

Bravo MaRDI: A Wikibase Knowledge Graph on Mathematics

Moritz Schubotz¹, Eloi Ferrer², Johannes Stegmüller¹, Daniel Mietchen¹,
Olaf Teschke¹, Larissa Pusch² and Tim Conrad²

¹*FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Berlin, Germany*

²*Zuse Institute Berlin, Berlin, Germany*

Abstract

Mathematical world knowledge is a fundamental component of Wikidata. However, to date, no expertly curated knowledge graph has focused specifically on contemporary mathematics. Addressing this gap, the Mathematical Research Data Initiative (MaRDI) has developed a comprehensive knowledge graph that links multimodal research data in mathematics. This encompasses traditional research data items like datasets, software, and publications and includes semantically advanced objects such as mathematical formulas and hypotheses. This paper details the abilities of the MaRDI knowledge graph, which is based on Wikibase, leading up to its inaugural public release, codenamed Bravo, available on <https://portal.mardi4nfdi.de>.

This paper is submitted to the ISWC Wikidata Workshop 2023 novel resource track.

1. The One-Stop-Portal Vision

The Mathematical Research Data Initiative (MaRDI) [10] strives to bring the FAIR data principles [23] to life for mathematics and its research data (MathRD). A critical aspect of this ambition involves creating a unified portal¹ that offers comprehensive access to all open research data within mathematics and related fields.

In this manuscript, we introduce the first Bravo release of the portal, now freely accessible through direct access or various APIs. Rather than constructing such a portal from scratch - a tactic adopted by parallel initiatives in Germany's National Research Data Initiative (NFDI) - we chose to utilize Wikimedia technology, predominantly MediaWiki, as a foundation, for several reasons:

- MediaWiki offers a thriving open-source environment with a myriad of free and open-source software.
- A considerable user base exists for MediaWiki; many mathematicians are already acquainted with Wikipedia and, to some extent, Wikidata.

Wikidata'23: Wikidata workshop at ISWC 2023

✉ moritz.schubotz@fiz-karlsruhe.de (M. Schubotz); ferrer@zib.de (E. Ferrer);


johannes.stegmueller@fiz-karlsruhe.de (J. Stegmüller); daniel.mietchen@fiz-karlsruhe.de (D. Mietchen);

olaf.teschke@fiz-karlsruhe.de (O. Teschke); pusch@zib.de (L. Pusch); conrad@zib.de (T. Conrad)

🌐 <https://schubotz.org> (M. Schubotz)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

¹Analogous initiatives name this type of service as 'Single Point of Access' [11].

- Upon the project’s completion, we have the option to incorporate our data into Wikidata or another compatible platform.
- Both the *Wikimedia mathematics community* and portal users, notably mathematicians based in Germany, mutually benefit from our advancements.
- We seamlessly inherit knowledge from Wikipedia and Wikidata, painstakingly curated by hundreds of thousands of contributors over the past two decades.

However, this decision comes with challenges; becoming part of the *Wikimedia movement* necessitates embracing their work culture and guidelines and engaging with the community to achieve consensus on sustainable solutions. Our contributions in this context are as follows:

1. We deliver a current ecosystem to house Wikibase and associated services on-premises.
2. We customize the general-purpose Wikibase infrastructure to cater to the specific needs of the mathematical community.
3. We have devised a mechanism for importing external data that aligns with the Wikidata data model.
4. We have actively enhanced the MediaWiki ecosystem by expanding and improving services such as *QuickStatements*, MediaWiki extensions like *Math* and *MathSearch*, and libraries like the *Wikibase Integrator*.
5. We have seeded the *MaRDI KG* with nearly 100 million triples² aggregated from various sources.

This paper is organized as follows: Section 2 begins with an overview of our technology stack. This is followed by Section 3, which presents our mathematics-specific extensions. Subsequently, Section 4 details the infrastructure for seeding and updating the knowledge graph. In Section 5, we illustrate the functionality of our system with an example. Finally, Section 6 provides a conclusion and discusses future prospects.

2. Technical Infrastructure Overview

The *MaRDI portal*’s infrastructure was initially constructed using the *Wikibase release pipeline*. The release pipeline incorporates a *docker-compose* file that delineates several services, including *MediaWiki* with *Wikibase* extensions, the *Wikidata Query Service*, and the *Blazegraph Knowledge Graph* backend. The *docker-compose* file allows for the launching of a pre-set group of services via a single command, thus optimizing reproducibility and system or web host independence.

