



Das *Electronic Research Archive for Mathematics* an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen

von Katharina Habermann

David Hilbert schrieb: „Fruchtbaren Begriffsbildungen und Schlußweisen wollen wir, wo immer nur die geringste Aussicht sich bietet, sorgfältig nachspüren und sie pflegen, stützen und gebrauchsfähig machen.“ (siehe [6]). Um Mathematikerinnen und Mathematikern die Spurensuche zu ermöglichen, sind Bibliotheken gefragt und gefordert, ihnen diese Spuren zu sichern und „gebrauchsfähig“ zu machen. Bei dieser Spurensicherung übt der Gedanke, alles mathematische Wissen in einer digitalen Bibliothek zusammenzuführen, mit Verknüpfungen zu versehen und auf diese Weise in einer allumfassenden Sammlung zu verschmelzen, eine unglaubliche Faszination aus. Mit der rasanten Technologieentwicklung im Medienbereich scheint die Idee einer Bibliotheca Alexandrina für die Mathematik prinzipiell realisierbar zu sein und hat daher neuen Auftrieb erhalten.

Einleitung

Die rasant wachsende Bedeutung des Internet für die wissenschaftliche Information und Kommunikation sowie die technische Entwicklung neuer Mediensysteme hat die Rahmenbedingungen wissenschaftlichen Publizierens erheblich verändert. Elektronisches Publizieren – einerseits von Preprints, andererseits parallel zu gedruckten Zeitschriftenartikeln und Büchern – ist im Wissenschaftsbetrieb heute Standard.

Wenn aktuelle wissenschaftliche Publikationen und andere Quellen mathematischer Fachinformation in elektronischer Form vorliegen und direkt vom Arbeitsplatz aus über den PC genutzt werden können, stellt sich die Frage, ob nicht auch die ausschließlich gedruckt vorliegende Forschungsliteratur im Nachhinein, d.h. retrospektiv, digitalisiert werden kann. So könnte man ein Informationssystem schaffen, welches durchgängig elektronischen Zugang zu verschiedensten Arten von Dokumenten und somit ein Arbeiten ohne Medienbruch ermöglicht.

Retrospektive Digitalisierung als Konvertierung ausschließlich gedruckt vorliegender Texte in eine maschinenlesbare Form ist auf verschiedene Weisen realisierbar. Die Digitalisierung kann sowohl manuell durch Abtippen als auch mit Hilfe von Scannern erfolgen. Das Abschreiben und Korrigieren ist da-

bei der teureren und aufwändigere Weg. Eine Variante ist das Abtippen durch Freiwillige, wie es im Gutenbergprojekt [16] geschieht. Als Mathematiker wünscht man sich dann auch noch, dass das Ergebnis im \TeX -Format vorliegt. Es sind tatsächlich mathematische Texte auf diesem Wege digitalisiert worden. Als Beispiel seien die Arbeiten Bernhard Riemanns genannt, die von David R. Wilkins bearbeitet worden sind und in der *Electronic Library of Mathematics* auf dem EMIS-Server [14] zur Verfügung stehen. Eine zweite Variante des Abtippens ist das sogenannte Double-Keying-Verfahren. Dabei wird der Text von zwei Personen unabhängig voneinander abgeschrieben, und ein Programm überprüft dann die Übereinstimmung beider Versionen. Bei Nichtübereinstimmung muss der Text einer manuellen Korrektur unterzogen werden.

Die heute verbreitetste Umsetzung der Digitalisierung ist das Einscannen der Texte. Dabei reichen die technischen Möglichkeiten von der Erstellung einzelner Bilddateien (Images) bis zur inzwischen meist üblichen und von Nutzern geforderten Volltext-Erfassung der Dokumente. Bei einer Volltext-Erfassung gescannter Images wird die bildhafte Darstellung zusätzlich durch Texterkennungssoftware (OCR) analysiert und mit einer durchsuchbaren Textdatei hinterlegt.

Trotzdem gibt es zahlreiche Archive, die retrospektiv digitalisiertes Material ausschließlich als Sammlung von Image-Dateien bereitstellen. Die Gründe dafür können vielfältig sein, auch wenn eine Sammlung von Image-Dateien nur sehr eingeschränkte Suchmöglichkeiten gestattet. So werden viele Materialien als Images digitalisiert, weil man das Original-Dokument „zeigen“ will. Dies ist z. B. bei Handschriften, alten Drucken (Drucke vor 1800) oder Bildtafeln der Fall. Außerdem gibt es Materialien, bei denen Texterkennungssoftware nicht (oder noch nicht) zufriedenstellend einsetzbar ist. Dazu gehören beispielsweise verschmutzte Vorlagen, Texte mit uneinheitlichem Schriftsatz oder Texte in Frakturschrift. Außerdem ist der Einsatz von Texterkennungssoftware mit erheblichen Kosten verbunden. Um trotzdem bessere Suchmöglichkeiten anbieten zu können, werden Dokumente in digitalen Archiven strukturiert erfasst und z. B. mindestens das Inhaltsverzeichnis suchbar gestaltet. Die Strukturbeschreibung von Zeitschriften und Zeitschriftenartikeln muss außerdem eine Kennzeichnung bzw. Markierung von Elementen wie Jahrgang, Heft-Nummer, Seitenangaben, Titel, Autor usw. enthalten.

