Die Transformation von zbMATH zu einer offenen Plattform für die Mathematik

Klaus Hulek und Olaf Teschke

Mit Beginn des Jahres 2021 wird die bisher kostenpflichtige Datenbank zbMATH in eine Open Access-Plattform übergeführt werden. Damit steht dieser Dienst weltweit allen Interessierten kostenfrei zur Verfügung. Diese Änderung des Geschäftsmodells wird es in Zukunft ermöglichen, die meisten Daten von zbMATH für Zwecke der Forschung und der Verknüpfung mit anderen nicht-kommerziellen Diensten unter den Bedingungen einer CC-BY-SA Lizenz frei zu nutzen. Im Folgenden beschreiben wir die Herausforderungen und die Vision, die sich für die Überführung von zbMATH in eine offene Plattform ergeben.

In den vergangenen Jahrzehnten ist die Mathematik von Beginn an Triebkraft und Nutznießer der Digitalisierung gewesen. Computer haben früh Einzug in mathematische Institute gehalten und dort Berechnungen und Resultate ermöglicht, die zuvor unerreichbar waren. Die mathematische Forschungsliteratur gehörte zu den ersten Korpora, die elektronisch verfügbar gemacht wurden – sowohl durch neue elektronische Journale als auch umfangreiche Retrodigitalisierungsanstrengungen. Review-Datenbanken ermöglichen seit langem einen Überblick über den Stand der Forschung, und umfangreiche Sammlungen mathematischer Objekte wie Zahlenfolgen oder spezielle Funktionen sind seit mehr als zwanzig Jahren elektronisch verfügbar. Plattformen werden zum kollaborativen Arbeiten und zur schnellen Verbreitung und Diskussion von Ergebnissen weit- hin genutzt.

Gleichzeitig ist die Mathematik wie kaum eine andere Wissenschaft auf Verlässlichkeit und Vollständigkeit ihres Wissenskorpus' angewiesen. Die gegenseitige Abhängigkeit mathematischer Konzepte und Sätze führt dazu, dass Fehler im Wissensbestand sich dramatisch fortpflanzen könnten. Regelmäßig sind auch zeitlich weit zurückliegende Resultate von hoher Relevanz; die Halbwertszeit mathematischen Wissens ist erheblich und liegt deutlich über der der meisten anderen Fachgebiete [1, 2, 4]. Die wachsende Quantität, aber vor allem auch Diversität der Ergebnisse bedeutet eine große Herausforderung für die Sicherung dieser Standards. Darüber hinaus ist klar, dass eine enge Verknüpfung der verschiedenen digitalen Ressourcen von hohem Wert für die mathematische Forschung ist – sie würde sowohl ein deutlich effizienteres Arbeiten ermöglichen als auch die Chancen für völlig neue Erkenntnisse eröffnen. Viele wertvolle Ressourcen sind historisch jedoch monolithisch gewachsen und durch ihr Geschäftsmodell eher abgeschlossen angelegt, so dass viele wünschenswerte Querverbindungen noch nicht existieren und schwierig zu realisieren sind.

zbMATH¹ (früher Zentralblatt für Mathematik) stellt seit langer Zeit einen umfangreichen Service mit Informationen über mathematische Publikationen, Autoren, Publikationen, Referenzen und Software dar. Die Nutzbarkeit dieser Daten für vernetzte Dienste oder zu Forschungszwecken war jedoch durch das Lizenzmodell bisher stark

eingeschränkt. Einer Empfehlung der Evaluierung von FIZ Karlsruhe 2017 folgend, hat die Gemeinsame Wissenschaftskonferenz des Bundes und der Länder 2019 entschieden, die Transformation von zbMATH zu einer Open-Access-Plattform zu unterstützen. Die Herausgeber von zbMATH – das FIZ Karlsruhe, die European Mathematical Society und die Heidelberger Akademie der Wissenschaften – haben dies begrüßt und verankern dieses Modell gegenwärtig in einem neuen Herausgebervertrag. Damit wird die Datenbank zbMATH ab 2021 frei zugänglich sein; darüber hinaus sollen über offene Schnittstellen Daten zugänglich gemacht werden. Durch die dadurch ermöglichte Vernetzung von Informationen soll die mathematische Community unterstützt werden. Wir wollen hier die nächsten Schritte und mögliche langfristige Perspektiven skizzieren.