The *MaRDI portal* operates on a solitary open-stack instance with 32 GB of main memory and 8 CPUs, alongside a 300 GB HDD. Only *Docker* and *Git* are installed on the host system. All

²Refer to the latest statistics from <https://portal.mardi4nfdi.de/wiki/Portal>.

configurations are housed in partially encrypted Git repositories, which feature a CI pipeline enabling direct deployment to production without necessitating access to the actual machine. If the data volume increases, we have the capability to transition from a single-node deployment to a Docker swarm with minimal modifications to our service specifications.

The following provides a concise description of selected services from our infrastructure, which currently comprises 28 services [8]:

Traefik As the primary entry point to our portal, all external network traffic is channeled through the Traefik reverse proxy. Traefik is responsible for managing SSL certificates properly via in-built lets-encrypt management and protective measures to guard individual services against unauthorized use. Additionally, Traefik collates access logs, enabling centralized usage examinations for our interconnected services.

MediaWiki In order to maintain compatibility with Wikipedia and manage only one version of mathematics-specific extensions, we chose to fork the infrastructure and update the container with the same versions utilized by Wikipedia in production starting from a previous effort [4].

WDQS WDQS is the abbreviation of Wikidata Query Service which we have adapted to query the MaRDI portal and renamed it to MaRDI Query Service. This service is deployed based on four Docker containers: the Blazegraph database storing the knowledge graph, an updater that synchronizes the Wikibase entries with the Blazegraph backend, a frontend GUI to perform SPARQL queries, and a proxy to protect the SPARQL endpoint. The GUI for this service can be accessed at query.portal.mardi4nfdi.de.

LaTeXML LaTeXML is a service that converts LaTeX input to content and presentation MathML. It is employed for MathSearch.

BaseX BaseX is an XML database utilized for formula search. The MathSearch extension is responsible for keeping this service current.

Backup A straightforward, custom backup process was installed to preserve precisely the data we wish to retain, including the information saved on the SQL database, the MediaWiki pages as XML and the uploaded files.

Scholia Scholia is a tool that facilitates scholarly exploration of subsets of Wikidata’s knowledge graph via a web browser [19]. Its backend queries the Wikidata Query Service through predefined SPARQL queries parametrized by Wikidata identifiers, and its frontend displays the results of several of these queries as profiles. A multitude of such profile types exist, including works, venues, topics, authors, awards, events, or organizations. MaRDI’s instance³ is deployed via a fork [3] of the original [2] deployed on the Wikimedia Toolforge⁴. The intent behind the fork is to ensure maximum customizability for mathematics, with general improvements fed back upstream wherever possible. The

³<https://scholia.portal.mardi4nfdi.de/>

⁴<https://scholia.toolforge.org/>

ongoing customization includes efforts towards federated queries that utilize both the MaRDI knowledge graph and Wikidata.

Portainer Portainer is a management tool for Docker containers, images, volumes, and networks that operates via a web interface. It provides multiple ways of interacting with containers, including viewing their logs and accessing the container console. Portainer is compatible with standalone Docker installations, Docker Swarm, and Kubernetes.

Watchtower Watchtower is a tool that keeps Docker images up to date and guarantees the early installation of critical patches, provided Docker containers follow semantic versioning.

Prometheus & Grafana Prometheus is a monitor and alerting toolkit to monitor the health of the portal and the infrastructure. We visualize the data gathered by Prometheus using the dashboarding tool Grafana.

To guard against unintended side effects arising from changes to this multifaceted service infrastructure, we implemented a *Testing* concept that is triggered by any modification. This helps increase our confidence in not deploying erroneous code to our production environment.

Additionally, Identity and Access Management (NFDIIAM) is a crucial aspect. MaRDI utilizes OAuth as a protocol, which is also supported by the NFDI's wide basic service, NFDIIAM.

3. Mathematics-specific Extensions

Our portal leverages two math-specific MediaWiki extensions, namely, Extension-Math [26, 22, 15, 14] and Extension-MathSearch [28, 20, 17]. The former extension, deployed across all Wikimedia projects, including Wikipedia, facilitates the rendering of mathematical expressions. The latter makes mathematical expressions searchable and links knowledge graph data to HTML5.

3.1. Rendering of Mathematical Expressions

The MaRDI portal, alongside Wikipedia and Wikidata, employs texvc markup to express mathematical formulas. This LaTeX variant does not permit runtime syntax changes, favoring context-free formula processing with a regular grammar. Mathematical expressions in wikitext are denoted by the wikitext tag `<math>` (to be distinguished from the HTML5 element `<math>`). These `<math>` tags enclose the texvc expressions, which are then rendered. In contrast, Wikibase introduces a data type specifically for mathematical expressions. Examples of such notations can be found in the DLMF defining formula in the MaRDI-KG [24]⁵ and in math-tagged elements in Wikitext⁶.