Mittlerweile ist eine Vielzahl ursprünglich nur gedruckt erschienener Texte retrospektiv digitalisiert und im Internet zur Verfügung gestellt worden. Bibliotheken und Archive bedienen sich der retrospektiven Digitalisierung einerseits, um durch Zerfall gefährdete Werke für die Nachwelt zu erhalten. Andererseits wird auf diese Weise eine breitere Nutzung der beeindruckenden Ressourcen wissenschaftlicher Bibliotheken ermöglicht, da seltene Werke so weltweit zugänglich sind. Insgesamt setzte mit dem Beginn der retrospektiven Digitalisierung der in Bibliotheken gesammelten Texte im letzten Jahrzehnt des 20. Jahrhunderts eine internationale Entwicklung ein, die den Zugriff auf eine immer größer werdende Zahl dieser Quellen auf elektronische Weise ermöglicht.

Bevorzugte Digitalisierungsobjekte sind vor allem Materialien mit Erscheinungsjahr bis einschließlich 1900, da hier geringere Hindernisse durch urheber- und vertriebsrechtliche Probleme auftreten. Ein Spezifikum der Mathematik, das eine retrospektive Digitalisierung von Altbeständen zusätzlich attraktiv macht, ist die unbegrenzte Gültigkeit mathematischer Forschungsergebnisse sowie die ausgeprägte Abhängigkeit mathematischer Forschung von zuvor erzielten Resultaten und damit von auch weit zurückliegenden mathematischen Veröffentlichungen. Auch wenn die mangelnde Verfügbarkeit historischer mathematischer Literatur als eine der wichtigsten Motivationen für deren Digitalisierung zu nennen ist, gibt es darüberhinaus zahlreiche weitere Gründe für die Notwendigkeit der Digitalisierung mathemati-

scher Texte. (Siehe dazu auch [11] im Kontext aktueller Initiativen.)

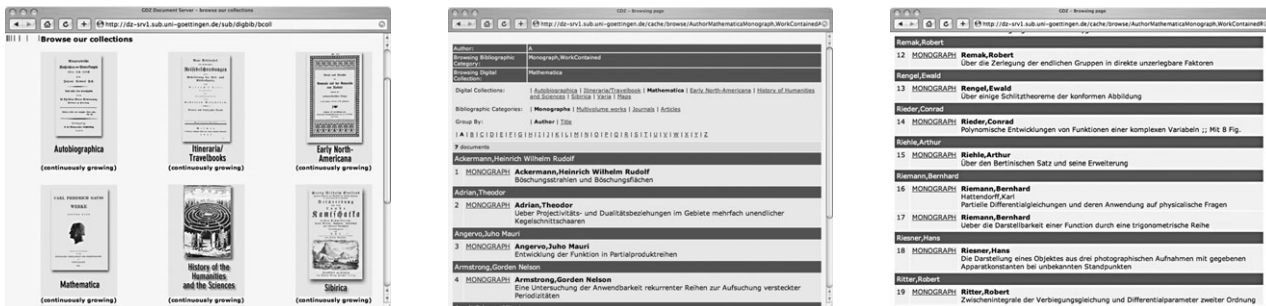
Die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB Göttingen) hat in einem der weltweit ersten Digitalisierungsprojekte das *Electronic Research Archive for Mathematics* (ERAM) geschaffen. Die Arbeiten zum „Aufbau eines Volltextarchivs forschungsrelevanter mathematischer Texte anhand des Jahrbuchs über die Fortschritte der Mathematik“ (auch kurz ERAM/Jahrbuch-Projekt) hatten im Juli 1997 mit finanzieller Förderung durch die Deutsche Forschungsgemeinschaft (DFG) begonnen. Das Projekt wurde in enger Kooperation mit den Herausgebern des *Zentralblatt MATH* realisiert.

Heute ist das Auslaufen der auch beeindruckend langen Projektförderung ein guter Anlass für einen Rückblick, einen Bericht über Erreichtes und natürlich einen Ausblick auf Zukünftiges.

Ausgangssituation

In den 1990er Jahren begann die mathematische Community, vielfältige Aktivitäten sowohl zur Entwicklung fachspezifischer Plattformen (z. B. MathNet [17] oder der EMIS-Server [13]) als auch im Bereich des elektronischen Publizierens (z. B. MPRESS [18] oder *Documenta Mathematica* [12]) ins Leben zu rufen [5]. Ziel war es, Mathematiker umfassend und direkt am Arbeitsplatz mit elektronischer Fachinformation – wie sie inzwischen alltägliches Arbeitsmittel ist – zu versorgen. Eine der damaligen Ideen war auch die Erfassung der Jahrbuch-Besprechungen, welche zunächst als eigenständige Aktivität geplant war. Dazu hatten die Chefredakteure der *Mathematical Reviews* Keith Dennis und des *Zentralblatt MATH* Bernd Wegner eine entsprechende Kooperation vereinbart und im Jahr 1996 (siehe [2]) interessierte Kollegen zur Mitarbeit als Experten aufgerufen. Zu diesem Zeitpunkt war jedoch die Finanzierung des Vorhabens überhaupt noch nicht geklärt und so dessen Realisierung unsicher.