Status quo und Bedarf

Mathematische Information ist traditionell in erster Linie in Publikationen abgelegt. Darüber hinaus sind heute aber mathematische Daten auch in Form von mathematischer Software und Forschungsdaten, Community-Plattformen, nicht-textuellen Materialien oder Bildungsinhalten in vielfältiger Form verfügbar. Sie zeichnen sich durch einen hohen Formalisierungsgrad und Langlebigkeit aus, wobei die Verknüpfung der Informationen eine hohe Qualität der Erschließung erfordert, nicht zuletzt bedingt durch die mathematischen Formeln. Noch immer sind Publikationen das wichtigste Medium zum Austausch und zur Sicherung wissenschaftlicher Ergebnisse. Mit zbMATH und MathSciNet² gibt es zwei unabhängige, hochwertige Nachweissysteme und damit exzellente Zugriffsmöglichkeiten auf die publizierte Literatur. Ein wachsender Anteil der mathematischen Publikationen steht heute über das arXiv³ und die European Digital Mathematics Library EuDML⁴ Open Access zur Verfügung.

Die klassischen Publikationen werden durch eine steigende Menge an Forschungsdaten ergänzt, die allerdings bisher noch weitgehend isoliert zur Verfügung stehen und kaum erschlossen sind. Dennoch werden sie von vielen Mathematikerinnen und Mathematikern erzeugt und genutzt.

Beispiele sind die Zahlenfolgen in der On-Line Encyclopedia of Integer Sequences⁵, Funktionen in der NIST Digital Library of Mathematical Functions⁶, die Calabi–Yau Daten⁷ oder die *L*-Functions and Modular Forms Database⁸. Das Gebiet der mathematischen Modellierung und Simulation benötigt und erzeugt Terabytes von Daten aus numerischen Berechnungen.

Community-Plattformen spielen für die Generierung und den Austausch von Informationen eine zunehmend wichtige Rolle. Ein exzellentes Beispiel ist das Diskussionsforum MathOverflow⁹. Besonders für den wissenschaftlichen Nachwuchs ist dies eine nicht zu unterschätzende Informationsquelle. Ein weiteres Beispiel sind die PolyMath-Projekte¹⁰, in denen mathematische Fragestellungen in kollaborativer Form erforscht werden. Wikipedia¹¹ oder Imaginary¹² erreichen ein breites Publikum jenseits der Fachwissenschaft und tragen essentiell zur Verbreitung mathematischen Wissens, der Popularisierung von Mathematik und wissenschaftlicher Bildung bei.

Gerade die letzten beiden Beispiele illustrieren ebenfalls gut die Stärke von offenen Plattformen – kollaborativ können Informationen verschiedenster Natur und Schwierigkeit integriert, miteinander verknüpft, und der breiten Öffentlichkeit zugänglich gemacht werden. Überhaupt hat sich die mathematische Community von Beginn an für offenen Zugang zum Wissen engagiert, etwa durch eine umfangreiche Nutzung des arXiv, die frühzeitige Gründung von offenen Zeitschriften oder Initiativen wie TheCostofKnowledge¹³. Nach wie vor hat der offene Zugang zu mathematischem Wissen einen sehr hohen Stellenwert in der Community, wie die Umfrage [13] zeigt – als wichtiger wird nur (und zwar mit deutlichem Abstand) die Qualitätssicherung angesehen. Hier spielen auch die Erfahrungen aus dem Wandel der Publikationslandschaft eine Rolle, wie etwa Qualitätsprobleme durch kommerzielle Fehlanreize; vor allem aber kommt hier der überragende Stellenwert gesicherter Resultate für darauf aufbauende mathematische Forschung zum Ausdruck.