Typically, portal users view page content in their web browsers. The Extension-Math processes the texvc math from the Wikitext or the Wikibase entities into HTML, allowing the

⁵<https://portal.mardi4nfdi.de/wiki/Item:Q1799>

⁶https://portal.mardi4nfdi.de/wiki/Non-negative_Matrix_Factorization_for_Time-Resolved_Raman_Spectroscopy_Data

browser to display the rendered formulas correctly. This rendering pipeline leverages an external REST-based service to generate SVG images from texvc within the MediaWiki ecosystem [26].

As of 2023, all Chromium-based browsers support MathML, a web standard defined by the W3C⁷. MathML is part of HTML5, akin to SVG, and enables native math rendering without needing images or browser extensions. Thus the latest versions of Edge, Chrome, Firefox, and Safari⁸ fully support HTML5, including MathML since January 2023. Due to changes in the MediaWiki ecosystem⁹ shifting away from microservices, and to augment processing capabilities for semantic annotation, the MaRDI Project developed a rendering pipeline written in pure PHP [5]. It generates MathML directly within the Math extension without the need for communication with REST-based services.

Initially, TexVC leverages a parsing expression grammar to extract a language-independent parse tree from LaTeX formulas. This parse tree verifies the formula’s syntactical correctness and provides feedback to authors in edit mode [12]. To validate this component’s functionality, we used 312,190 formulas extracted from the English Wikipedia [27].

The parse tree then undergoes root-first traversal, translating each item to a corresponding MathML item. The native MathML component creates the correct MathML representation for the 724 LaTeX commands supported by TexVC. The MathML used for automatic validation is generated from Mathoid [26] and LaTeXXML [29].

The MaRDI project contributed to the open-source community by developing a PHP version¹⁰ of the TypeScript-based mhchemParser [1, 13], integrating it into the MediaWiki ecosystem [9]. The *mhchem* syntax is employed alongside the mathematical formula notation to write chemical equations within the Wikibase ecosystem. Implementing this component in PHP enables the processing of these equations within the PHP-based math processing pipeline. The component’s functionality was validated using 116 automated tests, which incorporate the mhchem specification¹¹.

The native MathML pipeline in the MaRDI Wikibase paves the way for further advancements, such as parsing formulas from arXiv, semantic annotation of formulas, enhancing accessibility for visually impaired users, and resolving disambiguation scenarios.

3.2. Formula Search in the MaRDI Portal

The MaRDI Portal employs the MathSearch extension[28]¹² to search mathematical expressions based on texvc input. The original version utilized the Math Web Search engine, also used in the formula search of zbMATH Open [21]. In the MaRDI project, we optimized the formula search components for the Docker-based infrastructure running the latest MediaWiki versions. The MaRDI setup¹³ uses the BaseX database to store the search index created for all formulae in the MaRDI portal.

⁷<https://www.w3.org/TR/MathML/>

⁸<https://www.lambdatest.com/web-technologies/mathml>

⁹<https://phabricator.wikimedia.org/T303822>

¹⁰<https://gerrit.wikimedia.org/r/c/mediawiki/extensions/Math/+923597>

¹¹<https://texdoc.org/serve/mhchem/0>

¹²<https://www.mediawiki.org/wiki/Extension:MathSearch>

¹³<https://github.com/MaRDI4NFDI/formulasearch>

4. Seeding Mathematical Research Data

4.1. Defining Mathematical Research Data (MathRD)

In the realm of mathematics, research data (MathRD) [16, 6] encompasses all data forms integral to the research process. These data forms facilitate the creation and analysis of mathematical models, substantiate proofs, test algorithms, and generally elucidate mathematical phenomena. MathRD presents itself in various types, structures, and degrees of accessibility and interoperability.

MathRD ranges from highly structured forms like symbolic or numerical data, used for encapsulating theorems, proofs, number sequences, or matrices. Geometric data, delineating objects like curves, surfaces, or polytopes, also fall into this category. In contrast, MathRD may also include less structured forms such as mathematical models or observational data. At the extreme end of the spectrum, plain text data like scientific papers, online resources, articles, and books are also encompassed under MathRD.

The MaRDI knowledge graph is engineered to effectively capture the intricate structure of this data. It does this by aggregating metadata from diverse sources, each describing mathematical research data, into a consolidated graph.