Auch Bibliotheken entfalten seit dieser Zeit zahlreiche Aktivitäten im Bereich der elektronischen Informationsversorgung. Neben der klassischen Literaturversorgung im Printbereich erobern die Bibliotheken neue Arbeitsfelder wie z. B. das elektronische Publizieren oder die Langzeitarchivierung elektronischer Dokumente. In diesem Spektrum stellt die retrospektive Digitalisierung von Bibliotheksbeständen einen wesentlichen Teilbereich dar. Dabei spielte das von der DFG im Jahre 1997 eingerichtete Förderprogramm „Retrospektive Digitalisierung von Bibliotheksbeständen“ für viele Digitalisierungsaktivitäten in Deutschland eine wichtige Rolle [9]. Da die SUB Göttingen als eines ihrer Sondersammelgebiete



„Browsing“ im *Electronic Research Archive for Mathematics*

die „Reine Mathematik“ betreut, hat sie großes Interesse daran, ihre mathematischen Altbestände zu digitalisieren. Für ein solches Vorhaben stellte das mathematische Rezensionorgan „Jahrbuch über die Fortschritte der Mathematik“, das von 1868 (1871) bis 1942 (1944) im Berliner Verlag Walter de Gruyter & Co erschien, eine ideale bibliographische Erschließungsquelle dar. Dank der Initiative Hans Jürgen Beckers, meines Vorgängers im Fachreferat an der SUB Göttingen, haben beide Ideen, die der Digitalisierung mathematischer Altbestände der SUB Göttingen und die des Aufbaus der Jahrbuch-Datenbank, „zueinander gefunden“ und sind als gemeinsames Projekt bei der DFG beantragt worden.

Die SUB Göttingen konnte bei der Beantragung des Projektes im Rahmen des genannten Förderprogramms bereits auf Vorarbeiten zur retrospektiven Digitalisierung und der Bereitstellung elektronischer Dokumente verweisen. Im Jahre 1997, d. h. rechtzeitig zu Beginn der Projektlaufzeit, wurde (ebenfalls mit Förderung durch die DFG) an der SUB Göttingen das Göttinger Digitalisierungszentrum (GDZ) gegründet. Damit waren alle notwendigen technischen Voraussetzungen zum Aufbau des ERAM geschaffen.

Zielstellung

Im Projektantrag vom 27. September 1996 heißt es dazu:

Ziel des Projekts ist der Aufbau einer digitalen Fachbibliothek wichtiger mathematischer Publikationen, die in der zweiten Hälfte des 19-ten Jahrhunderts und der ersten Hälfte des 20-ten Jahrhunderts erschienen sind. Auf der Basis des JFM wird eine recherchierbare Literaturdatenbank erstellt. Diese Datenbank soll zum einen die mathematischen Publikationen von 1868–1943 möglichst vollständig auflisten, zum anderen aber auch eine Zugangsmöglichkeit zu den wichtigen Originalarbeiten, die in der digitalen Fachbibliothek abgelegt sind, bieten. Das Volltextarchiv soll sich zu einer wichtigen Informationsquelle für die mathematische Literatur von 1868–1943 entwickeln.

Die Datenbank des JFM bildet zudem – nach Abschluß des Projektes – zusammen mit der vom Fach-

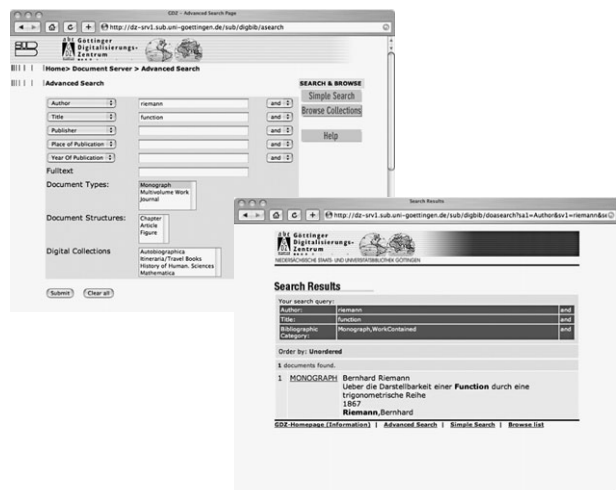
informationszentrum Karlsruhe (FIZ Karlsruhe) herausgegebenen Datenbank MATH des Zentralblatts für Mathematik eine einzigartige und lückenlose Bibliographie für die Mathematik von 1868 bis zum heutigen Tag, die über eine modernen Recherchefunktionen entsprechende Oberfläche zugänglich ist.

Ich habe diese Passage hier deshalb so ausführlich zitiert, weil sie deutlich macht, dass die damals formulierten und zukunftsweisenden Zielstellungen ausnahmslos erfolgreich umgesetzt worden sind.

Dem aufmerksamen Leser, der außerdem die Funktionalitäten des ERAM nutzt und kennt, müsste spätestens an dieser Stelle auffallen, dass sowohl im ausführlichen Projektantrag als auch in der gerade zitierten Passage des Antrages das Wort „Volltext“ benutzt wird, dass ERAM zur Zeit jedoch keine Volltexte in dem in der Einleitung geschilderten Verständnis anbietet. Ein weiterer Blick auf den Antragstext schafft hierzu Klarheit. Bei Beantragung des Projektes hat man unter einem Volltext offenbar die Existenz eines Dokumentes ausdrücklich als gesanntes Image verstanden. Die Anwendung von OCR und die Erstellung des ERAM als Volltextarchiv nach heutigem Verständnis waren in dem Projekt nicht vorgesehen. Im Antrag heißt es zum Aufbau des Archivs: „Die Publikationen werden durch Einscannen digitalisiert und auf einem Speichermedium (Server an der SUB Göttingen) abgelegt.“

Das Archiv

Das ERAM stellt als Content-Komponente gewissermaßen die materielle Grundlage des Projektes dar. Dank der Projektförderung durch die DFG konnte es als sehr umfangreiches und auf Grund seiner inhaltlichen Auswahl international sehr angesehenes Archiv mathematischer Literatur aufgebaut werden. Derzeit enthält es eine über eine Million Seiten umfassende Sammlung von Retrodigitalisaten ausgewählter Texte (über 400 Bücher und 40 Zeitschriften). Die eingescannten mathematischen Monographien und Zeitschriften stehen über das Internet weltweit kostenfrei zur Verfügung.



