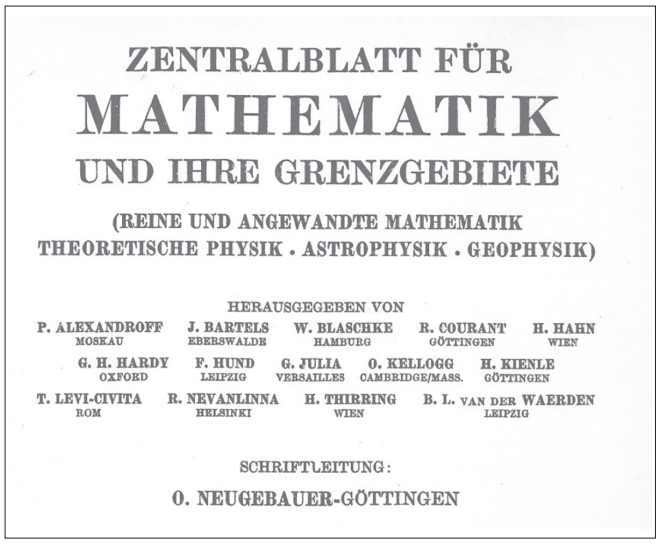
Editorial
zbMATH – Looking to the Future*

Klaus Hulek (Gottfried Wilhelm Leibniz Universität Hannover, Germany), Editor-in-Chief of zbMath

After 9 years in office as Vice-President for Research at Leibniz Universität Hannover, it had not been my intention to take on another administrative position. Consequently, I was somewhat reluctant (but still honoured) when representatives of FIZ Karlsruhe and Springer Verlag contacted me with the offer to become the new Editor-in-Chief of zbMATH. However, discussions with colleagues soon convinced me that this was not only an interesting and challenging task but also one that could provide a useful service to the mathematical community. I took over my new responsibilities in April 2016, succeeding (after an interlude with Olaf Teschke as acting Editor-in-Chief) Gert-Martin Greuel, who was Editor-in-Chief 2012–2015, and Bernd Wegner, who had held the office for 37 years before that.


ZENTRALBLATT FÜR
MATHEMATIK
UND IHRE GRENZGEBIETE
(REINE UND ANGEWANDTE MATHEMATIK
THEORETISCHE PHYSIK . ASTROPHYSIK . GEOPHYSIK)
HERAUSGEGEBEN VON
P. ALEXANDROFF MOSKAU · J. BARTELS EBERSWALDE · W. BLASCHKE HAMBURG · R. COURANT GÖTTINGEN · H. HAHN WIEN
G. H. HARDY OXFORD · F. HUND LEIPZIG · G. JULIA VERSAILLES · O. KELLOGG CAMBRIDGE/MASS. · H. KIENLE GÖTTINGEN
T. LEVI-CIVITA ROM · R. NEVANLINNA HELSINKI · H. THIRRING WIEN · B. L. VAN DER WAERDEN LEIPZIG
SCHRIFTLEITUNG:
O. NEUGEBAUER-GÖTTINGEN

The Editorial Institutions of zbMATH

zbMATH was founded as "Zentralblatt für Mathematik und ihre Grenzgebiete" by Otto Neugebauer and other prominent mathematicians in 1931 on the initiative of Harald Bohr and Richard Courant, among others. In 1939, it amalgamated with the even older "Jahrbuch über die Fortschritte der Mathematik", which was founded in 1868 by Carl Ohrtmann and Felix Müller. After the Second World War, Zentralblatt was re-established jointly by the Academy of Sciences and Springer Verlag. Even after the partitioning of Germany, Zentralblatt at first remained as one of the few German-German collaborations of that time, edited by both the Academy of Sciences in Berlin (East) and the Heidelberg Academy of Sciences, with Springer remaining the commercial partner. This collaboration was terminated by the GDR in 1977. Currently, zbMATH has three editorial institutions:

- European Mathematical Society (EMS).
- Heidelberg Academy of Sciences.
- FIZ Karlsruhe.