4.2. Constructing the MaRDI Knowledge Graph

As previously alluded to in Section 1, the data model for the MaRDI knowledge graph is an extension of the existing data model in Wikidata. Practically, this indicates that entities for the MaRDI knowledge graph are sourced from Wikidata if pre-existing there. Once integrated into the MaRDI knowledge graph, these entities can be further enriched with additional statements, though the detailed modalities of whether and how to keep the MaRDI data in sync with Wikidata remain to be worked out.

This approach is particularly vital in the case of properties, as it necessitates the design of our data model based on the available properties in Wikidata. New properties are defined only when a corresponding one is absent in Wikidata. As a result, many of the properties found in Wikidata are also present in the MaRDI knowledge graph. We anticipate the introduction of new properties as the mathematical knowledge graph expands, which will be specifically linked to mathematics.

To facilitate entity import from Wikidata, we have established an entity importer functionality built upon the Python module `WikibaseIntegrator`¹⁴. This module forms the basis for a Python class capable of transferring any given entity or list of entities from Wikidata to our Wikibase instance. This importer class copies not only the label, description, and aliases of the entity but also its statements, with the option to import additional connected entities. The importer class includes a parameter that determines the desired depth level of the import. By default, one level of depth is selected, importing all statements for each imported entity, and importing only label, description, and aliases for the entities mentioned in these statements.

The MaRDI knowledge graph further includes properties such as *wikidata PID*¹⁵ and *wikidata*

¹⁴<https://github.com/LeMyst/WikibaseIntegrator>

¹⁵<https://portal.mardi4nfdi.de/wiki/Property:P11>

*QID*¹⁶, which store the Wikidata identifier for each imported property and item, respectively. This information is also stored in an internal SQL table, mapping the MaRDI knowledge graph identifiers to Wikidata identifiers. This table additionally incorporates a parameter for each entity indicating whether all its statements have been imported or only its label, description, and aliases.

This approach facilitates a comprehensive understanding of the overlap between the MaRDI knowledge graph and Wikidata at any given time. Moreover, it ensures synchronization between imported entities and Wikidata, and maintains consistency between the knowledge graphs, thus making the eventual integration of the MaRDI knowledge graph into Wikidata seamless.

The properties imported from Wikidata bolster the creation of new items in the MaRDI knowledge graph. In addition to the Wikidata items, metadata from eight distinct sources have been imported to varying degrees. These include formula metadata from DLME, software metadata from swMATH, publication metadata from zbMATH Open, arXiv and crossref, metadata from resources stored at Zenodo, metadata on discrete geometric objects stored at polyDB¹⁷ and metadata on R packages published at CRAN¹⁸.

4.3. Persistent Identifiers and Compliance with FAIR Principles

Adherence to FAIR principles necessitates the assignment of persistent identifiers to each linked resource in the MaRDI knowledge graph. It's essential to differentiate between extrinsic and intrinsic identifiers:

- **Extrinsic identifiers:** These identifiers are linked to a specific resource through a register. DOIs, Wikibase IDs or ORCID IDs are instances of this category.
- **Intrinsic identifiers:** These identifiers are innately linked to the resource and do not necessitate an external register for association. An example is the name of a software package, if used as its identifier.

Every resource in the MaRDI knowledge graph is assigned an extrinsic identifier by being created as a Wikibase entity. These follow the format used in Wikidata, with item identifiers beginning with 'Q' and property identifiers with 'P'. Additionally, entities imported from a given source incorporate a statement that includes the identifier from the original source. Currently, the MaRDI knowledge graph supports the following extrinsic identifiers:

- DOI
- ORCID iD
- Digital Library of Mathematical Functions ID
- swMATH work ID

¹⁶<https://portal.mardi4nfdi.de/wiki/Property:P12>

¹⁷<https://polydb.org/>

¹⁸<https://cran.r-project.org/>

- zbMATH Open document ID
- zbMATH author ID
- arXiv ID
- arXiv author ID
- Mathematics Subject Classification ID
- Zenodo ID

Further, two types of resources are linked to their source using extrinsic identifiers:

- CRAN Project: This identifier, imported from Wikidata¹⁹, links to an R package published at the Comprehensive R Archive Network. The identifier is a string that corresponds to the name of the R package.
- PolyDB ID: This identifier, created in the MaRDI knowledge graph, links to the original metadata source of a polyDB collection. The identifier string corresponds to the name of the collection.

5. Case study: Orthogonal Polynomials and Special Functions

In this section, we showcase some capabilities of the MaRDI portal using semantically enhanced Mathematical formulas as an example, specifically focusing on Orthogonal Polynomials and Special Functions (OPSF).