In diesem Archiv findet man nicht nur die gesammelten Werke von Carl Friedrich Gauß und David Hilbert, sondern beispielsweise auch die Habilitationsschrift von Bernhard Riemann oder die Doktorarbeit Emmy Noethers. Neben mathematischen Altbeständen aus dem Monographienbereich beinhaltet das Archiv auch neuere Monographien, für die ein entsprechendes Einverständnis zur Digitalisierung vorliegt. Bei den mathematischen Zeitschriften wird in der Regel der Zeitraum vom Beginn des Erscheinens an bis zu dem Zeitpunkt, von dem an die Zeitschrift elektronisch publiziert wird, abgedeckt. Auch hier liegen entsprechende Digitalisierungslizenzen vor, die das Einscannen der betreffenden Zeitschriften und die Bereitstellung der Digitalisate im Internet in dem genannten Umfang erlauben. Darunter befinden sich das *Journal für die Reine und Angewandte Mathematik (Crelles Journal)* von 1826 (Band 1), dem Geburtsjahr Bernhard Riemanns, bis 1997 (Band 493) und das *Forum Mathematicum* 1 (1989)–9 (1997) aus dem Berliner Walter de Gruyter Verlag sowie zahlreiche Springer-Zeitschriften wie z. B. *Manuscripta mathematica* 1 (1969)–95 (1998), *Mathematische Annalen* 1 (1869)–306 (1996), *Inventiones mathematicae* 1 (1966)–123 (1996) und klassische Schriftenreihen wie die *Abhandlungen der Königlichen Gesellschaft der Wissenschaften in Göttingen* 1 (1838)–38 (1892).

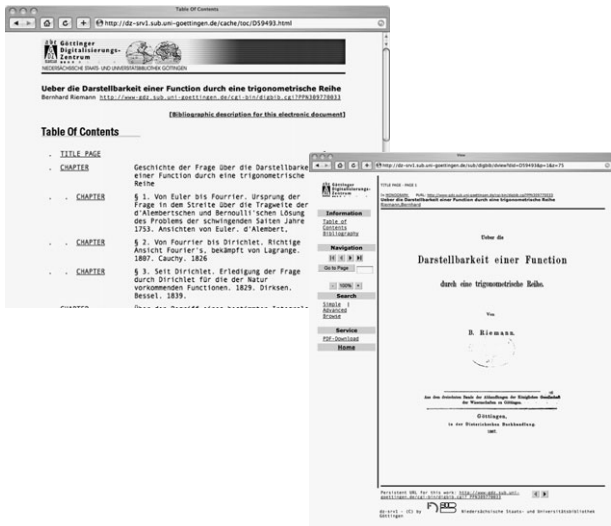
Das Göttinger Digitalisierungszentrum (GDZ), an dem das ERAM angesiedelt ist, ist eine Abteilung der SUB Göttingen und als Service- und Kompetenzzentrum für Bibliotheken und wissenschaftliche Einrichtungen auf dem Gebiet der retrospektiven Digitalisierung sowohl national als auch international aktiv. Bei der Realisierung der Digitalisierungsarbeiten innerhalb des kontinuierlichen Produktionsprozesses am GDZ werden ständig neue Erfahrungen gesammelt, einzelne Arbeitsschritte immer wieder hinterfragt und regelmäßig entsprechend technischer Gegebenheiten optimiert. So wurden auch während der Projektlaufzeit einzelne Arbeitsschritte genauer defi-

niert und aktuellen technischen Entwicklungen angepasst. Ein Beispiel für eine solche Entwicklung war die anfängliche Diskussion über die optische Auflösung der Scans. Mit der auch bei Druckern häufig verwendeten Maßeinheit dpi (dots per inch) wird die Anzahl der Bildpunkte angegeben, die innerhalb eines Zoll (2,54 cm) Platz finden. Je mehr Punkte pro Zoll angegeben werden können, um so besser ist schließlich die optische Qualität des Bildes. Für die Mathematik spielt diese Problematik bei der korrekten Wiedergabe von Formeln (z. B. Symbole + und –) und bei Indizes eine wichtige Rolle. In dieser Richtung wurde am GDZ anfangs mit unterschiedlichen Auflösungen experimentiert. Diese Experimente machten deutlich, dass unter den heutigen Gegebenheiten für mathematische Texte 600 dpi der Standard sein sollte. Dieser wurde dann vom GDZ durchgängig realisiert. Damit folgte man auch internationalen Empfehlungen für die Behandlung mathematischer Texte. Denn auch in Mathematikerkreisen, auf internationaler Ebene initiiert vom CEIC (Committee for Electronic Information and Communication) der International Mathematical Union (IMU), setzte das Bestreben um einheitliche Standards bei der Erfassung der Images ein. (Siehe dazu auch die vom CEIC veröffentlichten „Some Best Practices for Retrodigitization“ [10].) Diese Qualität der Images ermöglicht insbesondere eine spätere Volltext-Erfassung.

