

A REST API for zbMATH Open access

Marcel Fuhrmann and Fabian Müller

1 Introduction

For some time now, zbMATH Open has been offering its services in various formats and through different media. Biweekly printed volumes for long decades in the past, distribution via CD-ROMs and DVDs, and the now standard (and only) access option as an online service, have been incarnations of the service over the years. Typography and design have changed many times in the printed volumes as well as the online version. Furthermore, the online version has offered new ways of distributing the data for specific purposes. The contents of zbMATH Open are available in HTML, B_IT_EX, and PDF form, as well as via an XML-based OAI-PMH API (more on these three-letter acronyms below). In 2023 a new option was added, a so-called REST API, which makes the data available in a machine-readable fashion, for automated use by researchers. We will outline the features of this new API and describe new use cases made possible by this new way of accessing zbMATH Open.

2 Why do we need APIs?

API is an acronym for “application programming interface.” It denotes any part of a software system that makes it accessible through programs and machines, as opposed to human users, and is an important part of technology for software developers. The “interface” component of an API is a system that allows two or more software programs to interact and understand each other without communication problems. One could think of it as a sort of intermediary that helps software programs share information with each other.

Creating an API establishes a common language for wildly divergent systems to communicate to each other. Instead of having to learn each other’s specifications, both entities can interact in a common format, which can also be used by any other application or service. The so-called OpenAPI Specification defines a standard, language-agnostic interface to HTTP APIs. This allows both humans and computers to discover and understand the capabilities of the service without access to source code. The user can understand

and interact with the remote service with a minimal amount of implementation effort. An OpenAPI definition can then be used by documentation generation tools to display the API, code generation tools to generate servers and clients in various programming languages, testing tools, and many other use cases. This provides advantages for users and data providers alike, granting access in a machine-readable, automatable way for interested partners, e.g., research groups.

3 The OAI-PMH API of zbMATH Open

In 2002, an informal organisation of providers of repository services called the “Open Archives Initiative,” or “OAI” for short, created a common protocol that allowed users to automatically harvest metadata from such a service in a unified way using programmatic tools. Unsurprisingly, this was called the “Protocol for Metadata Harvesting,” or “PMH.” It is supported by numerous providers of scientific or other databases. In mathematics, of course, by far the largest of them is the arXiv. In September 2020, zbMATH Open started to support this protocol as well.¹

The obvious advantages of using such a common protocol are that it is well established, easy to implement, and has a substantial positive impact on working mathematicians. Aggregators, archives, and bibliometric researchers commonly use the OAI-PMH, and there are ready-made client libraries available in every major programming language. Furthermore, this protocol is well suited for harvesting the entire open collection of zbMATH Open document data.² Moreover, it is possible to restrict the harvesting process to well-defined subsets, and to only consume updates since the last download, obliterating the need to download all the data every time.

¹ Available at oai.zbmath.org. For further information, see [2].

² The data made available through the OAI-PMH API is provided under a [CC-BY-SA 4.0 license](https://creativecommons.org/licenses/by-sa/4.0/), as are all other data offers by zbMATH Open made through APIs. Legal constraints by publishers that are outside the control of zbMATH Open are currently restricting the amount of data that can be distributed through APIs.

4 Why do we need another API?

Although the OAI-PMH API has numerous advantages – not least among them its widespread adoption – it also has various shortcomings:

- The standard format supports only a very limited set of metadata (although additional formats can be defined).
- It is XML-based only.
- It only supports filtering by time and by subset.
- It is only designed to retrieve data, not to update or add any.

Due to this lack of flexibility, we decided to create a new REST API for zbMATH Open, with the OAI-PMH API as a starting point. The acronym “REST” stands for “representational state transfer” and is a style of software architecture that defines how different systems communicate, again improving reusability and interoperability. An API that complies with these REST principles is known as a REST API.

The new zbMATH Open REST API³ now gives access to a number of additional benefits. First, improving the data set offered by OAI-PMH, a complete set of bibliographic metadata for zbMATH documents is available to any interested party. The complex metadata structure of any document covers a vast variety of information. A quick overview of the available data sets is shown in Figure 1.

```
result:
  0:
    biographic_references: []
    contributors:
      authors:
        0:
          aliases: []
          checked: "1"
          codes: [...]
          name: "Daubechies, I."
        1:
          aliases: []
          checked: "1"
          codes: [...]
          name: "Teschke, G."
          author_references: []
          editors: []
          document_type: "journal article"
          editorial_contributions: [...]
          id: 2189397
          keywords: [...]
          language: {...}
          links: [...]
          msc: [...]
          references: [...]
          source: {...}
          states: [...]
          title: {...}
          year: "2005"
          zbmath_url: "https://zbmath.org/2189397"
```

Figure 1. Screenshot of data structure of an example result, as seen in the browser.

Furthermore, the API offers access to information about published authors and their work, their awards, any external IDs, and so on. Other data endpoints cover information about the current version of Mathematics Subject Classification codes for documents (see [1]), metadata information about books and journals, any published document available within the zbMATH Open database, and metadata information about research software. All these available data sets are similar to what is available at the zbMATH Open website. The difference is that the data is machine-readable and therefore can be easily used for any purpose, e.g. any bibliographic research efforts.