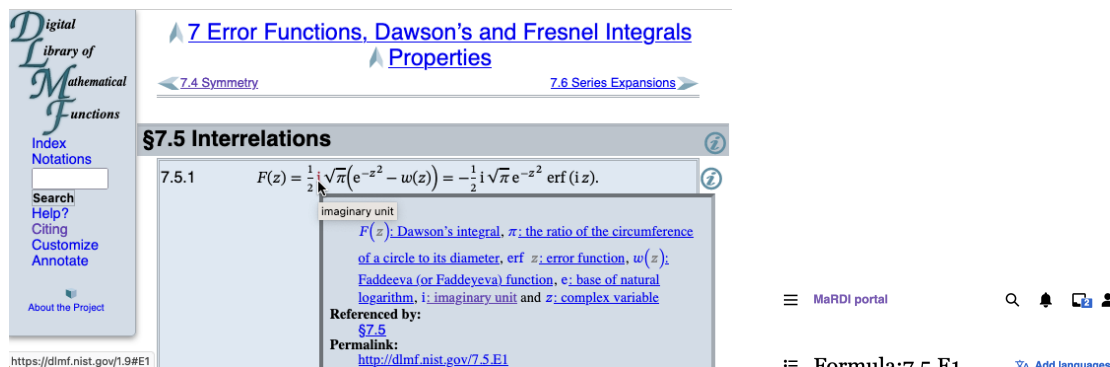
In the field of OPSF, relationships between functions are essential. The NIST Digital Library of Mathematical Functions, along with its companion, the Digital Repository of Mathematical Formulae [25, 24], is an indispensable source of identifiers and formulas pertaining to this field. Their main function is as aggregators of well-established knowledge, rather than providers of novel work.

A special variant of LaTeX has been developed by the team to enable a higher level of machine-readable semantics. For example, one typically writes i in LaTeX to represent the imaginary unit. In the context of DLMF/DRMF, however, one would use `\iunit`. Both versions render the same output. However, when the LaTeX source code is web-rendered rather than PDF-generated, the `\iunit` version offers the added advantage of being clickable. This improves accessibility by linking to a human-readable explanation of what the imaginary unit represents.

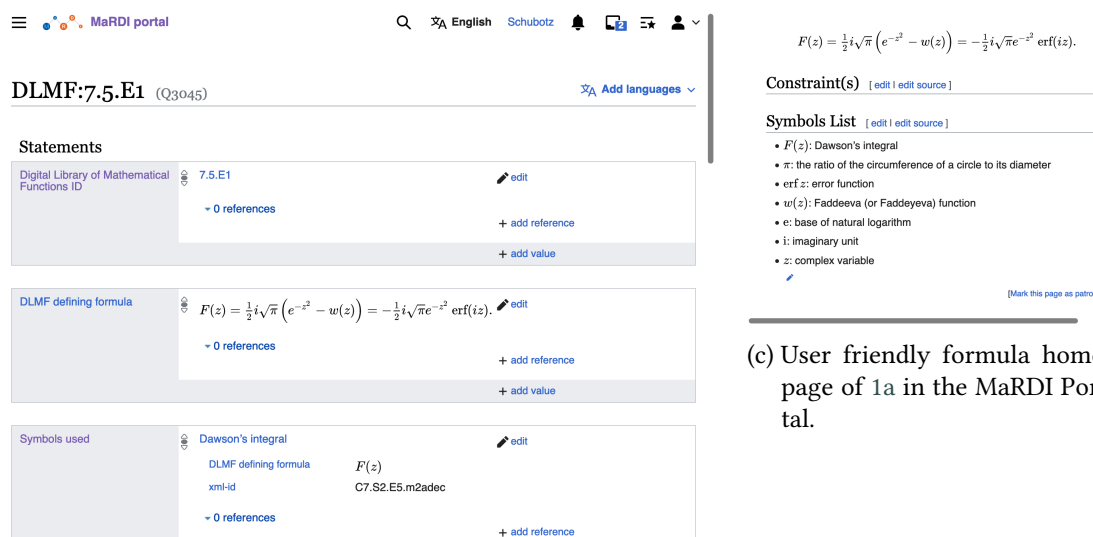
In Figure 1a, `\iunit` links to <https://dlmf.nist.gov/1.9#E1>, which is considered a permanent identifier. This means that the numbering remains consistent, irrespective of the introduction of additional material.

To make DLMF data accessible from the MaRDI portal, we have imported the formulas into our knowledge graph. Each formula has a corresponding Wikibase item. For example, the formula displayed in Figure 1a corresponds to `Q3045` and links to the definition `Q1399`.

