MathDiss International

Vorgeschichte 1
Dissertationen Online und seine Ergebnisse 2
Spezifika mathematischer Dissertationen 3
Das Konzept von Math*Diss* International 3
Zielsetzungen und Umsetzung des Projektes Math*Diss* International 5
Internationale Kooperation 6
Fazit 6
Ansprechpartner 7

Vorgeschichte

Die Entwicklung einer scientific Community wird wesentlich auch durch ihre wissenschaftliche Produktivität geprägt; die Aktualität der in den wissenschaftlichen Communities *heute* diskutierten und bearbeiteten Forschungsfragen - insbesondere im mathematisch-naturwissenschaftlichen Bereich - entscheidet wesentlich über die *morgigen* technologischen Erfolgsaussichten dieser Gesellschaft. Dieses wird gerade durch die Tatsache unterstrichen, dass die Mathematik mittlerweile den Rang einer Schlüsseltechnologie einnimmt. Hier ist das Halten der Anschlussfähigkeit der länderspezifischen wissenschaftlichen Forschung eine stets neu einzulösende Aufgabe.

In diesem Kontext stellen Dissertationen[1] einen wesentlichen Indikator für wissenschaftliche Forschungen dar. Durch diese Beiträge werden einerseits die aktuellen Forschungsfragen in den Forschungslabors wie an den Schreibtischen dokumentiert, andererseits spiegeln sich in ihnen auch die eingesetzten Methoden wider. Daher muss jede Wissenschaftsgesellschaft ein originäres Interesse haben, diese Informationen zu *erschließen* und *offen zugänglich* zu machen, damit[2] Information als Rohstoff für Innovation dienen kann.

Informationen bergen jedoch auch Gefahren in sich, wenn sie ausschließlich als ‚Flut' wahrgenommen werden. Es müssen Konzepte entwickelt werden, dass der einzelne Wissenschaftler der damit verbundenen enormen Informationsflut adäquat begegnen kann. Dies ist möglicherweise eher ein Charakteristikum unserer westlichen Gesellschaften, während Wissenschaftler in den Ländern der Dritten Welt eher das Gegenteil beklagen könnten. In diesem Zusammenhang kommt der masselosen elektronischen Information und Kommunikation (IuK) eine entscheidende Funktion zu: sie kann beiden Defiziten Rechnung tragen, sie kann die gezielte Selektion optimieren, sie kann aber auch ‚schwerelos' Zugänge eröffnen, die klassisch nur mit großem Aufwand - wenn überhaupt - realisiert werden könnten.

Es überrascht daher nicht, dass gerade die wissenschaftlichen Fachgesellschaften der Mathematik und Physik in der Bundesrepublik schon vor Jahren sowohl den Handlungsbedarf, wie auch die Möglichkeiten erkannt haben. Sie haben die IuK-Initiative[3] ins Leben gerufen, um die Veränderung der Informations- und Kommunikationsprozesse in den Wissenschaften, nicht zuletzt in den exakten Wissenschaften, gestaltend begleiten zu können. Als Beispiele für die Entwicklungen in der Deutschen Mathematiker Vereinigung (DMV)[4] dient das *Math-Net*[5] - der Internet-

---

[1] In der Bundesrepublik Deutschland zählt Die Deutsche Bibliothek (DDB) mehr als 20.000 Exemplare pro Jahr
[2] um ein Programm der Bundesregierung 1996 - 2000 zu zitieren
[3] http://www.iuk-initiative.org
[4] http://www.mathematik.uni-bielefeld.de/DMV/mathnet_de.html
[5] http://www.Math-Net.org

Informationsdienst für Mathematiker - als Referenz und für die Deutsche Physikalische Gesellschaft (DPG)[6] sei auf das *PhysNet*[7]- das Physics Department and Document Network - verwiesen.