Zur Erfüllung dieser Bedürfnisse wird in Zukunft ein möglichst umfangreicher und freier Zugriff auf die unterschiedlichen, qualitätsgesicherten, verknüpften und erschlossenen Ressourcen unabdingbare Grundlage mathematischer Forschung sein. Eine umfassende Vision einer solchen Infrastruktur wurde von der International Mathematical Union (IMU) als Global Digital Mathematics Library [7] formuliert. Ein solcher Service würde Mathematikerinnen und Mathematiker bei der Suche, Einordnung und Bewertung dieser Informationen unterstützen und ermöglichen, Zusammenhänge zu erkennen und zu analysieren. Dabei muss auch die besondere Natur mathematischer Information berücksichtigt werden: die Mathematik hat in Form von mathematischen Formeln ihre eigene Sprache entwickelt, die in sehr kompakter Form eine hohe Präzision der Aussage sichert. Häufig findet man die Informationen, die in Formeln enthalten sind, in keiner anderen Form. Dies gilt für alle mathematikrelevanten Daten, in besonderem Maß jedoch für Forschungsdaten, die wesentlich formelbasiert sind und somit eine Formelsuche zur Erschließung benötigen.

Was behindert derzeit die Verwirklichung?

Entsprechend stellt sich als wichtige Zukunftsaufgabe, diese beschriebenen Ressourcen miteinander zu verbinden, inhaltlich zu erschließen und eine kohärente, umfassende, offene und nachhaltige Plattform aufzubauen. Diese Plattform muss Informationen verschiedenster Niveaus beinhalten um einen umfassenden Wissenstransfer zu gewährleisten. Durch einen standardisierten Nachweis von Forschungsdaten wird deren Reproduzierbarkeit erhöht, was der Qualitätssicherung von Publikationen dient. Zudem wird eine systematische Zitierbarkeit von Forschungsdaten ermöglicht, dies wiederum sichert die Anerkennung der darin enthaltenen Forschungsleistung. Es ist zudem essenziell, hochgradig formalisierte mathematische Inhalte wie Formeln zu erschließen. Auch wenn diese Entwicklungen noch ganz am Anfang stehen, lassen sie das große Potential moderner Informationstechnologie für die Mathematik erkennen.

Solch ein freier Zugriff auf miteinander vernetzte und annotierte Ressourcen wird in Zukunft die unverzichtbare Basis mathematischer Forschung sein. Dadurch erhalten Mathematikerinnen und Mathematiker Hilfe bei der Suche und Bewertung dieser Informationen, Zusammenhänge werden verdeutlicht und können analysiert werden. Wesentliche Voraussetzung bleibt dabei die hohe Qualität aller verfügbaren Informationen. Ein großer Teil der oben beschriebenen vorhandenen Informationen steht aktuell jedoch nicht frei zur Verfügung oder ist nicht oder nur gering erschlossen. Dadurch sind Informationen teilweise nicht auffindbar oder gehen auf Grund der fehlenden Vernetzung verloren.

Funktionen und Inhalte einer offenen Plattform müssen weit über die derzeitigen Möglichkeiten abgeschlossener Systeme wie zbMATH, MathSciNet, Google Scholar¹⁴, Researchgate¹⁵ oder WolframAlpha¹⁶ hinausgehen. Aktuell ist ein großer Teil der oben beschriebenen vorhandenen Informationen weder frei verfügbar noch ausreichend erschlossen und dadurch teilweise nicht auffindbar, viele Informationen gehen auf Grund der fehlenden Vernetzung verloren. Durch die Verlinkung wird das vorhandene Wissen gebündelt. Jeder einzelne Informationsservice liefert einzelne Bausteine an Informationen, die, korrekt analysiert, zusammengeführt zu neuen Erkenntnissen führen werden. Die Abbildung auf der folgenden Seite zeigt eine Auswahl möglicher unterschiedlicher Informationsquellen auf, die jeweils innerhalb der einzelnen Segmente und zusätzlich untereinander vernetzt werden müssen.

Diese Verknüpfungen sind nur der Anfang: durch offene Schnittstellen erhält die Community die Möglichkeit, vorhandene Daten für ihre eigene Forschung zu nutzen und anschließend Forschungsergebnisse wieder auffindbar einzubringen. Dies können neben Publikationen oder Forschungsdaten auch Tools zur Nutzung oder Analyse der Daten sein. Damit leistet die Community selbst innovative Beiträge für die Entwicklung eines umfassenden mathematischen Forschungssystems. Alle bisherige Erfahrung zeigt, dass durch eine solche Öffnung erhebliche und jetzt oft noch unvorhersagbare Innovationspotentiale erschlossen werden.

