The transition of zbMATH towards an open information platform for mathematics (II): A two-year progress report

Klaus Hulek and Olaf Teschke

Two years ago, we outlined in this column [*Eur. Math. Soc. Newsl.* 116, 44–47 (2020)] *the vision of zbMATH Open as an open service for mathematics research. Here we give a report of the achievements since then.*

1 Introduction

Two years ago, we described in this column [3] the vision to transform zbMATH into an open service for mathematics research. This has now become reality. For this, we first received a special transformation grant from the German government, with the perspective that this could be made permanent after a successful evaluation after two years. In our original application we outlined several goals which are essential requirements of the mathematical community. We described and discussed these also in [3]. Naturally, the two-year period is not long enough to expect that these long-term goals could have been fully completed. In addition, the pandemic created its own problems, which needed to be addressed. In spite of this, we were able to achieve several important milestones, and the evaluation at the end of the transformation period confirmed that the grant should be made permanent. This now provides zbMATH Open with sustainable funding through the Leibniz Association. In this column, we report on the progress made since the beginning of 2020.

2 Preparatory work (I): Legal aspects

What has been known as the reviewing service *Zentralblatt*, or later the zbMATH database, has been, since its foundation by Springer Verlag in 1931, a commercial enterprise for many decades. Naturally, all legal documents, from the editorial contract to the indexing agreements with the many publishers active in mathematics, were based on the model of a subscription service, distributed to a well-defined and controllable set of customers (with their own licensing contracts of various terms). Transition to an Open Access, and beyond this, Open Data platform, required a complete replacement of these agreements, and related negotiations. One result of the new editorial contract among the editorial institutions of the European Mathematical Society, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, and Heidelberg Academy of Sciences was that the role of a commercial distributor, which had been faithfully fulfilled by Springer-Nature AG, became obsolete, leading to separation from a partner which had been very supportive of zbMATH, especially in difficult times. On this occasion, we would like to thank our Springer colleagues for their decades of commitment, which was concluded by enabling a coordinated transition from subscriptions to Open Access by the end of 2020. The role of the European Mathematical Society, as well as the Heidelberg Academy, is to ensure the scientific quality of the service and to further the involvement of the mathematical community.

Retrospectively, the amount of legal preparations achieved in 2020 is amazing. This includes a large number of renewed indexing contracts with a majority of mathematics publishers, a considerable fraction of which agreed also to Open Data services within the new zbMATH Open platform. While it had already been decided that all data generated within the zbMATH editorial process (such as reviews, author disambiguation data, or semantic and interlinking data) would be made available under the CC BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/), not all publisher data would fit into this framework (e.g., abstracts which would usually come along with a different copyright). Nevertheless, this could be achieved for a considerable part of the information, and there is ongoing activity to expand this further.

Also, the terms and conditions, both for users and reviewers, needed to be adapted and agreed upon accordingly. Finally, also the interface was revised, with a special focus on minimizing storage and processing of user data. While subscriptions required extensive user tracking and detailed usage reports for libraries, Open Access allowed for a platform built upon the principle of data avoidance and data minimization. As of 2021, zbMATH Open is indeed one of the few complex sites that can be used completely without cookies (and consequently, without the need for cookie approvals), with only optional cookies needed to store user preferences if required.

With these issues being completed in the course of 2020, it was possible to start zbMATH Open at the beginning of 2020.

EMS MAGAZINE 125 (2022) — DOI 10.4171/MAG/91

3 Becoming open: Usage and feedback

2021 being the first full year of zbMATH Open as a free service, it might be interesting to look at some experiences concerning usage figures and user feedback. Without doubt, the special situation of the still looming COVID-19 pandemic and its impact on working behaviour drove both the interest in and the willingness to contribute to open services. Historically, zbMATH had about 1,200 subscriptions, where access was often channelled via institutional proxies. This resulted in about 22 million customer searches per year. In an average month of 2021, more than 60,000 unique visitors used the site, with more than 32 million searches in 2021. This indicates that Open Access facilitated a much broader user base, which is still growing (2022 March figures point toward 40 million searches this year).