Wissenschaftler nehmen in diesem Prozess eine doppelte Funktion war: sie sind Informationslieferanten wie auch Informationskunden. Zum ersten Mal relativiert sich auf der elektronischen Ebene die Maklerrolle von Verlagen und Literaturproduzenten, zumal die Finanzierung der Anschaffungen bibliothekarischer Informationen zusehends auch für bundesdeutsche Universitäten unbezahlbar wird. Insofern sind die Wissenschaftler gut beraten, ihre Interessen partiell selbst in die Hand zu nehmen. Die Deutsche Mathematiker Vereinigung (DMV) hat in Kooperation mit der österreichischen Mathematiker Gesellschaft (ÖMG)[8] diesen Fragen in einer eigenständigen Fachgruppe eine Plattform[9] gegeben, die regelmäßig auf den Jahrestagungen der Gesellschaften Sektionen betreut.

Hier spielen nun die elektronisch vorgehaltenen wissenschaftlichen und allgemein zugänglichen Dokumente eine zentrale Rolle. Gerade der Nonprofit-Sektor eröffnet Möglichkeiten der Gestaltung, wobei den elektronisch repräsentierten Dissertationen eine zentrale, ja exemplarische Rolle zukommt. Im internationalen Kontext ist für den Bereich der elektronischen Qualifikationsarbeiten ein neues Schlagwort etabliert worden, man spricht von ‚Electronic Theses and Dissertations' (kurz: ETDs). Einerseits sind ETDs ein geeignetes Objekt, die neuen Workflows und damit verbundene Probleme der elektronischen Publikation zu studieren, andererseits geht es um die nicht geringe inhaltliche Bedeutung dieser Dokumente. Die jährlichen ETD-Symposien der Networked Digital Library of Theses and Dissertations (NDLTD)[10] belegen, dass diese Chancen von den Communities erkannt wurden und nun auch Gegenstand vielfältiger Untersuchungen und Förderungen geworden sind.

Dissertationen Online und seine Ergebnisse

Im abgelaufenen DFG-Projekt Dissertationen Online[11], dessen Initiierung von der DMV mitgetragen worden ist, wurden Lösungen entwickelt, Dissertationen elektronisch zu archivieren und zu publizieren. Das erwähnte Projekt beschäftigte sich schwerpunktartig mit

- der Erschließung von digitalen Dissertationen durch Metadaten,
- der Anpassung einer Suchmaschine an weitere Metadatentypen,
- der fächerübergreifenden Vernetzung von Servern als Suchmaschinen,
- der Entwicklung von Document Type Definitions (DTDs) und der automatisierten Generierung der Informationen zum Zwecke einer Konvertierung nach XML,
- der Einbringung multimedialer Elemente (Chemie) in digitale Dissertationen,
- dem Aufbau eines benutzerorientierten Informationssystems und
- der Anpassung an den bestehenden Geschäftsgang in den Bibliotheken.

Bei der Durchführung des Projektes wurden in den Jahren 1998-2000 deutlich, dass es Barrieren gibt, die einer alle Fachgebiete übergreifenden einheitlichen Lösung, Dissertationen elektronisch zu handhaben, entgegenstehen. Dabei wollen wir die gewachsenen Gepflogenheiten in den verschiedenen Fakultäten, Dissertationsabläufe zu definieren, als eher marginal (und lösbar) außer Acht lassen.

---

[6] http://www.dpg-physik.de
[7] http://physnet.uni-oldenburg.de/PhysNet/physnet.html
[8] http://www.mat.univie.ac.at/~oemg/
[9] http://elib.zib.de/IuK-DMV/
[10] http://www.ndltd.org
[11] http://www.dissonline.org

Als zentrales Problem muss nämlich weiterhin der Aspekt der *Datenformate* gesehen werden. Benutzte Datenformate orientieren sich zwangsläufig an den Präsentationsnotwendigkeiten für die in Frage stehenden wissenschaftlichen Ergebnisse. Insofern kann hier wohl auch mittelfristig nicht auf Einheitlichkeit gesetzt werden. Jeder Doktorand wird eine dem Komplexitätsniveau angemessene Sprache verwenden. Schwerpunkt im oben erwähnten Projekt *Dissertationen Online* waren Dokumente, die im Textverarbeitungssystem Microsoft Word erstellt worden sind, einem System, das den Formatanforderungen im geisteswissenschaftlichen Bereich angepasst ist. Mit nachfolgender Priorität wurden Dokumente bearbeitet, die im LaTeX-Format, dem Standardformat für mathematisch-naturwissenschaftliche Dokumente, erstellt worden sind.

