

Digitale Mathematische Bibliotheken, zbMATH und Linked Data

1. Referatedienste vor 150 Jahren und heute

Vor 150 Jahren wurde der erste deutsche Referatedienst der Mathematik, das Jahrbuch über die Fortschritte der Mathematik, ins Leben gerufen. Das Antrittsprogramm formulierte:

„Das Ziel, das uns vorschwebte, war einerseits: Demjenigen, der nicht in der Lage ist, alle auf dem umfangreichen Gebiete der Mathematik vorkommenden Erscheinungen selbständig zu verfolgen, ein Mittel zu geben, sich wenigstens einen allgemeinen Überblick zu verschaffen; andererseits: Dem gelehrten Forscher seine Arbeit bei Auffindung des bereits Bekannten zu erleichtern.“

Dem schmalen 1868er Band mit Kurzreferaten zu den 880 Arbeiten jenes Jahres stehen heute gewaltige Zahlen gegenüber – die Tagesstatistik von zbMATH (der Datenbank in der Nachfolge des Jahrbuchs) zeigt u.a. an:

3.828.247 Dokumente (3.322.334 MSC-erschlossen)

992.466 Personen (+Publikations- und Zitationsprofile)

6.363 Zeitschriften/Serien (+Publikations- und Zitationsprofile)

17.879 Software packages zitiert in 146.140 Dokumenten

160.809.572 Formeln indexiert

2.593.496 Volltextlinks, darunter 2.174.679 DOI, 164.517 arXiv, 170.314 EuDML, 23.396 EMIS, 17.978 Project Euclid, 25.425 Numdam...

26.371.061 Referenzen, davon 14.290.699 verlinkt zu 1.210.519 zbMATH-Dokumenten

2. Was benötigen Mathematiker(innen)?

Trotz der diesen Zahlen zugrundeliegenden effektiven technologischen Erschließung und Verknüpfung in einer schnell recherchierbaren Datenbank darf jedoch bezweifelt werden, dass das 1868 formulierte Ziel – das nach wie vor aktuell ist – gleichermaßen gut erreicht wird wie durch das Durchblättern jenes Bandes. Der Grund ist natürlich das inzwischen erheblich gewachsene Publikationsvolumen. Ein modernes Informationssystem muss dem Rechnung tragen; zentral ist hierbei

- Ein möglichst **umfangreicher** und **freier Zugriff** auf Ressourcen

- Hohe **Qualität** der Informationen
- Möglichkeit der **Verknüpfung** und **Annotation** dieser Ressourcen
- Hilfe bei der **Suche**, **Einordnung** und **Bewertung** der Informationen
- Unterstützung beim Erkennen von **Zusammenhängen**

In einer Welt, wo mathematische Ergebnisse längst weit jenseits reiner Publikationen in höchst diverser Form generiert werden (z.B. auch in Software, Forschungsdaten, Videos, mathematischen Modellen, Datenbanken von Objekten oder formalisierter Mathematik, Forschungsplattformen etc.) bedeutet dies, dass die Funktionen und Inhalte eines idealen Informationssystems weit über die derzeitigen Möglichkeiten abgeschlossener Systeme (wie Verlage, zbMATH, MathSciNet, Google Scholar, Researchgate, WolframAlpha...) hinausgehen müssen.

3. Linked Open Data und Global Digital Mathematics Library

Die wesentliche Struktur, um diesen Herausforderungen künftig zu begegnen, wurde 2007 von Tim Berners-Lee als Gegenentwurf zu den riesigen proprietären Datensammlungen von Monopolisten wie Google oder Facebook formuliert:

- „1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things.“

Für die Mathematik wurde dieser Ansatz im 2014 erschienenen NAS-Report „Developing a 21st Century Global Library for Mathematics Research“ konkretisiert und als Ziel formuliert:

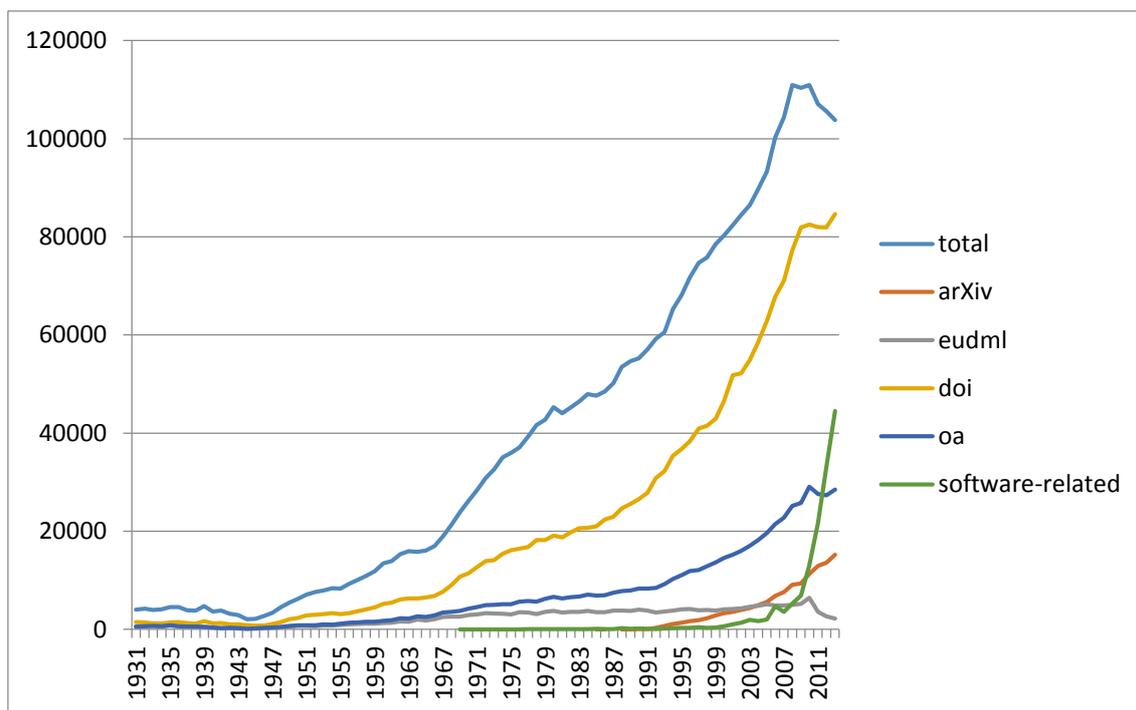
„Beyond digitization, more value is expected by creating connected information resources which are of greater value than the sum of its contributing parts.“