Aber nicht nur bei der technischen Umsetzung der Scanarbeiten, auch bei der Bereitstellung der gescannten Materialien für die Nutzer müssen immer wieder neue technische Anforderungen berücksichtigt und umgesetzt werden. Heute steht den Nutzern insbesondere die Möglichkeit eines PDF-Downloads der Dokumente zur Verfügung. Auch eine Lieferung von Materialien per CD-ROM wird angeboten.

Das Archiv ist im Internet direkt unter <http://gdz.sub.uni-goettingen.de/> erreichbar. Dort wählt man die Kollektion „Mathematica“. Für dieses Repository wird ein Dokumentmanagement-System genutzt, welches einfache und erweiterte Suchmöglichkeiten zur Verfügung stellt. Daneben ist auch ein Browsing-Einstieg in die Bestände des Archivs möglich. Dabei muss man beachten, dass das Browsing über zwei Facetten erfolgt. Dies sind die bibliographischen Kategorien *Monographs*, *Multivolume works*, *Journals* und *Articles* sowie die Gruppen *Author* und *Title*.

Bei den Monographien und mehrbändigen Werken liefert die Wahl des Anfangsbuchstabens des gesuchten Autors bzw. des gesuchten Titels eine Ergebnisliste, in der man wiederum nach Auswahl des entsprechenden Links die zugehörige bibliographische Beschreibung des Werkes erhält. In der Folge kann man sich das Inhaltsverzeichnis anzeigen lassen, von wo



aus man dann direkt zu den gescannten Images des Textes gelangt.

Zum Browsing in den Zeitschriftenbeständen des ERAM wählt man in den oben genannten Browsing-Facetten die Gruppierung *Title* und kann sich zu entsprechenden Anfangsbuchstaben die Zeitschriftentitel auflisten lassen. Nach Auswahl einer Zeitschrift erscheint eine Liste aller zur Verfügung stehenden Bände, bevor man sich die Inhaltsverzeichnisse einzelner Hefte anzeigen lassen kann. Von den Inhaltsverzeichnissen aus gelangt man schließlich zu den einzelnen Artikeln.

Zur Zeit ist das ERAM das umfangreichste mathematische Archiv dieser Art. Gemessen am gesamten Altbestand mathematischer Texte, der allein an der SUB Göttingen ca. 5 000 Titel umfasst, stellt das ERAM gegenwärtig aber nur einen Ausschnitt dieser reichen Ressourcen zur Verfügung. Lohnend ist ein Besuch und ein Stöbern im ERAM allemal, da in Kooperation mit Mathematikern für dieses Kernarchiv eine sehr schöne Auswahl historisch wertvoller mathematischer Texte getroffen werden konnte.

Die Datenbank

Mit der Gründung des *Jahrbuch über die Fortschritte der Mathematik* durch Carl Ortmann und Felix Müller als erste Herausgeber trat 1868 das mathematische Referatewesen in Deutschland erstmals in Erscheinung. Die Herausgeber schildern in der „Vorrede“ (siehe [15]) im ersten Band ihre Intention wie folgt:

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

Ein „grosser Uebelstand“, der von den Herausgebern bereits in der „Vorrede“ im ersten Band aufgezeigt wird und sich später als von immer größerem Gewicht erwies, war der beträchtliche Aktualitätsverlust vom Erscheinen der Publikation bis zum Referat im Jahrbuch.

Die zum Aktualitätsrückstand dazukommende starke Entfaltung der mathematischen Forschung, schlechte finanzielle Rahmenbedingungen und kriegsbedingt auflaufende Rückstände waren weitere Faktoren, die dazu führten, dass das Jahrbuch 1944 sein Erscheinen beenden musste. Mit seinen über 200 000 Rezensionen in 68 Bänden ist es eine der weltweit meistgenutzten Quellen mathematischer Forschung.

Mit dem Konzept einer schnelleren Information über neue mathematische Publikationen wurde 1931 das *Zentralblatt für Mathematik (und ihre Grenzgebiete)* gegründet. Während die zu Beginn der 90er Jahre angebotene elektronische Datenbank MATH sich zunächst als dem Zentralblatt assoziiert betrachtete, werden heute die elektronische (auf CD-ROM und im Internet) und die Printausgabe als technisch verschiedene Formen des *Zentralblatt MATH* entsprechend der aktuellen Publikationsformen, d. h. parallel sowohl gedruckt als auch elektronisch, herausgegeben.

Den vom *Zentralblatt MATH* gesetzten technischen Standards folgend, sollte eine zeitgerechte Aufarbeitung des Jahrbuches zu einer Homogenisierung in der Nutzbarkeit der angebotenen Informationen führen. Dies wurde als Service-Komponente des hier geschilderten ERAM/Jahrbuch-Projektes realisiert. Dazu wurden die Referate des Jahrbuches in einer Datenbank, der so genannten Jahrbuch-Datenbank, erfasst und – soweit Digitalisate vorhanden sind – mit Links auf das Göttinger ERAM versehen. Die Struktur der Jahrbuch-Datenbank entspricht dabei im wesentlichen der des *Zentralblatt MATH*. So konnte für die Jahrbuch-Datenbank im Prinzip die gleiche Software wie für die Datenbank des *Zentralblatt MATH* Verwendung finden. Auch die vorgesehene Integration der Jahrbuch-Daten in das *Zentralblatt MATH* konnte auf diese Weise einfach realisiert werden.