Spezifika mathematischer Dissertationen

Während davon ausgegangen werden kann, dass Word-basierte Dissertationen, die keinen mathematikspezifischen Formelanteil beinhalten, ohne wesentliche Schwierigkeiten[12] auf SMGL als Archivierungsformat gebracht werden können, ist eine Transformation von LaTeX-Dokumenten nach SGML zwar prinzipiell machbar, aber derzeit nur eingeschränkt sachgemäß. Man beachte, dass z.B. Formeln im relativ klar strukturierten LaTeX zu Bildern bzw. Graphiken degenerieren. Andererseits besteht eine mathematische Dissertation nicht selten zu ca. 90% aus Formeln, damit wird deutlich, dass eine Transformation nach SMGL bis auf weiteres eine nicht vertretbare Komplexität verursacht.

Dieses Problem kann nicht übergangen werden, da die aktuellen Zahlen (Stand September 2001) im Bereich der Online-Hochschulschriften, die bei Der Deutschen Bibliothek (DDB)[13] gemeldet sind, ein eindeutiges Bild aufzeigen. Der Anteil der Online-Dissertationen an allen Dissertationen im Fach Mathematik liegt bei knapp 20% und wird nur noch von den Dissertationen aus der Chemie (23%) und der Physik (22%) überboten. Insgesamt ergibt sich interdisziplinär ein Online-Anteil von ca. 10%.

In zunehmendem Maße wird diese Problematik auch in den internationalen mathematischen bzw. naturwissenschaftlichen Communities gesehen. Diesem Defizit tragen derzeit erste Neuentwicklungen Rechnung, die den Gebrauch und die Wiedergabe von mathematischem und wissenschaftlichem Inhalt auf dem World Wide Web (WWW) erleichtern sollen und dabei die semantische Information des LaTeX-Codes beibehalten, z.B. OpenMath[14] oder auch MathML (Mathematical Markup Language)[15]. Der in Berkeley veranstaltete internationale Kongress *The Future of Mathematical Communication*[16] hat gezeigt, dass diese ‚Sprachen' entwickelt werden. Da MathML jedoch zusätzliche semantische Informationen benötigt, die im LaTeX-Code nicht enthalten sind, ist eine mathematisch korrekte automatische Konvertierung auf absehbare Zeit hinaus illusorisch. Es wird daher noch Jahre dauern, bis diese Sprachen ausgereift sind und die *Akzeptanz* und *Verbreitung* erreichen (was wissenschaftliche Texterstellung anbetrifft), die derzeit LaTeX in der naturwissenschaftlichen Welt genießt.

Das Konzept von Math*Diss* International

Dem Faktum der Formate und der Archivierung eines Optimums an Informationen muss *heute* die Langzeitarchivierung Rechnung tragen und für eine, noch nicht abzuschätzende Zeit auch ASCII-LaTeX-Files vorhalten, die dann vorläufig (und nicht(!) SGML-Derivate) als Quellfiles auch für die Präsentation dienen sollten. Durch Rückgriff der Textersteller auf selbstentwickelte

---

[12] Jede neue Word-Version birgt allerdings unvorhersehbare Eventualitäten in sich!
[13] http://www.ddb.de
[14] http://www.nag.co.uk/projects/openmath/omsoc/
[15] http://www.w3.org/TR/REC-MathML/
[16] http://www.msri.org/activities/events/9900/fmc99/