¹⁹<https://www.wikidata.org/wiki/Property:P5565>



(a) Screenshot from the NIST Digital Library of Mathematical Functions (DLMF) <https://dlmf.nist.gov/7.5> demonstrating the use of `\iunit`.



(b) Representation of 1a in the MaRDI Wikibase Instance.

(c) User friendly formula home page of 1a in the MaRDI Portal.

Figure 1: A DLMF formula in different representations.

Our SPARQL query interface enables users to retrieve all formulas that include a link to the imaginary unit with the following command²⁰:

```
SELECT ?item WHERE { ?item wdt:P4 wd:Q1399 . }
```

More complex queries are also supported, such as searching for all formulas indirectly dependent on the gamma function.

We have additionally created formula homepages, similar to the DRMF project, to make the data available for classic searches and to offer a more convenient view compared to the standard Wikidata item view. Once the import of zbMATH articles is completed, bibliographic references related to the formula, software models, and algorithms will be linked from here.

²⁰<https://tinyurl.com/2bd4jv8k>

Our software can also translate the expression from DLMF to Maple and Mathematica [7]. However, due to potential licensing issues, we only display the formulas on the separate site <https://lct.wmflabs.org>, and not in the portal.

6. The road ahead

The Bravo Release of the MaRDI platform served as a proof of concept for leveraging MediaWiki and associated services from the Wikimedia ecosystem as a platform for the National Research Data Infrastructure, focusing on the field of mathematics. We successfully demonstrated the platform’s capability to handle domain-specific requirements, such as the manipulation of mathematical expressions, proving the adaptability of this general-purpose platform for specialized needs.

However, we recognize that several aspects of the platform’s usability require enhancement. The current interface design, reminiscent of a Wikipedia fork, needs to be reimaged to reflect the nuances of a research data management platform. Furthermore, while advanced users might find formulating SPARQL queries straightforward, the average user might need additional support. Although Wikidata provides visual aids to improve query generation, we aim to go a step further by integrating a conversational interface to facilitate SPARQL query generation following our previous MathQA efforts [18].

The platform’s citation capabilities also leave room for improvement. While MediaWiki has long provided permanent links (using the oldid flag), these could be further enhanced. For instance, we are contemplating the implementation of a feature that stores a snapshot of the PDF on Zenodo for convenience. However, the issue of citing query results remains an open research problem, posing a challenge for reproducibility in data-intensive mathematics.

The road ahead is not without its obstacles, but we are committed to continuous improvement and development. Our aim is to deliver a platform that not only adheres to the FAIR principles but also becomes an integral part of the mathematical research data ecosystem, thereby facilitating collaboration and advancement in mathematical research.

Acknowledgments

This work was funded by the MaRDI project [10] under DFG grant number 460135501. The text of this manuscript was improved with several AI tools, including Grammarly and ChatGPT.

References

- [1] [Software excerpt] M. Hensel and M. van den Berg, *mhchemParser*. SWHID: [⟨https://github.com/mhchem/mhchemParserswh:1:cnt:ddf2c75241bf79a175e655176b7808b07f7e9783;origin=https://github.com/mhchem/mhchemParser;visit=swh:1:snp:7ec88857e3585721a5525a67085731cde73bfb51;anchor=swh:1:rev:a4456994f67c21819affc40e415ce5b404df05ab;path=/src/mhchemParser.ts⟩.](https://github.com/mhchem/mhchemParserswh:1:cnt:ddf2c75241bf79a175e655176b7808b07f7e9783;origin=https://github.com/mhchem/mhchemParser;visit=swh:1:snp:7ec88857e3585721a5525a67085731cde73bfb51;anchor=swh:1:rev:a4456994f67c21819affc40e415ce5b404df05ab;path=/src/mhchemParser.ts)

