Sustaining the swMATH project: Integration into zbMATH Open interface and Open Data perspectives

Maxence Azzouz-Thuderoz, Moritz Schubotz and Olaf Teschke

1   Introduction

The swMATH database was launched in 2013 as the main result of a joint 2011–2013 project of FIZ Karlsruhe and Mathematisches Forschungsinstitut Oberwolfach supported by the Leibniz Association, aiming to increase the visibility of research software contributions in mathematics, as well as to provide quality measures by evaluating the software usage in peer-reviewed publications indexed in the zbMATH database [1, 2]. Since then, the service has been produced by FIZ Karlsruhe, further developed in collaboration with the Zuse Institute Berlin (ZIB) supported by the Forschungscampus MODAL [3–5]. Due to its history, the service has been, on the one hand, closely connected to zbMATH, and, on the other hand, it has been developed and maintained on a single, independent platform, which grew out of the initial project. This resulted in both advantages (some features could be implemented independently) and disadvantages (full, and ever-growing, functionalities of zbMATH could not be transferred easily to swMATH). We describe here the new version of swMATH fully integrated to the zbMATH Open framework, which has been made possible by the transition to an open service [6, 7].

2   swMATH as part of zbMATH Open transition

Since its inception, swMATH has been a free service, although its publication part has always been based on data contained in zbMATH, which was a subscription-based service until 2020. That led to some restrictions in designing swMATH at that time: as a general rule, only limited paywalled information was made available via swMATH. This resulted in a restriction of certain functions; e.g., swMATH was not interlinked with the author and journal databases of zbMATH, and the retrieval lacked the options of full logical combinations of the various search fields. Moreover, and perhaps even more unsatisfactory, swMATH was open access, but not open data. On the other hand, the much reduced data allowed for a rather lean, independent front-end implementation employing the Django framework. This facilitated the addition of some small extra features.
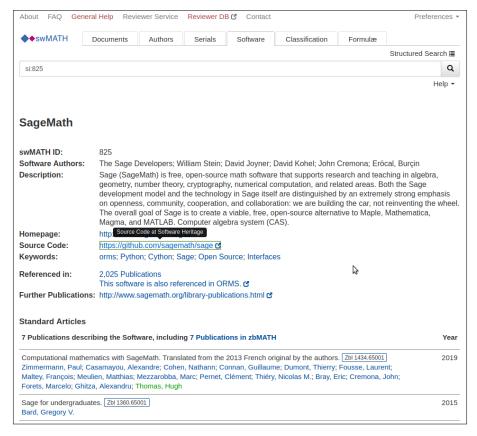
With the transition toward zbMATH Open, the main obstruction causing reduced swMATH functions became obsolete [6]. Simultaneously, in 2020/2021, a lot of internal preparatory work was completed that allowed a more flexible development, like the replacement of the indexing component or the migration of the code to Python 3 [7]. While before the system had been quite specialised, with a focus on bibliographic data, it became now much easier to add additional layers. At the same time, the life cycle of the Django-based software came to an end, making a replacement necessary.

The natural next step of swMATH at this point involved three directions of development: First, to make the swMATH data available through an API, adding a truly open data layer to the service and enabling its use in various interfaces. Second, to establish a fully integrated software facet within zbMATH Open. And third, to provide an independent platform for features incompatible with the integrated version. Here, we will mainly concentrate to report on the first two aspects.

3   Integrated functions available for software search

The search in the software layer of zbMATH Open allows now, as in the case of the other search facets, any logical combination of expressions in the various available search fields (software, name, authors, classification, keywords, and identifier are indexed) in the one-line search. Likewise, wildcard search with * is now available. Results are sorted in descending order by the number of articles referencing a software package (default), or alphabetically. The information for a single package is arranged in the detailed zbMATH Open profile standard, as shown in the figure below.

The profile information contains not just all information familiar from the old swMATH platform (though arranged in a different manner), but also a granular and interlinked breakdown of the publications using the software with respect to authors, journals, and subjects. Furthermore, while the old swMATH platform had just a static list of these documents, this is now directly interlinked to the dynamic result page in zbMATH Open, where the results can be further filtered, refined, or extended.

Profile of SageMath software

4   swMATH and API

As part of the FAIRCORE4EOSC project, the new independent MediaWiki-based swMATH version will become an integrated component of the European open science cloud (EOSC). By standardised API specifications such as [9, 10] OAI-PMH and CodeMeta [8], swMATH data will become an integral component of the EOSC. This contributes to five of nine core components of FAIRCORE4EOSC.

1. The EOSC Metadata Schema and Crosswalk Registry (**MSCR**) aims to support publishing, discovery and access of metadata schemata and provides functions to operationalise metadata conversions by combining crosswalks. Through the new API, research software programs and their metadata will be easily accessible. This component will be exposed so that one can use it to convert metadata of mathematical research software to ease the querying process.

2. For each mathematical software program, the EOSC Research Software APIs and Connectors (**RSAC**) will ensure the long-term preservation of research software. We will demonstrate it easily, as the Software Heritage identifier SWHID of the archived project will be displayed on the new swMATH website.

3. Moreover, swMATH will integrate to the EOSC PIDGraph, a knowledge graph which improves the way of interlinking research entities across domains and data sources on the basis of PIDs (Persistent Identifiers); this can play a key role in helping applied mathematicians to identify the most convenient mathematical research software programs.