These organisations are jointly responsible for the content and scientific quality of the database. The financial and organisational responsibilities lie with Fachinformationszentrum (FIZ) Karlsruhe, whereas Springer Verlag remains responsible for distribution and marketing.

FIZ Karlsruhe (Leibniz Institute for Information Infrastructure) is a member of the Leibniz Association, a group of independent German research institutes. As such, it receives funding from the federal and state governments but also generates substantial revenues of its own. The German Mathematical Society (DMV) was a founding member of FIZ Karlsruhe (established in 1977). Mathematicians are typically familiar with the Leibniz Association through the Oberwolfach Research Institute (MFO) or the Weierstrass Institute for Applied Analysis and Stochastics (WIAS) in Berlin, both of whom are member institutes.

This structure ensures that zbMATH is not run primarily as a commercial enterprise. The income generated is used to finance the Berlin office of zbMATH with its approximately 30 full-time positions, as well as to maintain and develop the necessary infrastructure for the database. After all, it is crucial that the (currently approximately) 120,000 mathematical publications that appear each year in different sources (such as research monographs, conference proceedings, collected volumes, more than 2,200 journals and roughly 1,000 book series) are documented and reviewed in a timely fashion. For this, the contributions of our 7,000 reviewers are essential. Without these reviewers, zbMATH would not be possible.

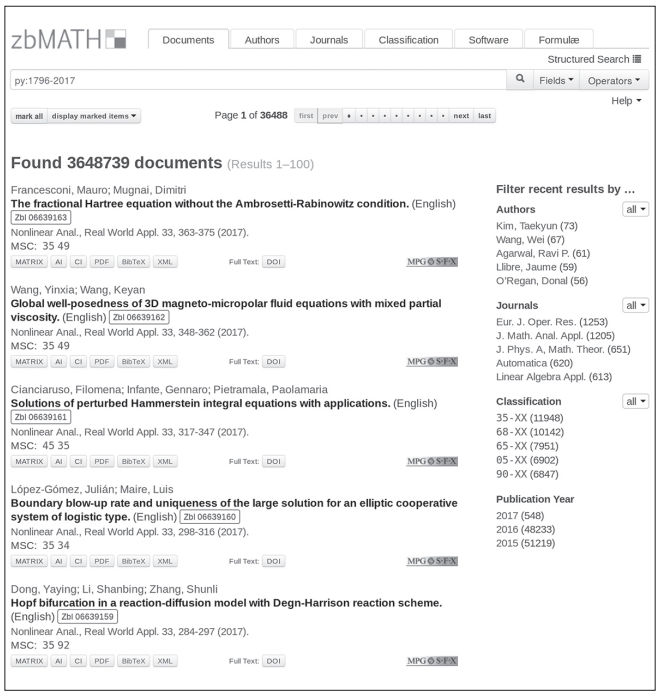
zbMATH caters for the needs of the mathematical community. As such it provides a number of free services:

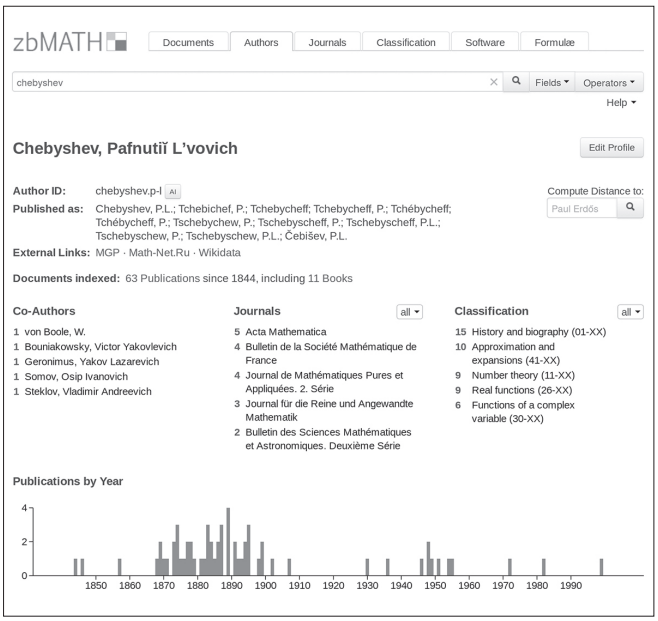
- Free access for developing countries.
- Free access for individual members of the EMS.
- Free access to author and journal profiles.
- Free information on mathematical software through the database swMATH.
- Free access to three search results, in particular to individual reviews.