Vernetzung mathematischer Informationen

Der Entwicklung eines solchen Netzwerks standen bisher verschiedene Hindernisse im Weg. Zum einen stehen historisch gewachsene Geschäftsmodelle (die früher oft auch nötig waren, um die Ressourcen für die Qualitätssicherung und Verbreitung wissenschaftlicher Information bereitzustellen) im Widerspruch zu offenen Lösungen. Weiterhin sind eine Reihe von Techniken, die für die automatisierte Verknüpfung, Indexierung und das Wissensmanagement in der digitalen Mathematik wünschenswert sind, noch nicht ausgereift oder in der idealen wünschbaren Mächtigkeit vorhanden (vom Formel-OCR bis zu automatisierten Verifikationstools).

Viele Bereiche eines solchen vernetzten, kollaborativen Systems werden ein umfangreiches Engagement erfordern, ähnlich dem, welches die Community seit Jahrzehnten für das aufwändige Peer Review Verfahren von Artikeln leistet. Das Engagement von Individuen oder Gruppen ist sicher die knappste und wertvollste Ressource. Es wird nur dann aktiviert und eingebracht werden, wenn damit offensichtliche Vorteile für die Forschung verbunden sind.

Was sind die nächsten Schritte für zbMATH Open?

Dank der Unterstützung des Bundes und der Länder wird zbMATH ab 2021 seine Dienste Open Access anbieten können und damit die bisherigen Einschränkungen aufgrund seines traditionellen Geschäftsmodells überwinden. Die meisten Daten werden über eine CC-BY-SA-Lizenz offen zugänglich sein (mit Ausnahme von Inhalten Dritter wie die Abstracts der Autoren, die mit einer anderen Lizenz verbunden sind).

Diese offenen Daten werden für die Forschung und Entwicklung frei verfügbar sein. Es gibt bereits viele Projekte (z.B. für die Geschichte der Mathematik), die zbMATH-Daten verwenden. Solche Projekte und daraus resultierende Dienste werden durch offene Daten wesentlich erleichtert. Ein weiteres Beispiel ist die Verwendung von zbMATH-Daten im ISC-Projekt „Gender Gap In Science“¹⁷, das von der IMU koordiniert wird. Ein Teil der Projektergebnisse ist die Visualisierungsplattform Gender Publication Gap¹⁸. Die nachhaltige Verfügbarkeit solcher Ergebnisse erfordert eine

langfristige Wartung dieser Projekte. Eine Schlüsselaufgabe wird es sein, wichtige Projektergebnisse unter Einbeziehung der zbMATH-Daten zu identifizieren, sie möglichst in das zbMATH Open-Netzwerk zu integrieren und nachhaltig verfügbar zu machen.

Publikationen sind nach wie vor der entscheidende Kern der mathematischen Forschung. In den letzten Jahrzehnten ist ein wachsender Anteil offen zugänglich geworden (siehe [15] für einen qualitativen und quantitativen Überblick). Viele Entwicklungen beschleunigen dies: der wachsende Anteil der arXiv-Preprints [12], die Einrichtung von moving walls, offene Retrodigitalisierungsprojekte, transformative Verlagsverträge und Initiativen wie Plan S¹⁹. Das technische Niveau (wie auch die verwendete Lizenz) der Ergebnisse ist jedoch sehr unterschiedlich und reicht von einfachen Scans bis hin zu verfügbaren $\text{\LaTeX}$ -Quellen, die bereits eine Volltext-Formelsuche ermöglichen [11]. Die Integration dieser verschiedenen Ressourcen mit Hilfe offener zbMATH-Daten wird die derzeitige Situation erheblich verbessern und einen Rahmen schaffen, der volltextbezogene Dienste für die Indexierung und Suche verfügbar macht.