Most importantly, it is now possible to search in the REST API in exactly the same way as on the website zbmath.org, including arbitrary combinations of logical operators and filtering by all supported fields (in particular, arbitrary time frames).

The search result is a machine-readable status report in JSON format, containing the requested data, but also including information about the search execution (for example, whether it was successful or not, and if not, why). In the future, other formats will be supported as well (e.g., XML).

All the information is documented in the Swagger UI (see the screenshot of Figure 2),⁴ so that, based on the information there, other developers may create their own APIs for their own

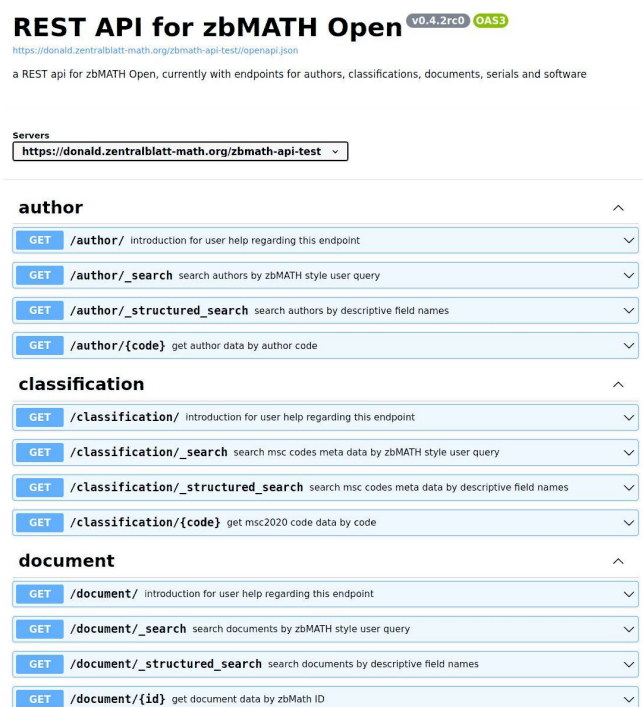


Figure 2. Screenshot of the user interface of zbMATH Open REST API.

³ Available at <https://api.zbmath.org>.

⁴ For more information, see <https://swagger.io/docs/>.

research projects. Like in the case of the OAI-PMH API, in some instances, parts of the information (e.g., abstracts or references, based on the agreements with the respective publisher) need to be redacted for legal reasons. The API takes care of that automatically.

5 Outlook: Pushing data via the REST API

Currently, we are working on adding more features to the zbMATH Open API. So far, data can only be retrieved from the database. A next step is to enable users to also add new data in an automated way. We identified a number of use cases for that, which would improve the updating process of the zbMATH Open database and facilitate the incorporation of input from the user community.

The most straightforward use case would be to enable the upload of bibliographic metadata information about documents from publishers directly. Although some larger publishers have their own data delivery processes, for many (especially smaller ones) it can be quite helpful to offer a well-documented way of adding their bibliographical data to zbMATH Open.

Other use cases involve uploading references or external IDs (like DOI or arXiv identifier) of existing documents to enrich the data contained within the zbMATH Open database. Moreover, the upload of, and therefore the access to, full-text documents is planned to be possible in the near future. This requires resolving any licensing issues, in particular the uploader (e.g., the original author) needs to certify that they have the right to upload the full text for distribution by zbMATH Open.

Finally, adding or correcting new information about authors and their connection to existing articles (if this information was missing before) is going to be possible via the zbMATH Open REST API.

6 Final thoughts

To summarize, the new zbMATH Open REST API will be a useful tool for automated access to the data stored within the zbMATH Open database. It will extend the possibilities currently offered by the OAI-PMH API (without superseding it). With the oncoming features of adding and/or updating information to the database in an automatable way, open access for zbMATH Open data will continue to expand.

References

- [1] E. Dunne and K. Hulek, [Mathematics Subject Classification 2020](#). *Eur. Math. Soc. Newsl.* 115, 5–6 (2020)
- [2] M. Schubotz and O. Teschke, [zbMATH Open: Towards standardized machine interfaces to expose bibliographic metadata](#). *Eur. Math. Soc. Mag.* 119, 50–53 (2021)

Marcel Fuhrmann studied physics at Humboldt-Universität zu Berlin. He completed his PhD in astrophysics and worked as a postdoctoral researcher in applied mathematics at the University of Potsdam. From 2017 to 2021 he worked at the Federal Institute for Risk Assessment. In 2021 he moved to FIZ Karlsruhe and works on the zbMATH Open REST API and other related projects.

marcel.fuhrmann@fiz-karlsruhe.de

Fabian Müller studied mathematics and computer science at Humboldt-Universität zu Berlin. After finishing his doctoral studies in algebraic geometry in 2013, he started working at zbMATH Open, where he is team lead for development and systems.

fabian.mueller@fiz-karlsruhe.de