The transition from print version to modern database

When I was a young researcher, I went to the library once a week, typically on Friday afternoon, to check Zentral-

* This is the adapted English version of an article that appeared in Mitteilungen der DMV 24, No. 3, 140–143 (2016).





blatt and Mathematical Reviews for recently published articles. At that time, this meant consulting the printed versions, which filled awe-inspiring yellow (Zentralblatt) and red (Mathematical Reviews) shelves in the library. Now that the print versions have been terminated, this is history. Also, it is now the mathematics arXiv that provides, at least in my field of research, much of the information about the latest articles – long before they are published. However, it should be noted that the maths arXiv so far covers 3.5% of all publications listed in zbMATH (even for current research, this figure only stands at 20% [1]).

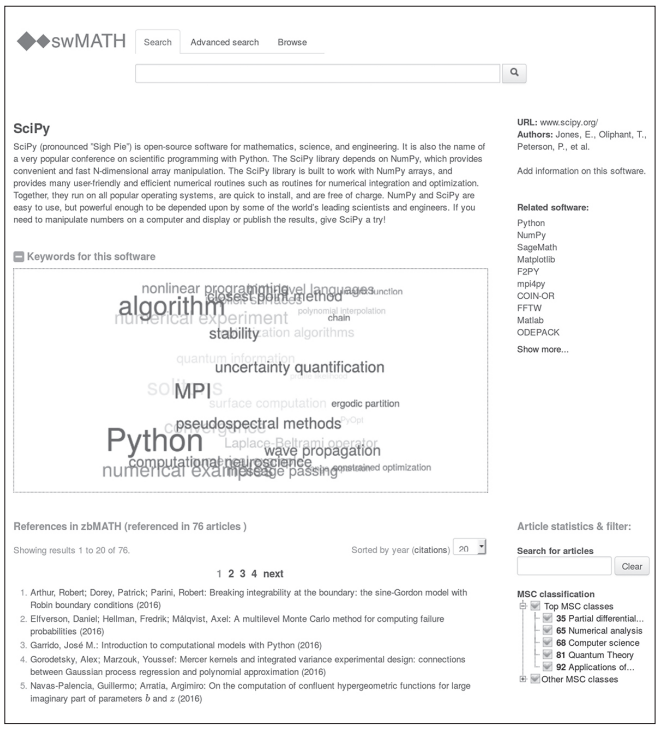
The transition from printed journal to modern database started under the aegis of Bernd Wegner and further important steps were taken in recent years when Gert-Martin Greuel was Editor-in-Chief. The challenges involved are considerable: the digitisation of the content from the 1980s and 1990s required a substantial investment and the same is true for making additional levels of information available.

The first of these is *author disambiguation*. Currently, the zbMATH author database comprises 950,000 entries with more than 6 million authorships, distributed among 3.6 million publications. In more than 20% of these cases, at least two authors share the same surname and abbreviated first name. Even with full names, it often happens that there are multiple possibilities. To take an example, the name "Alexander Schmidt" could be any of seven existing authors or, as an eighth possibility, it could refer to a new author. Similarly, there are many variants of names that could refer to the same person. This could, for example, be because a person has changed their name through marriage or it could be due to different transcriptions: the name Chebyshev, for example, can be found as 15 different variations.