Dies mündete in die Gründung einer Arbeitsgruppe der IMU zur Verwirklichung einer Global Digital Mathematics Library (GDML) mit dem Mission Statement:

„The Mission of the GDML is to construct, as a global public good, an effective knowledge base encompassing the results of the world's mathematics through collaborations deploying both present and new technology, and to foster a supporting community.“

Die Verwirklichung der GDML beruht auf der Durchführung verschiedener Digitalisierungsschritte auf dem geeigneten Korpus. Anfang 2018 wurde bei zbMATH durch ein umfangreiches bibliographisches Matching mit den Mathematical Reviews die bisher genaueste Korpusabschätzung erreicht. Die Schnittmenge beträgt (bei einer Unsicherheit von ca. 5%) ca. 60% von zbMATH (bzw. 66% von MR), dies ergibt für den Korpus der Mathematik zu diesem Zeitpunkt seit 1868 ca. 190.000 Bücher mit durchschnittlich 350 Seiten und ca. 3,9 Mio. Artikel mit durchschnittlich 14,5 Seiten, also ca. 120 Mio. Seiten Mathematik, derzeit jeweils ca. zur Hälfte in Büchern und Artikeln (wobei sich dieser Anteil in Richtung der Artikel verschiebt).

Davon liegen ca. 80% der Dokumente und ca. 60% der Seiten als Scan oder sogar PDF vor, jedoch sind nur ca. 10% der Dokumente / 20% der Seiten offen verfügbar. Für weniger als 5% der Dokumente / 3% der Seiten stehen offene LaTeX/XML-Quellen zur Verfügung, die eine Konversion in (Presentation-)MathML erlauben. Eine echte Konversion in semantisches MathML – letztlich Grundlage für die avisierte inhaltliche Verknüpfung – gibt es nur für eine verschwindend geringe Menge; noch kleiner ist die Menge an formalisierter Mathematik. Dennoch sind die Fortschritte deutlich – das folgende Diagramm veranschaulicht die Entwicklung:



4. Beispiele und Ansätze für Linked-Data in zbMATH

- zbMATH hat schon relativ früh Teile der Datenbank zur Ermöglichung einer besseren Vernetzung im oben beschriebenen Sinn verfügbar gemacht.

- Die 2010 eingeführte neue Version der Mathematical Subject Klassifikation wurde umgehend in SKOS transformiert und als Linked Open Data veröffentlicht.
- Seit 2012 gibt es eine offene Schnittstelle zum Erzeugen von zbMATH-Identifiern und damit zur Verknüpfung von Dokumenten.
- Zusätzlich wird freien digitalen Bibliotheken auch eine Schnittstelle zur Identifizierung von Autoren in zbMATH angeboten.
- Die Dokumentenverknüpfung mit dem arXiv und der Einsatz von Open Source Software zur Formelindexierung erlaubt seit 2016 eine Volltextformelsuche.
- Die Dokumentenschnittstelle wurde 2017 zu einem Interface zur gegenseitigen Verknüpfung mit Diskussionen im MathOverflow-Forum ausgebaut.

Dies deckt natürlich noch längst nicht alle Bedürfnisse ab; in den kommenden Jahren steht etwa die Einführung von entsprechenden Angeboten für Softwaredaten, Formelaggregationen, Mathematischen Modelle oder Knowledge graphs an. Langfristig können über verfügbar gemachte Daten und geeignete Schnittstellen

- Forschungsdaten und nichttextuelle Materialien erschlossen und integriert werden.
- die Community innovative Beiträge zur Weiterentwicklung leisten.
- Qualitativ hochwertigen Informationen für die Forschung in geeigneter Weise frei und nachnutzbar verfügbar gemacht werden.

Literatur

Berners-Lee, T. (2006 ff.). Linked Data.

<https://www.w3.org/DesignIssues/LinkedData.html>

MSC2010 (Mathematics Subject Classification) (2010 ff.) im SKOS Format.

<http://msc2010.org/resources/MSC/2010/info/>

National Research Council (2014). *Developing a 21st Century Global Library for Mathematics Research*. Washington, DC: The National Academies Press.

doi:10.17226/18619.

Ohrtmann, C., Müller, F. (1868). *Vorrede*, Jahrbuch über die Fortschritte der Mathematik, Band I, Berlin: Verlag Reimer.

zbMATH (1990 ff.), Facts & Figures. https://zbmath.org/about/#id_4