4. Standardised access to swMATH data is provided by integrating swMATH in the EOSC PID Meta Resolver (**PIMDR**).

5. Eventually, links between software and publications will also be discoverable via the EOSC Research Discovery Graph (**RDGraph**) which ensures that whenever a paper associated to a software that is indexed in swMATH is found in the Graph, the link to the software can be used in the EOSC context.

The only drawback of this version is that the classical swMATH contained also some additional references to documents not (yet) indexed in zbMATH Open, as, e.g., from arXiv, which do not fit into this format, since they are not yet indexed with respect to authors, journals, or MSC (Mathematics Subject Classification). This information is, however, not lost, but will be on display in the upcoming independent swMATH platform based on the MediaWiki framework.

5 Conclusion and outlook

We outlined the new opportunities gained from the new integrated swMATH version, which first and foremost improves the situation for mathematicians who frequently use zbMATH Open. Software is now natively supported by the zbMATH Open website. At the same time, we outlined how we will improve the accessibility to swMATH data for researchers in Europe who are used to interact with the European open science cloud. Everyone using the EOSC can benefit from swMATH data without even knowing that swMATH exists or understanding the design of the swMATH website. Utilising standards and processes developed in EOSC, the data will also be an integral component of the common research data good within Europe. Also, swMATH will benefit from other EOSC data sources that can be used to improve the quality of the service further. We envision that the integration with the EOSC will also pave the ground for manifesting the role of swMATH, among many domain-specific aggregators for research software, as one that takes into account the particularities of mathematics, and at the same time takes advantage of the technology and methodological insights and achievements from aggregators in different domains.

References

[1] S. Böhnisch, G.-M. Greuel and W. Sperber, Building an information service for mathematical software – the SMATH project. *Eur. Math. Soc. Newsl.* **83**, 51–52 (2012)

[2] S. Böhnisch, G.-M. Greuel and W. Sperber, The software information service swMATH – release of the first online prototype. *Eur. Math. Soc. Newsl.* **87**, 48–49 (2013)

[3] H. Chrapary, W. Dalitz, Software products, software versions, archiving of software, and swMATH. In *Mathematical Software – ICMS 2018, Proceedings of the 6th International Conference (South Bend, 2018)*, edited by J. H. Davenport et al., Lect. Notes Comput. Sci. 10931, Springer, Cham, 123–127 (2018)

[4] H. Chrapary, W. Dalitz, W. Neun and W. Sperber, Design, concepts, and state of the art of the swMATH service. *Math. Comput. Sci.* **11**, 469–481 (2017)

[5] H. Chrapary, Y. Ren, The software portal swMATH: A state of the art report and next steps. In *Mathematical Software – ICMS 2016, Proceedings of the 5th International Conference (Berlin, 2016)*, edited by G.-M. Greuel et al., Lect. Notes Comput. Sci. 9725, Springer, Cham, 397–402 (2016)

[6] K. Hulek and O. Teschke, The transition of zbMATH towards an open information platform for mathematics. *Eur. Math. Soc. Newsl.* **116**, 44–47 (2020)

[7] K. Hulek and O. Teschke, The transition of zbMATH towards an open information platform for mathematics (II): A two-year progress report. *Eur. Math. Soc. Newsl.* **125**, 44–47 (2022)

[8] M. B. Jones, C. Boettiger, A. C. Mayes, A. Smith, P. Slaughter, K. Niemeyer, Y. Gil, M. Fenner, K. Nowak, M. Hahnel, L. Coy, A. Allen, M. Crosas, A. Sands, N. C. Hong, P. Cruse, D. S. Katz and C. Goble, CodeMeta: An exchange schema for software metadata. Version 2.0, KNB Data Repository, DOI 10.5063/schema/codemeta-2.0 (2017)

[9] M. Petrera, D. Trautwein, I. Beckenbach, D. Ehsani, F. Müller, O. Teschke, B. Gipp and M. Schubotz, zbMATH Open: API solutions and research challenges. In *Proceedings of the Workshop on Digital Infrastructures for Scholarly Content Objects (Online, 2021)*, edited by W.-T. Balke et al., CEUR Workshop Proceedings, 4–13 (2021)

[10] M. Schubotz and O. Teschke, zbMATH Open: Towards standardized machine interfaces to expose bibliographic metadata. *Eur. Math. Soc. Newsl.* **119**, 50–53 (2021)

———

Maxence Azzouz-Thuderoz joined FIZ Karlsruhe in 2022 to develop core components of the European Open Science Cloud after many years as a natural language processing engineer in the industry. His goal is to make the swMATH database a connected resource of the EOSC ecosystem. As a researcher, his main research interests cover natural language processing, open science for mathematical software and recent development in artificial neural networks.

maxence.azzouz-thuderoz@fiz-karlsruhe.de

Moritz Schubotz is a senior researcher for mathematical information retrieval and decentralised open science at zbMATH Open. He maintains the support for mathematical formulae in Wikipedia and is an off-site collaborator at NIST.

moritz.schubotz@fiz-karlsruhe.de

Olaf Teschke is managing editor of zbMATH Open and vice-chair of the EMS Committee on publications and electronic dissemination.

olaf.teschke@fiz-karlsruhe.de