Clearly, this work cannot be done solely intellectually. zbMATH combines advanced algorithms, information provided by users and editorial efforts in order to ensure optimally precise information. While the first version of the author database from 2008 still required substantial corrections, zbMATH has since advanced its algorithms and adapted them specifically to the needs of the zbMATH data. In addition, many manual checks have been carried out. The result is that the percentage of uncertain attributions has dropped to below 4% – and this refers not only to current publications but also to historic data, many of which come with a high share of uncertainty [2].

Another area of increasing importance concerns citation data. Combining references available in the database with full texts provides an additional tool, permitting us to trace mathematical developments. This often allows us to recognise unexpected connections or applications of our own research. Admittedly, it was MathSciNet which initiated this development in 2001, with zbMATH to follow somewhat later. However, in the last few years, the situation has improved significantly: while there were 3 million references available in 2010 for 150,000 articles, of which 30% were linked within the database, the figure now stands at 16 million references for more than 750,000 publications in the period 1873–2016, with 10.5 million linkages within the database. Due to the introduction of a new search interface, the information available through zbMATH can be accessed much more easily and comfortably, e.g. by using the new filter and profile functions.

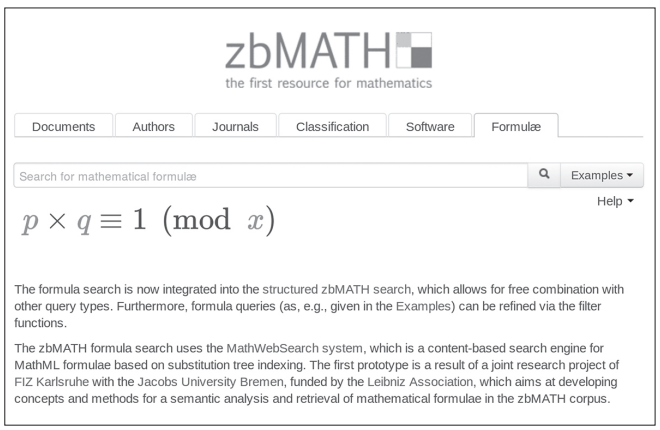
Meanwhile, zbMATH has introduced new functions that allow access to information beyond the realm of classical literature search. One example is the widely used platform swMATH, which originated from a joint project of the Leibniz Association in 2011–2013, involving MFO, FIZ Karlsruhe and other partners. There have been previous collections of mathematical software, such as the "Guide to Mathematical Software" of NIST or the "Oberwolfach References on Mathematical Software". However, it is only the timely connection with the accessible research literature that makes it possible to provide a service allowing not only the researching and highlight-

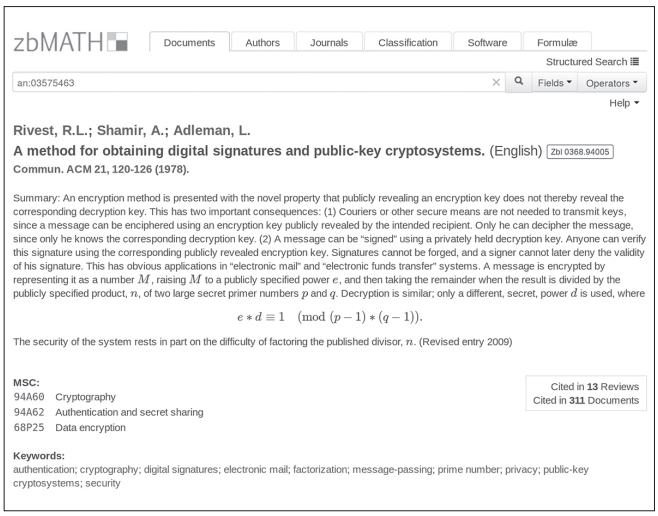


ing of software use but also access to information about what software has been used for comparable research problems.