Wie in [9] dargelegt, werden mathematische Forschungsdaten zu einem Thema von extrem wachsender Bedeutung werden. Diese Daten sind jedoch von Natur aus sehr vielfältig, und es ist eine große Herausforderung, sie gemäß den FAIR-Prinzipien (*Findability, Accessibility, Interoperability, Reusability*²⁰) nachhaltig verfügbar zu machen. Auch hier wird ihre Integration in ein vernetztes System mit offenen Publikationsdaten einen großen Schritt nach vorn bedeuten. Auffindbarkeit und Sichtbarkeit werden dadurch verbessert und Mechanismen zur Qualitätssicherung ermöglicht. Vor einigen Jahren hat zbMATH einen solcher Ansatz für mathematische Software begonnen und daraus den freien Dienst swMATH²¹ entwickelt. Ein vergleichbarer Dienst für die Gesamtheit der mathematischen Forschungsdaten ist eine Aufgabe, die weit über den zbMATH Open-Rahmen hinausgeht. Daher werden wir zunächst bereits technisch reife Ressourcen in zbMATH einbinden und diese Ergebnisse für weitere vernetzte Aktivitäten verfügbar machen, wie sie im derzeit gebildeten MaRDI-Konsortium (Mathematical Research Data Initiative²²) im Rahmen der nationalen Forschungsdateninfrastruktur (NFDI²³) vorgesehen sind.

Wie im Fall der mathematischen Software wird das Engagement zur Erstellung, Entwicklung und Pflege von Forschungsdatensammlungen oft nicht ausreichend gewürdigt. Eine vernetzte Plattform würde ihren Urhebern zudem sehr helfen, Anerkennung für Beiträge zu erhalten, die nicht direkt mit Publikationen in Zusammenhang stehen. Wenn die breite Verwendung in einem vernetzten System sichtbar wird, wird es für die Mathematikerinnen und Mathematiker wie auch für ihre Institutionen viel einfacher sein, die entsprechende Anerkennung für ihre Bemühungen zu erlangen und hoffentlich ausreichende Mittel für Dienstleistungen zu erhalten, die gegenwärtig oft auf freiwilliger und daher nicht immer nachhaltiger Basis erbracht werden.

Ebenso hat sich die Rolle von nicht-textuellem Material wie Videos erheblich verändert – die mathematische Kommunikation unter den Bedingungen des Kontaktbeschränkungen haben dies besonders deutlich gemacht. Auch wenn die Bedrohung durch SARS-COV-2 zurückgegangen sein wird, werden die Erfahrungen der letzten Monate sehr wahrscheinlich einen nachhaltigen Einfluss haben. Es erscheint unwahrscheinlich (und auch nicht wünschenswert, da etwa die viel größere Bedrohung durch die globale Erwärmung nach wie vor besteht), schnell wieder zum alten, dicht gedrängten Konferenzkalender zurückzukehren. Stattdessen werden wir sicher die digitale Kommunikation in der Mathematik weit stärker nutzen. Die so geschaffenen Ressourcen erfordern jedoch die gleiche Integration, Vernetzung, nachhaltige Bewahrung und Indexierung wie Forschungsdaten.

Eine ähnliche Arbeit muss für die enorme Menge an mathematischem Wissen geleistet werden, das von den bereits erwähnten digitalen mathematischen Communities zusammengetragen wird. Zwar wurden bereits erste Schritte unternommen, z. B. die Verknüpfung von Wikidata-Informationen in zbMATH-Autorenprofilen mit dem Vorteil der Erstellung von Zeitleisten oder die Verknüpfung von Publikationen mit MathOverflow-Diskussionen [10], doch wird noch viel mehr möglich sein, wenn die Kreativität der Mathematikerinnen und Mathematiker mit geeigneten Werkzeugen unterstützt wird. Es liegt an uns allen, die dafür nötigen Werkzeuge zu definieren, um die offenen Daten zu nutzen und einzubinden – Ihre Vorschläge und Ihr Feedback an editor@zbmath.org sind willkommen!