The user survey conducted mid-2021 supported this impression. While a lot of specific questions pertaining to zbMATH Open features, service and data quality confirmed a significant positive development in comparison with the already very good results of 2016 (see the report [1]), the main emphasis of the feedback was an unanimously enthusiastic appreciation of becoming open. Moreover, numerous ideas for the further improvement have been advanced, some of which could already be incorporated, such as an upgrade of the BibTeX output toward a more standardized format and the improvement of the site's accessibility. We experienced also a new overwhelming willingness to contribute reviews. Despite the universal limitations caused by the pandemic, almost 1,200 new reviewers joined the service (about twice as many as in previous years), and 13 % more reviews were contributed in 2021 compared to 2020.

Though it might be unfair to pick a single example, a good illustration of the increased reach is perhaps Peter Scholze's review of the *Publ. Res. Inst. Math. Sci.* volume containing Shinichi Mochizuki's work *Inter-universal Teichmüller theory* (https://zbmath.org/1465.14002). Openly available, is was within days distributed, linked, and discussed on a large variety of platforms (Reddit, Twitter, MathOverflow, …), and was retrieved within a few days more than 10,000 times, making it likely the most-read zbMATH review of all time.

4 Preparatory work (II): Backend upgrades

While the transition to an open service required also some accompanying developments, the main efforts were focused on new features that would be enabled by being an open data platform allowing to interlink with other free sources. While most of such features could not be deployed before 2021, there was a considerable amount of necessary preparations done in 2020. Among the essential upgrades in the backend was the replacement of the indexing software. Since going online in the mid-90s, zbMATH was based on an in-house code optimized for the specific data that traditionally formed the service. However, this came along with limitations as a growing interconnectedness lead to the import of heterogeneous data from various sources. Hence, the complete indexing code was replaced by Elasticsearch (www.elastic.co/de/elasticsearch/), which not just allowed for much more flexibility, but also led to a significant speed-up of update as well as search time. Moreover, the challenging part of retaining the traditional features of the service could be achieved.

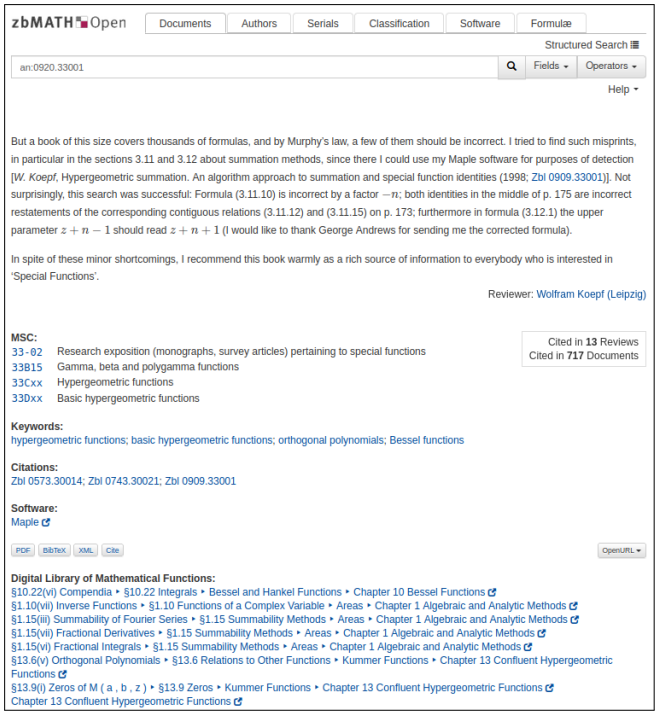

Figure 1. References to DLMF from the review of the Andrews, Askey and Roy's book *Special Functions*

Another key component was the development of a new reviewer backend that was also optimized with respect to the experiences of mobile work. It would have been impossible to handle the growth of the numbers of reviewers and reviews in 2021 without this system. Simultaneously, the backend and frontend components were recoded from Python 2 to Python 3, keeping up with the status of the programming language in which the system is developed.