When Gert-Martin Greuel took over as Editor-in-Chief, this database was developed further. Since 2013, swMATH has been run by FIZ Karlsruhe, which provides it as a free service to the community. Since then, the service has grown tremendously – it now covers over 15,000 software packages and provides approximately 125,000 records of use in research publications [3]. Under the aegis of the research campus MODAL, the Zuse Institute Berlin (ZIB) has been a cooperation partner of swMATH since 2015.

Another new feature of zbMATH is the search for mathematical formulae. Based on the MathML standardisation of zbMATH contents, this allows a search for contents that are often difficult to describe or locate based solely on a textual description. Variables allow the user to search for potentially substitutable terms. Recently, this feature has been extended to a formula search in more than 120,000 full arXiv texts (see the zbMATH





column in this issue [4]). Even though this development is in its infancy, it allows us to glimpse the enormous potential that modern information technology can provide for mathematics. This has found expression in the vision of the Global Digital Mathematics Library: to provide an as complete as possible collection of all mathematical research in a highly standardised form, which, for example, relieves the mathematician of tedious routine tasks by providing access to suitable mathematical software via adapted interfaces [5].

Another important feature of a database such as zbMATH is completeness. This is a challenge that must not be underestimated. Each year, around 300 new journals apply to zbMATH asking to be listed, all of them claiming to publish serious research whose quality is assured by a peer review process. Clearly, some very good new journals have been founded in recent years. At the same time, it is also obvious that not all of the new journals conform to these strict requirements. Indeed, there are fewer than 100 journals a year that are added to zbMATH, although the overall number of journals is growing at an even lower rate, since some journals are discontinued and other journals no longer fulfil the required quality standards. It is a complex challenge to observe the ever-changing landscape of publications and to adjust the necessary decisions concerning their indexing. In addition to this, roughly 3,000 research monographs and conference proceedings are published every year and it has always been a top priority of zbMATH to provide coverage of these as complete as possible.

The Future

With the rapid development of information technology, the conditions under which a database like zbMATH operates are constantly changing. Thus, it is vital that zbMATH constantly reassesses where it stands and where it wants to go. An immediate task in hand is updating the MSC classification. While this system may have lost some of its historic importance as a tool for bibliographic classification, its importance has grown in connection with derived semantics, data analysis and profiling information. For this reason, it is of great importance that new branches of mathematics are adequately represented.

This revision will be undertaken jointly with MathSciNet, as announced at 7ECM and in the Newsletter of the EMS and the Notices of the AMS. The restructuring will be done on the basis of comments received from the mathematical community. You are invited to submit your suggestions either at msc2020.org or by email to feedback@msc2020.org.

There are, however, a number of much more *fundamental challenges*. Information and relevant data on mathematical publications can nowadays be obtained in many ways. Apart from the specialised service provided by the mathematics arXiv, this includes such different sources as Google, Google Scholar, Researchgate, Scopus or the Web of Science. Why then do we still need zbMATH or MathSciNet?

There are several special features concerning information about mathematical research. One is the longevity of mathematical results: the average halftime regarding the citation of mathematical publications is currently 15 years and indicators predict that this will grow even further [6]. Another is the immense importance of a consistent corpus and the necessity of error detection and correction. Finally, mathematical publications are characterised by a high degree of formalisation and an enormous density of information encoded in the mathematical language. It is high quality reviews, written by mathematical experts, summarising the results and putting them into the context of current research, that have hitherto proven an excellent tool in making this information accessible. In future, this will be further supported by the database for software (swMATH) and formula search. But there are many other aspects, and I would like to conclude by briefly mentioning some, which, I believe, will play an important role in the future.

Bibliometric data

We are all acutely aware that bibliometric data are playing an increasingly important role. Many of us are sceptical, and with good reason. At the same time, if we are honest, most of us also use them in some form or another, not least in hiring processes or peer reviews of research proposals. What is important is to know which data are reliable, and for which purposes and in which form they can be used without giving wildly distorted results. On one hand, this requires a reliable high quality database, such as zbMATH, and, on the other hand, it needs the expertise of how these data can be used and how they are to be interpreted. These are questions that can only be answered in collaboration with the mathematical community.