- [2] [Software Release] D. Mietchen, F. Å. Nielsen, C. MacKenzie, E. Willighagen, A. Rutz, and J. Poelen, *Scholia*. SWHID: `<swh:1:dir:a90e241734badc22ae4154488d320b1087fd4263;origin=https://github.com/WDScholia/scholia;visit=swh:1:snp:475bacef8c802eefdad22c940284b89767fa8e37;anchor=swh:1:rev:6d14811f1a52ec68c2b359f64807b306fd82d257>`.
- [3] [Software Release] D. Mietchen and M. Schubotz, *MaRDI Fork of Scholia*. SWHID: `<swh:1:dir:b19ebb2df2a5a706b9ed0123e2e4c9847bb3b802;origin=https://github.com/MaRDI4NFDI/scholia;visit=swh:1:snp:293cd1547e8700fb806efabd54a8fa272538c061;anchor=swh:1:rev:cd9083745656c676edb1c81f8655d5d7a333a5b3>`.
- [4] [Software Release] M. Schubotz, *mediawiki-docker*. SWHID: `<swh:1:dir:c574bf0d9641a95d65cb2eb676b9cfb161781b60;origin=https://github.com/physikerwelt/mediawiki-docker;visit=swh:1:snp:1996cb919e55391cc057329bf311c28a5f3be4e7;anchor=swh:1:rev:56e6f1f27d4d1bde6fe2cdfd8f004c15636d6a29>`.
- [5] [Software Module] J. Stegmüller and M. Schubotz, *Extension:Math*. LIC: GPL. SWHID: `<swh:1:dir:96973202c47de62339be4e20031446171c0e78eb;origin=https://github.com/wikimedia/mediawiki-extensions-Math;visit=swh:1:snp:5615bfab6c6038fb784d9b86419783bfbeefd826;anchor=swh:1:rev:73f2bd4e7008960b3e97e1de16f34e6a5816b8ac;path=/src/TeXVC/>`.
- [6] P. Benner, K. Lund, and J. Saak. “Towards a Benchmark Framework for Model Order Reduction in the Mathematical Research Data Initiative (MaRDI)”. In: *CoRR* abs/2307.00137 (2023). DOI: 10.48550/arXiv.2307.00137. arXiv: 2307.00137.
- [7] A. Greiner-Petter, M. Schubotz, C. Breitingner, P. Scharpf, A. Aizawa, and B. Gipp. “Do the Math: Making Mathematics in Wikipedia Computable”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 45.4 (Apr. 2023), pages 4384–4395. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2022.3195261.
- [8] [Software excerpt] M. Schubotz, A. Ortiz, J. Stegmüller, L. Pusch, E. Ferrer, and D. Nolte, *Docker-compose repository for MaRDI 2023*. SWHID: `<swh:1:cnt:0a0d351f48352fa08626eb293ed856f036636caa;origin=https://github.com/MaRDI4NFDI/portal-compose;visit=swh:1:snp:476d7e7e164fe7f054feff57b4df4b98326af2e;anchor=swh:1:rev:2e1f77a314a041c736eda37853b63f54274d915;path=/docker-compose.yml;lines=24-508>`.
- [9] J. Stegmüller. “mhchemParser in PHP”. In: Zenodo, June 2023. DOI: 10.5281/zenodo.8189613.
- [10] M. Hintermüller. “MaRDI: Mathematical Research Data Initiative Proposal”. In: (May 2022). DOI: 10.5281/ZENODO.6552435.
- [11] R. Shigapov. *Knowledge graphs in BERD and in NFDI*. English. 2022. DOI: 10.5281/ZENODO.7373258.
- [12] J. Stegmüller and M. Schubotz. “Formula validation in MediaWiki with TexVC”. In: Zenodo, Nov. 2022. DOI: 10.5281/zenodo.7305299.
- [13] M. Hensel. “The mhchem Bundle”. In: *available via CTAN, eg, at the site www.dante.de/cgi-bin/ctan-index* (2021).