5 New features

The first newly available open data facet even predated the open access of the interface: In September 2020, the first version of the zbMATH Open OAI-PMH API was released, based on the data of the *Jahrbuch über die Fortschritte der Mathematik* (JFM). This was

possible thanks to the European Mathematical Society and SUB Göttingen, which had made these data already available under a CC BY-SA 4.0 license (more details about JFM as a part of zbMATH Open have been given earlier in this column, see [11]). The experience acquired in this initial version facilitated its expansion to zbMATH Open data in a stable version in 2021 [7, 9], which forms now a core component in distributing zbMATH Open data. Based on this interface, further APIs with the aim of supporting extended standards and interlinking specific services are being developed.

An immediate application was the development of a linking API enabling the interlinking of zbMATH Open and the NIST Digital Library of Mathematical Functions (DLMF). The system, which can be adapted in the future to interlink with further research data resources, was described in this column before [2], and the resulting interlinking data are now also visible in zbMATH Open.

The more traditional tasks of providing literature and author information also benefitted considerably from the open approach. The improved and extended zbMATcH API allowed for the interlinking with many more open fulltext sources, not just twice as many integrated arXiv links, but also almost 100,000 new links to Unpaywall and CiteSeer^X resources. Author profiles contain now identifiers from and links to as much as 15 external services. Vice versa, additional information from these sources can be included, enabling, e.g., the display of various non-ASCII spellings. Furthermore, the profiles differentiate now between various roles, like author, editor, or further contributions (like appendix authors). Likewise, additional information from the coauthor graph is displayed (both features were frequently asked for in the user surveys).

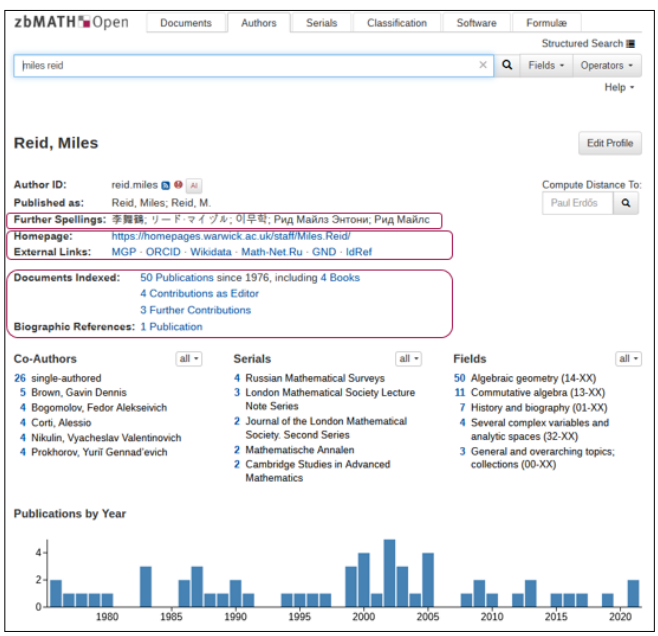

Figure 2. 2022 zbMATH author profile

It is perhaps also an interesting aspect that the underlying author disambiguation data have been further improved, with significant support by the community interface [5] as well as various external community platforms. The latter comprises such different examples as the correction of several mistakes in the attribution of works of Renée Peiffer (of the Peiffer identity) on Twitter [4], as well as the insight that the Rabinowitsch trick was most likely not, as commonly assumed, discovered by G. Y. Rainich [12].