Bemerkenswert ist, dass die Projektförderung die komplette manuelle Neuerfassung und Aufbereitung des Jahrbuches im in der Mathematik gängigen $\text{T}_{\text{E}}\text{X}$ -Format umfasste.

Die Erfassung der Daten wurde von drei Schreibbüros in Berlin, Sofia und Prag übernommen. Die Weiterbearbeitung und Ergänzung der von den Schreibbüros erfaßten Daten erfolgte dann teilweise in Berlin und teilweise in Göttingen. In Berlin beispielsweise wurden die Daten sortiert, Dubletten entfernt,

Kapitelüberschriften ergänzt, die vollständigen Namen der Referenten hinzugefügt und Jahreszahlen eingetragen. Außerdem wurden interne Links gesetzt, die manuell nachgeschlagen und kontrolliert werden mussten. In Göttingen wurden die Bände bibliographisch nachbereitet. Diese bibliographische Nachbearbeitung umfasste zum Beispiel für viele Monographien die Aufnahme vollständiger Autorennamen mit Lebensdaten, Herausgeberjahr, Verlag, Ort, Signatur etc. in die Datenbank.

Nach Beendigung der ersten Bearbeitung des Datamaterials wurden die Dokumente kapitelweise an etwa 70 Experten geschickt, die bis heute ehrenamtlich an diesem Vorhaben mitarbeiten. Diese Experten übersetzen die Titel der Artikel ins Englische, vergeben Keywords, klassifizieren die Arbeiten nach dem mathematischen Klassifikationsschema MSC 2000, korrigieren Schreibfehler und beurteilen, ob die Artikel auch heute noch wichtig sind. Der Bearbeitungszeitraum dafür liegt zwischen einem Monat und einem Jahr und die Anzahl der freiwilligen Mitarbeiter ist von Gebiet zu Gebiet sehr unterschiedlich. Generell ist der freiwillige Beitrag der Fachgemeinschaft an solchen Projekten schwer kalkulierbar und die Beteiligung hier sehr erfreulich. Da die Zuarbeiten der Experten noch nicht abgeschlossen sind und auch zukünftig eingearbeitet werden müssen sowie weitere Links auf externe Digitalisate gesetzt werden sollen, sind für die Nachhaltigkeit der Projektergebnisse regelmäßige Updates der Jahrbuch-Datenbank notwendig.

In der Jahrbuch-Datenbank sind alle 68 Bände des *Jahrbuch über die Fortschritte der Mathematik* erfasst. Aus der Datenbankrecherche heraus kann sich der Nutzer direkt bis zum gescannten Text – soweit vorhanden – durchklicken. Neben zahlreichen Links auf das Göttinger ERAM führen weitere von den jeweiligen Referaten auf inzwischen weltweit vorhandene retrodigitalisierte mathematische Originalquellen. Insgesamt sind dies zur Zeit 17 035 Links zu gescannten Texten in Göttingen, Cornell, Michigan, Paris und Grenoble.

Da die Jahrbuch-Datenbank mit finanzieller Förderung durch die DFG aufgebaut worden ist, besteht für den Nutzer die Möglichkeit einer freien Recherche. Beim Sucheinstieg kann man hier neben der Recherche in den Besprechungen zusätzlich die Option „Documents with facsimiles“ wählen. Dann werden nur die Treffer aus der Jahrbuch-Datenbank angezeigt, für die eine Verlinkung zu dem entsprechenden gescannten Dokument zur Verfügung steht. Dieser Link ist ein mit einem roten Pfeil versehener Button und führt entweder ins ERAM oder auch zu einem der anderen digitalen Mathematikarchive weltweit.

Da die Daten der Jahrbuch-Datenbank zusätzlich in die des *Zentralblatt MATH* integriert wurden, werden sie automatisch in die Recherche im *Zentralblatt MATH* einbezogen. Auf diese Weise werden im *Zentralblatt MATH* Literaturrecherchen über einen Zeitraum von fast 140 Jahren ermöglicht.

Internetadressen

Allgemeine Projektinformationen:

<http://www.emis.de/projects/JFM/JFM.html>

Für die freie Recherche in der Jahrbuch-Datenbank findet man unter <http://www.emis.de/MATH/JFM/JFM.html> den direkte Sucheinstieg. Um eine ständige Verfügbarkeit zu sichern, wird dieser Service zusätzlich unter <http://jfm.sub.uni-goettingen.de/> angeboten.

Neben einer Verlinkung über die Jahrbuch-Datenbank auf Teile des Archivs kann auf das gesamte ERAM autonom über die so genannte Mathematica-Kollektion des GDZ unter <http://gdz.sub.uni-goettingen.de/> zugegriffen werden.

An dieser Stelle sei nochmals ausdrücklich darauf hingewiesen, dass das ERAM wesentlich mehr Material bereit hält, als über die Suche in der Jahrbuch-Datenbank erreicht werden kann.

Resonanz

Das gesamte Projekt wurde unter lebhafter Beteiligung der mathematischen Community und einer Reihe von Kommissionen, nationalen Arbeitsgruppen und internationalen Gremien realisiert. Auch bei den internationalen Aktivitäten zum Aufbau der World Digital Mathematics Library (WDML) ist das Projekt präsent und gefragter Partner.

Die aufgebauten Services sind heute fester Bestandteil des mathematischen Forschungsalltags. Monatlich gibt es zwischen 5 000 und 10 000 Zugriffe auf die Jahrbuch-Datenbank. Dabei ist bemerkenswert, dass die Zahlen weiter angestiegen sind, obwohl viele Nutzer auch den Zugang über die gleichzeitige Recherche des Materials im *Zentralblatt MATH* nutzen.