Research data

Research data and big data are buzz words that many mathematicians do not immediately connect with their own research. But this is no longer a tenable attitude. Clearly, mathematics can and will play a major role in the handling and analysis of big data. At the same time, mathematics has started producing its own research data on a large scale. Mathematical statistics is obviously concerned with data from diverse areas, ranging from medicinal data to data from highly sophisticated physical experiments; modelling and simulation requires and generates terabytes of data. Mathematical software, and the resulting research and benchmarks, has become its own ecosystem of data. In the meantime, we have a huge set of mathematical objects that have been collected and described in various forms: these range from integer sequences in the *On-line Encyclopedia of Integer Sequences*, mathematical functions in the *NIST Digital Library of Mathematical Functions*, manifolds in the Manifold Atlas, and lists of Calabi-Yau varieties and modular forms in the *L-functions and modular forms database*, to collections of formalised mathematical definitions, theorems and proofs in systems such as Coq, HOL or Mizar. To produce sustainable access to and linking of this information produces many questions to which we currently have, at best, rudimentary answers.

Non-textual material

Printed or electronic articles and books are now far from being the only ways in which mathematical information is documented and distributed. Numerous lectures (which have always been an essential form of mathematical communication) are now available on the internet, be it as videos or presentations. Some of these, but by no means all, are directly associated to traditional publications and can thus be linked with these. Should zbMATH take this up and incorporate this material in its database? The answer to this question will also depend on how the behaviour of users and working styles develops. The questions raised here concern most of the mathematical community and will, in the future, be discussed at various levels, in private discussions as well as in official committees of professional societies. zbMATH is prepared to face these challenges and I am looking forward to any comments or suggestions you may have.

In the wake of 7ECM, the EMS has conducted a zbMATH user survey. While a detailed analysis will be given in the zbMATH column of a subsequent issue, I would like to mention two interesting conclusions here. The first, and for us a gratifying result, is that the community values the great effort zbMATH has made over the last few years to update its web interface and to provide new services such as author and citation profiles and swMATH. The second, and this was not at all clear to us, is that formula search seems to be considered a valuable asset, in particular in connection with search in full texts. We look forward to giving a detailed analysis of the feedback in a future column of this newsletter.

References

[1] F. Müller and O. Teschke, Will all mathematics be on the arXiv (soon)? Eur. Math. Soc. Newsl. 99, 55–57 (2016).

[2] M. Jost, N. Roy and O. Teschke, Another update on the collaboration graph, Eur. Math. Soc. Newsl. 100, 58–61 (2016).

[3] G.-M. Greuel and W. Sperber: swMATH – ein neuer Service für die Suche nach mathematischer Software, Mitteilungen der DMV, 21, 12–13 (2013).

[4] F. Müller and O. Teschke, Discovering formulae in arXiv full-texts via zbMATH, Eur. Math. Soc. Newsl. 102, (2016).

[5] https://en.wikipedia.org/wiki/Global_digital_Mathematics_Library.
[6] Th. Bouche, O. Teschke and K. Wojciechowski, Time lag in mathematical references. Eur. Math. Soc. Newsl. 86, 54-55 (2012).

Klaus Hulek studied mathematics at the University of Munich and the University of Oxford. He completed his PhD in Erlangen in 1979 and spent the academic year 1982/83 at Brown University, Providence, RI, USA. He received his habilitation in Erlangen in 1984. Klaus Hulek was a professor of mathematics in Bayreuth from 1985 to 1990, when he moved to Hanover. He held the position of Vice-President of Research of Leibniz Universität Hannover from 2005 to 2014, and from 2007 to 2014 he represented the German Rectors' Conference (HRK) at the Research Policy Working Group of the European University Association (EUA). Klaus Hulek was a member of the Institute for Advanced Study (IAS) in Princeton in 2015. His field of research is algebraic geometry.