184.
[17] T. Mossakowski, Hets: the Heterogeneous Tool Set, web page at http://www.informatik.uni-bremen.de/cofi/hets/, seen July 2005.
URL http://www.informatik.uni-bremen.de/cofi/hets/
[18] F. Pfenning, C. Schürmann, System description: Twelf — a meta-logical framework for deductive systems, in: H. Ganzinger (Ed.), Proceedings of the 16th Conference on Automated Deduction, no. 1632 in LNAI, Springer Verlag, 1999, pp. 202–206.
[19] C. David, D. Ginev, M. Kohlhase, B. Matican, S. Mirea, A framework for modular semantic publishing with separate compilation and dynamic linking (2011).
URL https://svn.mathweb.org/repos/planetary/doc/sepublica11/paper.pdf
[20] PlanetMath Redux.org – math for the people, by the people, http://planetmath.mathweb.org, seen January 2011.
URL http://planetmath.mathweb.org
[21] PlanetMath.org – math for the people, by the people, http://www.planetmath.org, seen January 2011.
URL http://planetmath.org
[22] J. Gardner, A. Krowne, L. Xiong, Nnexus: Towards an automatic linker for a massively-distributed collaborative corpus, IEEE Transactions on Knowledge and Data Engineering 21 (6).
[23] A. M. Society, Mathematics Subject Classification MSC2010, http://www.ams.org/mathscinet/msc/ (2009).
[24] A. Kohlhase, M. Kohlhase, C. Lange, sTeX – a system for flexible formalization of linked data, in: A. Paschke, N. Henze, T. Pellegrini, H. Weigand (Eds.), Proceedings of the 6th International Conference on Semantic Systems (I-Semantics) and the 5th International Conference on Pragmatic Web, ACM, 2010.
URL http://kwarc.info/kohlhase/papers/isem10.pdf
[25] D. Ginev, C. Jucovschi, S. Anca, M. Grigore, C. David, M. Kohlhase, An architecture for linguistic and semantic analysis on the arXMLiv corpus, in: Applications of Semantic Technologies (AST) Workshop at Informatik 2009, 2009.
URL http://www.kwarc.info/projects/lamapun/pubs/AST09_LaMaPUn+appendix.pdf
[26] Executable paper challenge system, http://trac.mathweb.org/planetary/wiki/EPCSystem.
URL http://trac.mathweb.org/planetary/wiki/EPCSystem
[27] M. Kohlhase, I. Şucan, A search engine for mathematical formulae, in: T. Ida, J. Calmet, D. Wang (Eds.), Proceedings of Artificial Intelligence and Symbolic Computation, AISC'2006, no. 4120 in LNAI, Springer Verlag, 2006, pp. 241–253.
URL http://kwarc.info/kohlhase/papers/aisc06.pdf
[28] C. Lange, U. Bojārs, T. Groza, J. Breslin, S. Handschuh, Expressing argumentative discussions in social media sites, in: J. Breslin, U. Bojārs, A. Passant, S. Fernández (Eds.), Social Data on the Web (SDoW), Workshop at the 7th International Semantic Web Conference, no. 405 in CEUR Workshop Proceedings, Aachen, 2008.
URL http://ceur-ws.org/Vol-405/paper4.pdf
[29] A. Kohlhase, M. Kohlhase, Maintaining islands of consistency via versioned links, submitted (2011).
URL http://kwarc.info/kohlhase/submit/mkm11-verlinks.pdf
[30] J. Carette, W. Farmer, High-level theories, in: Autexier et al. [42], pp. 232–245. doi:10.1007/978-3-540-85110-3_19.
[31] S. Autexier, C. David, D. Dietrich, M. Kohlhase, V. Zholudev, Workflows for the management of change in science, technologies, engineering and mathematics, submitted (2011).
[32] E. Hilf, M. Kohlhase, H. Stamerjohanns, Capturing the content of physics: Systems, observables, and experiments, in: J. Borwein, W. M. Farmer (Eds.), Mathematical Knowledge Management (MKM), no. 4108 in LNAI, Springer Verlag, 2006, pp. 165–178.
URL http://kwarc.info/kohlhase/papers/mkm06physml.pdf
[33] A. Kohlhase, M. Kohlhase, Compensating the computational bias of spreadsheets with MKM techniques, in: Carette et al. [41], pp. 357–372.
URL http://kwarc.info/kohlhase/papers/mkm09-sachs.pdf
[34] A. Kohlhase, M. Kohlhase, Spreadsheet interaction with frames: Exploring a mathematical practice, in: Carette et al. [41], pp. 341–356.
URL http://kwarc.info/kohlhase/papers/mkm09-framing.pdf
[35] A. Kohlhase, M. Kohlhase, What we understand is what we get: Assessment in spreadsheets, in: S. Thorne (Ed.), Symp. of the European Spreadsheet Risks Interest Group (EuSpRIG 2010), European Spreadsheet Risk Interest Group, 2010, pp. 111–121.
URL http://www.kwarc.info/kohlhase/papers/eusprig10-coverage.pdf
[36] A. Kohlhase, M. Kohlhase, Semantic transparency in user assistance systems, in: B. Mehlenbacher, A. Protopsaltis, A. Williams, S. Slatterey (Eds.), Proceedings of the 27th annual ACM international conference on Design of communication (SIGDOC), ACM Special Interest Group for Design of Communication, ACM Press, New York, NY, USA, 2009, pp. 89–96. doi:10.1145/1621995.1622013.
URL http://kwarc.info/kohlhase/papers/sigdoc09-semtrans.pdf
[37] M. Kohlhase, et al., Planet GenCS, web page at http://gencs.kwarc.info, seen January 2011.
URL http://gencs.kwarc.info
[38] Logic atlas and integrator, web page at logicatlas.omdoc.org, seen January 2011.
URL http://logicatlas.omdoc.org
[39] Planetary developer forum, http://trac.mathweb.org/planetary/.
URL http://trac.mathweb.org/planetary/
[40] Planetary sandbox, web page at http://planetbox.kwarc.info, seen January 2011.
URL http://planetbox.kwarc.info
[41] J. Carette, L. Dixon, C. Sacerdoti Coen, S. M. Watt (Eds.), MKM/Calculemus Proceedings, no. 5625 in LNAI, Springer Verlag, 2009.
[42] S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, F. Wiedijk (Eds.), Intelligent Computer Mathematics, no. 5144 in LNAI, Springer Verlag, 2008.
[43] Intelligent Computer Mathematics, submitted.