Erwähnung findet das Projekt im Internet an den verschiedensten Stellen. Beispielsweise wird es auf den Internet-Seiten zahlreicher Bibliotheken und mathematischer Institute als Fachinformationsquelle für die Mathematik angeboten und kommentiert.

Hervorzuheben ist die Auszeichnung mit dem Award 2005 der Physics-Astronomy-Mathematics (PAM) Division der amerikanischen Special Libraries Association (SLA), einem amerikanischen Bibliothekspreis. Die PAM Division der SLA begründete ihre Entscheidung mit den Worten



The award should be given for a significant contribution to the literature of physics, mathematics or astronomy or to honor work that demonstrably improves the exchange of information in physics, mathematics or astronomy. The contribution should also significantly benefit libraries or enhance the ability of librarians to provide service. It should be special – above and beyond the normal job requirements of the individual(s) or group concerned. The ERAM or Jahrbuch Project is perfectly suited for this award. Mathematics librarians and researchers around the world depend daily on this enhancement to our research tools.

Weitere Informationen zu dieser Auszeichnung findet man unter <http://www.sla.org/division/dpam/manual/awards/2005division.html> sowie auf der Homepage der Jahrbuch-Datenbank.

Ausblick

Technisch sind die Möglichkeiten zur Realisierung und die Ansprüche der Nutzer in den fast zehn Jahren seit Beginn der Arbeiten an dem Projekt erheblich gewachsen. Die Erfassung der Digitalisate mittels OCR gehört bei zahlreichen heute beginnenden Digitalisierungsaktivitäten zum Produktionsprozess. Dies wird auch in naher Zukunft am GDZ realisiert sein. Die bisherigen Scans liegen außerdem in solch einer Qualität vor, dass die OCR-Erfassung bei diesen auch noch nachträglich durchgeführt werden kann. Ein weiteres Element, welches in der sich entwickelnden digitalen Bibliothek von wachsender Bedeutung ist und immer stärker gewünscht wird, ist der Ausbau

bisheriger Verlinkungen zu einem umfassenden „Referenzlinking“ [1]. Darunter versteht man die Verknüpfung von Literaturzitationen eines Artikels oder Buches mit entsprechenden digitalen Dokumenten im Internet, mit Bibliothekskatalogen, mit weiteren Suchsystemen, mit zugehörigen Besprechungen oder dem betreffenden Volltext, der ein direktes Weiterlesen in der zitierten Quelle ermöglicht. Zur Realisierung dieser und weiterer möglicher und wünschenswerter Funktionalitäten, zu denen auch in Zukunft weitere Neuerungen hinzukommen werden, bedarf es auch immer der entsprechenden finanziellen und personellen Ressourcen. Die Bibliotheken sind sich ihrer Verpflichtung durchaus bewusst und nicht nur das, sie bemühen sich auch vor dem Hintergrund der massiven Vorstöße kommerzieller Unternehmungen zielstrebig um die Umsetzung entsprechender Aktivitäten und die Mobilisierung aller Partner [8].

Weitere Planungen und Aktivitäten im Bereich der mathematischen Fachinformation sehen unter anderem die Einbindung des ERAM in das jüngst von der DFG bewilligte Projekt zum Aufbau der „Virtuellen Fachbibliothek Mathematik“ vor. Über dieses Projekt, welches zur Zeit an der SUB Göttingen realisiert wird, gilt es später zu berichten.

Weiterhin sind das ERAM sowie die Jahrbuch-Datenbank in weltweite Aktivitäten zum Aufbau einer globalen digitalen Mathematik-Bibliothek einbezogen. Die World Digital Mathematics Library (WDML) ist das ambitionierte Vorhaben, möglichst alle mathematischen Publikationen in digitaler Form (digitalisiert bzw. bereits elektronisch produziert) und über das Internet weltweit zur Verfügung zu stellen (siehe [3]). Die Mathematik wäre dann eine der ersten Wissenschaften, die über eine vollständige digitale Bibliothek verfügen würde. Schätzungen gehen dabei von ca. 50 Millionen retrospektiv zu digitalisierenden Seiten aus. Zur Realisierung der WDML sind in Zukunft immer noch eine Reihe genereller Fragen und Probleme zu lösen. Dies sind nicht nur konzeptionelle und technische Probleme, sondern auch Fragen der Lizenzierung, zu Geschäftsmodellen usw. Die International Mathematical Union (IMU) als Vertreterin der internationalen Gemeinschaft der Mathematiker ist an der Realisierung der WDML sehr stark interessiert und setzt sich für deren Realisierung nachdrücklich ein.

Schlussbemerkungen

Sicher haben viele von Ihnen mittlerweile auch einen leistungsstarken Scanner auf dem Schreibtisch und möglicherweise auch das eine oder andere wertvolle mathematische Werk im Bücherschrank stehen. Vielleicht haben Sie sogar schon einmal daran gedacht,

so ein Werk einfach einzuscannen und ins Netz zu stellen. Scannen kann doch heute jeder!