Anmerkungen

1. zbmath.org
2. mathscinet.ams.org
3. arxiv.org
4. eudml.org
5. OEIS, www.oeis.org
6. DLMF, dlmf.nist.gov
7. hep.itp.tuwien.ac.at/~kreuzer/CY/
8. LMFDB, www.lmfdb.org
9. mathoverflow.net
10. polymathprojects.org

11. www.wikipedia.org
12. imaginary.org
13. thecostofknowledge.com
14. scholar.google.de
15. www.researchgate.net
16. www.wolframalpha.com
17. gender-gap-in-science.org
18. gender-publication-gap.f4.htw-berlin.de
19. www.scienceeurope.org/our-priorities/open-access
20. www.go-fair.org/fair-principles/
21. swmath.org, siehe auch [3, 5]
22. wias-berlin.de/mardi/
23. www.dfg.de/foerderung/programme/nfdi/

Literatur

- [1] A. Bannister, O. Teschke, An update on time lag in mathematical references, preprint relevance, and subject specifics. *Eur. Math. Soc. Newsl.* 106, 37–39 (2017).
- [2] A. Bannister, K. Hulek, O. Teschke, Das Zitationsverhalten in mathematischen Arbeiten. Einige Anmerkungen. *Mitt. Dtsch. Math.-Ver.* 25, No. 4, 208–214 (2017).
- [3] S. Bönisch, G.-M. Greuel, W. Sperber, The software information service swMATH – Release of the first online prototype. *Eur. Math. Soc. Newsl.* 87, 48–49 (2013).
- [4] T. Bouche, O. Teschke, K. Wojciechowski, Time lag in mathematical references. *Eur. Math. Soc. Newsl.* 86, 54–55 (2012).
- [5] H. Chrapary, W. Dalitz, W. Neun, W. Sperber, Design, concepts, and state of the art of the swMATH service. *Math. Comput. Sci.* 11, No. 3–4, 469–481 (2017).
- [6] K. Hulek et al., Mathematical research data – an analysis through zbMATH references. *Eur. Math. Soc. Newsl.* 113, 54–57 (2019).
- [7] NAS Committee on Planning a Global Library of the Mathematical Sciences. *Developing a 21st Century Global Library for Mathematics Research* (2014).
- [8] G.-M. Greuel, A. Matt, A. Mey, IMAGINARY – Mathematics communication for the 21st century. *Eur. Math. Soc. Newsl.* 92, 3–6 (2014).
- [9] K. Hulek, F. Müller, M. Schubotz, O. Teschke, Mathematical research data – An analysis through zbMATH references. *Eur. Math. Soc. Newsl.* 113, 54–57 (2019).
- [10] F. Müller, M. Schubotz, O. Teschke, References to research literature in QA forums – A case study of zbMATH links from MathOverflow. *Eur. Math. Soc. Newsl.* 114, 50–52 (2019).
- [11] F. Müller, O. Teschke, Full text formula search in zbMATH. *Eur. Math. Soc. Newsl.* 102, 51 (2016).
- [12] F. Müller, O. Teschke, Will all mathematics be on the arXiv (soon)? *Eur. Math. Soc. Newsl.* 99, 55–57 (2016).
- [13] C. Neylon, D. M. Roberts, M. C. Wilson, Mark, Results of a worldwide survey of mathematicians on journal reform. *Eur. Math. Soc. Newsl.* 103, 46–49 (2017).
- [14] D. H. J. Polymath et al., The “Bounded gaps between primes” Polymath project – A retrospective analysis. *Eur. Math. Soc. Newsl.* 94, 13–23 (2014).
- [15] O. Teschke, Green, gold, platinum, nickel: On the status of open access in mathematics. *Eur. Math. Soc. Newsl.* 110, 60–63 (2018).

Eine englische Version dieses Artikels ist zuerst im *Newsletter of the European Mathematical Society* 116, 44–47 (2020), erschienen. Die Mitteilungen danken für die Genehmigung zum Druck der deutschsprachigen Fassung.

Prof. Dr. Klaus Hulek, Gottfried Wilhelm Leibniz Universität Hannover,
Institut für Algebraische Geometrie, Welfengarten 1, 30167 Hannover
hulek@math.uni-hannover.de

Dr. Olaf Teschke, FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur,
Abteilung Mathematik, Franklinstraße 11, 10587 Berlin
olaf.teschke@fiz-karlsruhe.de