Another frequently requested facet that is under active development is an affiliation information. There already exists an internal database of about 15,000 disambiguated entities of mathematical institutions, which is currently matched to publications and authors. When completed, this will allow us to release a transparent, open data institution facet of zbMATH Open – we will be happy to report on its progress soon!

6 Future developments and projects

The ongoing internal development to integrate and interlink further publications, author and affiliation information, research data and community platforms is only one side of the evolving network. The other is the use of zbMATH Open data in projects conducted worldwide. Currently, there are several projects which already make extensive use of zbMATH Open data, especially in semantic analysis and the history of mathematics. For example, Norbert Schappacher's book [8], commissioned on the occasion of the IMU centenary, contains a detailed analysis of ICM speakers, their networks and working fields based on zbMATH Open data. For already several years, the ISC project Gender Gap in Science (https://gender-gap-in-science.org/), led by the IMU, is supported by zbMATH Open data which are an integral part of its visualization platform Gender Publication Gap (https://gender-publication-gap.f4.htw-berlin.de/; see also [6] in this column).

Through zbMATH Open data, FIZ Karlsruhe, the institution which produces, develops, and maintains zbMATH Open, was able to engage in several Open Science projects. Perhaps the most important is the Mathematical Research Data initiative (MarDI) (https://mardi4nfdi.de), which has been approved as the mathematics consortium within the evolving German National Research Data Infrastructure and started to work by the end of 2021. Other projects, pertaining to the EOSC cloud and the development of a math-specific plagiarism detection system based on [10], have already been approved, while others are in preparation.

But, above all, it is the mathematics community that drives the further development of zbMATH Open by providing ideas and contributions. We highly appreciate the ever-increasing number of valuable reviews, as well as your suggestions and feedback to editor@zbmath.org!

References

[1] I. Brüggemann, K. Hulek and O. Teschke, Results of the 2016 EMS user survey for zbMATH. *Eur. Math. Soc. Newsl.* **104**, 67–68 (2017)

[2] H. S. Cohl, O. Teschke and M. Schubotz, Connecting islands: bridging zbMATH and DLMF with Scholix, a blueprint for connecting expert knowledge systems. *Eur. Math. Soc. Mag.* **120**, 66–67 (2021)

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

[4] P. Koushik, https://twitter.com/PraphullaK/status/1368190495676108800

[5] H. Mihaljević-Brandt and N. Roy, zbMATH author profiles: open up for user participation. *Eur. Math. Soc. Newsl.* **93**, 53–55 (2014)

[6] H. Mihaljević and L. Santamaría, Mathematics publications and authors' gender: Learning from the Gender Gap in Science project. *Eur. Math. Soc. Mag.* **123**, 34–38 (2022)

[7] 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, A. de Waard, Y. Fu, B. Hua, J. Schneider, N. Song and X. Wang, CEUR Workshop Proceedings, 4–13 (2021)

[8] N. Schappacher, *Framing global mathematics: The International Mathematical Union between theorems and politics.* Springer, Cham (2022)

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

[10] M. Schubotz, O. Teschke, V. Stange, N. Meuschke and B. Gipp, Forms of plagiarism in digital mathematical libraries. In *Proceedings of the 12th International Conference of Intelligent Computer Mathematics, CICM 2019 (Prague, 2019)*, edited by C. Kaliszyk, Lect. Notes Comput. Sci. 11617, Springer, Cham, 258–274 (2019)

[11] O. Teschke, The "Jahrbuch über die Fortschritte der Mathematik" as a part of zbMATH Open. *Eur. Math. Soc. Mag.* **122**, 62–64 (2021)

[12] O. Teschke, Identity of J. L. Rabinowitsch (of Rabinowitsch Trick). https://mathoverflow.net/questions/416577/identity-of-j-l-rabinowitsch-of-rabinowitsch-trick

---

Klaus Hulek is professor of mathematics at Leibniz University Hannover and editor-in-chief of zbMATH Open. His field of research is algebraic geometry.

hulek@math.uni-hannover.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