Wer sich jedoch ein wenig mit der Materie vertraut macht, wird bald die Komplexität solch eines Unterfangens erkennen. Digitalisierung gedruckter Dokumente ist ein umfassender Prozess, der auch eine ausgereifte Planung von Prozessabläufen benötigt und bei dem Umfang, von dem hier die Rede ist, eine logistische Herausforderung darstellt. Der Vorgang des eigentlichen Scannens macht nur einen Bruchteil der zu leistenden Arbeiten aus. Hinzu kommen zahlreiche vorbereitende und nachbereitende Tätigkeiten, die unter anderem die Auswahl der Quellen, die bibliothekarische und technische Vorbereitung des Materials, die Nachbearbeitung und Qualitätskontrolle der Scans, die Erfassung der Meta- und Strukturdaten (Titel, Autor, Inhaltsverzeichnis, Kapitelüberschriften etc. – wobei größte Sorgfalt angezeigt ist), die permanente Speicherung und Archivierung der Scans, die eindeutige Zuordnung von Identifikatoren für die Dokumente und schließlich die ganzheitliche Aufbereitung und Bereitstellung des Materials für die Nutzung (Internet, CD-ROM) umfassen.

Die Realisierung eines Digitalisierungsprojektes ist eben nicht einfach „Vorlage auflegen, Scanbutton anklicken und ins Netz stellen“.

Nachdem ich David Hilbert zu Beginn zitiert habe, lassen Sie mich auch mit einem Zitat von ihm (siehe [7]) zur Mathematik schließen:

Für uns Mathematiker ist es die Aufgabe, sie wie ein Heiligtum zu hüten, damit einst *alles* menschliche Wissen überhaupt der gleichen Präzision und Klarheit teilhaftig wird. Daß dies kommen muß und geschehen wird, ist meine feste Überzeugung.

Literatur

- [1] Priscilla Caplan, William Y. Arms: *Reference Linking for Journal Articles*. <http://www.dlib.org/dlib/july99/caplan/07caplan.html>
- [2] R. Keith Dennis, Bernd Wegner: *Die Erfassung der mathematischen Literatur von 1868 bis 1942*. Mitteilungen der DMV Heft 2, 1996, Seite 55.
- [3] John Ewing: *Twenty centuries of mathematics: Digitizing and disseminating the past mathematical literature*. http://www.ams.org/ewing/Twenty_centuries.pdf
- [4] Silke Göbel, Bernd Wegner: *Das Zentralblatt als Zugang zur mathematischen Literatur*. Mitteilungen der DMV Heft 1, 2004, Seiten 48–49.
- [5] Martin Grötschel: *Der Weg der Mathematik in die Informationsgesellschaft*. Mitteilungen der DMV Heft 3, 1997, Seiten 48–57. <http://www.math-net.de/project/papers/dmv-mitteilungen/dmv-mitteilungen97.html>
- [6] David Hilbert: *Über das Unendliche*. Mathematische Annalen, 95 (1926), Seite 170. http://www-gdz.sub.uni-goettingen.de/cgi-bin/digbib.cgi?PPN235181684_0095
- [7] David Hilbert: *Die Grundlagen der elementaren Zahlentheorie*. Mathematische Annalen, 104 (1931), Seite 494. http://www-gdz.sub.uni-goettingen.de/cgi-bin/digbib.cgi?PPN235181684_0104
- [8] Hilmar Schmudt: *Schwatzen, Schrillen, Schreien*. Der Spiegel Heft 34, 2005, Seiten 118-120
- [9] *Zum Förderprogramm der DFG zur Retrodigitalisierung von Bibliotheksbeständen*: http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/retro_digitalisierung_eval_050406.pdf http://www.sub.uni-goettingen.de/ebene_2/vdf/endfas.pdf
- [10] http://www.ceic.math.ca/Publications/retro_bestpractices.pdf
- [11] <http://www.msri.org/local/DMLP/1-Eisenbud-Moore70.pdf>
- [12] <http://www.mathematik.uni-bielefeld.de/documenta/>
- [13] <http://www.emis.de>
- [14] <http://www.emis.de/classics/Riemann/>
- [15] <http://www.emis.de/projects/JFM/vor2.html>
- [16] <http://gutenberg.spiegel.de>
- [17] <http://www.math-net.org>
- [18] <http://mathnet.preprints.org/>

Adresse der Autorin

PD Dr. Katharina Habermann
 Fachreferat für Mathematik und Informatik
 Nds. Staats- und Universitätsbibliothek Göttingen
 Platz der Göttinger Sieben 1
 37073 Göttingen
habermann@sub.uni-goettingen.de



Katharina Habermann, Jahrgang 1966, begann ihre wissenschaftliche Laufbahn an der Humboldt-Universität zu Berlin. Sie besuchte an dieser Einrichtung die Spezialschule für Mathematik und Physik, studierte von 1984 bis 1989 Mathematik und promovierte 1993. Nach der Promotion ging sie an die Ruhr-Universität Bochum, wo sie sich im Jahre 1999 habilitierte. Weitere Stationen waren das Max-Planck-Institut für Mathematik in den Naturwissenschaften in Leipzig und die Ernst-Moritz-Arndt-Universität Greifswald. Im Jahr 2000 erhielt sie den Gerhard-Hess-Preis der Deutschen Forschungsgemeinschaft, mit dem sie ein eigenes Forschungsprojekt auf dem Gebiet der symplektischen Geometrie realisieren konnte. Seit 2004 arbeitet sie als Fachreferentin für Mathematik und Informatik an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen und betreut dort insbesondere das Sondersammelgebiet „Reine Mathematik“. Katharina Habermann ist außerdem zweite Sprecherin der IuK-Fachgruppe der DMV/ÖMG.