- [14] P. Scharpf, M. Schubotz, and B. Gipp. “Mathematics in Wikidata”. In: *Proceedings of the 2nd Wikidata Workshop (Wikidata 2021) co-located with the 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24, 2021*. Edited by L. Kaffee, S. Razniewski, and A. Hogan. Volume 2982. CEUR Workshop Proceedings. CEUR-WS.org, 2021.
- [15] M. Schubotz, A. Greiner-Petter, N. Meuschke, O. Teschke, and B. Gipp. “Mathematical Formulae in Wikimedia Projects 2020”. In: *JCDL '20: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, Virtual Event, China, August 1-5, 2020*. Edited by R. Huang, D. Wu, G. Marchionini, D. He, S. J. Cunningham, and P. Hansen. ACM, 2020, pages 447–448. DOI: 10.1145/3383583.3398557.
- [16] K. Hulek, F. Müller, M. Schubotz, and O. Teschke. “Mathematical Research Data – An Analysis Through zbMATH References”. In: *EMS Newsletter 2019–9.113 (Aug. 2019)*, pages 54–57. ISSN: 1027-488X. DOI: 10.4171/NEWS/113/14.
- [17] P. Scharpf, M. Schubotz, and B. Gipp. “Representing Mathematical Formulae in Content MathML Using Wikidata”. In: *Proceedings of the 3rd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL) 2018 Co-Located with the 41st International (ACM) (SIGIR) Conference on Research and Development in Information Retrieval (SIGIR) 2018*. Edited by P. Mayr, M. K. Chandrasekaran, and K. Jaidka. Volume 2132. CEUR Workshop Proceedings. CEUR-WS.org, 2018, pages 46–59.
- [18] M. Schubotz, P. Scharpf, K. Dudhat, Y. Nagar, F. Hamborg, and B. Gipp. “Introducing MathQA - a Math-Aware Question Answering System”. In: *Proceedings of the Annual International ACM/IEEE Joint Conference on Digital Libraries (JCDL), Workshop on Knowledge Discovery*. Fort Worth, USA, June 2018. DOI: 10.1108/idd-06-2018-0022.
- [19] F. Å. Nielsen, D. Mietchen, and E. Willighagen. “Scholia and scientometrics with Wikidata”. In: *Scientometrics 2017*. Nov. 2017, pages 237–259. DOI: 10.1007/978-3-319-70407-4_36.
- [20] M. Schubotz. *Augmenting Mathematical Formulae for More Effective Querying & Efficient Presentation*. Epubli Verlag, Berlin, 2017. ISBN: 978-3-7450-6208-3. DOI: 10.14279/depositonce-6034.
- [21] F. Müller and O. Teschke. “Full text formula search in zbMATH”. In: *Eur. Math. Soc. Newsl* 102 (2016), page 51.
- [22] M. Schubotz and A. P. Sexton. “A Smooth Transition to Modern mathoid-based Math Rendering in Wikipedia with Automatic Visual Regression Testing”. In: *Joint Proceedings of the FM4M, MathUI, and ThEdu Workshops, Doctoral Program, and Work in Progress at the Conference on Intelligent Computer Mathematics 2016 co-located with the 9th Conference on Intelligent Computer Mathematics (CICM 2016), Bialystok, Poland, July 25-29, 2016*. Edited by A. Kohlhase, P. Libbrecht, B. R. Miller, A. Naumowicz, W. Neuper, P. Quaresma, F. W. Tompa, and M. Suda. Volume 1785. CEUR Workshop Proceedings. CEUR-WS.org, 2016, pages 132–145.

- [23] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 'T Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. "The FAIR Guiding Principles for Scientific Data Management and Stewardship". In: *Scientific Data* 3.1 (Mar. 2016), page 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18.
- [24] H. S. Cohl, M. Schubotz, M. A. McClain, B. V. Saunders, C. Y. Zou, A. S. Mohammed, and A. A. Danoff. "Growing the Digital Repository of Mathematical Formulae with Generic Sources". In: *Intelligent Computer Mathematics, Lecture Notes in Artificial Intelligence 9150*. Edited by M. Kerber, J. Carette, C. Kaliszyk, F. Rabe, and V. Sorge. Volume 9150. LNCS. Springer, 2015, pages 280–287. DOI: 10.1007/978-3-319-20615-8_18.
- [25] H. S. Cohl, M. A. McClain, B. V. Saunders, M. Schubotz, and J. C. Williams. "Digital Repository of Mathematical Formulae". In: *Intelligent Computer Mathematics - International Conference, CICM 2014, Coimbra, Portugal, July 7-11, 2014. Proceedings*. Edited by S. M. Watt, J. H. Davenport, A. P. Sexton, P. Sojka, and J. Urban. Volume 8543. Lecture Notes in Computer Science. Springer, 2014, pages 419–422. DOI: 10.1007/978-3-319-08434-3_30.
- [26] M. Schubotz and G. Wicke. "Mathoid: Robust, scalable, fast and accessible math rendering for wikipedia". In: *Intelligent Computer Mathematics: International Conference, CICM 2014, Coimbra, Portugal, July 7-11, 2014. Proceedings*. Springer, 2014, pages 224–235.
- [27] Schubotz and Obrovac. *LaTeX formulae from English Wikipedia*. Version 1.0.0. Zenodo, Nov. 2014. DOI: 10.5281/zenodo.7494266.
- [28] M. Schubotz. "Making Math Searchable in Wikipedia". In: arXiv:1304.5475 (Apr. 2013). DOI: 10.48550/arXiv.1304.5475. arXiv: 1304.5475 [cs].
- [29] D. Ginev, H. Stammerjohanns, B. Miller, and M. Kohlhase. "The LaTeXXML Daemon: Editable Math on the Collaborative Web". en. In: 6824. Proceedings of the Conference on Intelligent Computer Mathematics, Bertinoro, IT, Dec. 2011. DOI: https://doi.org/10.1007/978-3-642-22673-1_25.