Input-Files ist selbst bei LaTeX-Files eine nicht unproblematische Vielfalt zu konstatieren. Hier sind auf jeden Fall Homogenisierungen zu diskutieren. Nichts desto weniger setzen in den Vereinigten Staaten Experten auf Übergangslösungen, nämlich LaTeX-Texte authentisch aufzubewahren, da eine Übersetzung solcher Texte für spätere marktfähige Produkte als Selbstverständlichkeit angesehen wird. Dieses Problem wurde bereits von der International Mathematical Union (IMU)[17], dem *Committee on Electronic Information and Communication* (CEIC)[18] und dem Los Alamos Servers[19] erkannt, was zu einem Aufbau einer Maschine führte, auf der mathematisch-naturwissenschaftliche, elektronische Arbeiten in LaTeX ‚konserviert' werden. Dissertationen spielen allerdings auf diesem Server nur eine Nebenrolle; darüber hinaus genügen die Rahmenbedingungen dieser amerikanischen Maschine keineswegs den hierzulande entwickelten Standards (Metadaten, Authentifizierung usw.), die in Zusammenarbeit von *Dissertationen Online* und den Bibliotheken, z.B. DDB und SUB Göttingen, vereinbart und hierzulande etabliert wurden.

File-Formate haben nicht nur problematische Konsequenzen für die Langzeitarchivierung, sondern spielen bei der automatisierten Erschließung des Textes eine entscheidende Rolle. Da die Volltextsuche kein adäquater Recherche-Ansatz ist und eher zu einer nichthandhabbaren Informationsinflation führt, kommt neben den originären Metadaten auch den *Document Type Definitions* (DTDs) eine zentrale Bedeutung zu. Unter pragmatischen Gesichtspunkten müssen solche DTD-Informationen automatisiert erstellt werden können. Mithin wären prinzipiell für jedes Texterstellungsprogramm (und möglicherweise für jedes Update) eigene Tools zu entwickeln bzw.diese anzupassen.

Dies ist im Projekt *Dissertationen Online* befriedigend für Word-Dokumente[20] geleistet worden, mit zweiter Priorität und nur bedingt befriedigend für LaTeX-Dokumente. Dabei spielte eine Rolle, dass für LaTeX-Files eine Word-parallele Lösung gesucht wurde, um eine grundsätzliche Vergleichbarkeit mit Word-Dokumenten zu gewährleisten. Unberücksichtigt blieb, dass LaTeX eigene Programmmöglichkeiten hat, Meta-Informationen zu generieren, allerdings in XML-affiner Form. Schließlich konnten wir auch nicht dem Umstand Rechnung tragen, dass Word in der Regel in anderen wissenschaftlichen Fachbereichen eingesetzt wird im Vergleich zu TeX resp. LaTeX.

Einmal mehr wird deutlich, dass ein fachgebietshomogener wie auch ein formathomogener Dokumentenserver in diesem Zusammenhang einfachere Lösungen zulässt, wenn man bereits beim Erstellen des Dokuments (z.B. durch gezielte Indexierung) entsprechende Vorbereitungen trifft. Es ist dann denkbar, dass man auf der Ebene der LaTeX-Sprache relativ problemlos Inhaltsverzeichnisse, Glossare, Indizes, Bibliographien usw. in eine Datenbank exportieren kann, in der dann gezielte Recherchen durchgeführt werden können. Da sich aber die Anforderungen in solchen homogenen ‚Fakultätswelten' international kaum unterscheiden, kann hier auf eine internationale Kooperation gesetzt werden, wodurch nebenbei die wünschenswerte Entwicklung von professionellen Konvertierungstools begünstigt wird.

Aufbauend auf den Ergebnissen von *Dissertationen Online* sind diese Überlegungen Grundgedanken für den Aufbau eines internationalen Mathematik-Dissertationen-Servers (Math*Diss* International)[21]. Dieses Projekt, das von der Deutschen Forschungsgemeinschaft (DFG) gefördert wird, ist eine Gemeinschaftsarbeit des Institutes für Mathematik der Gerhard-Mercator-Universität Duisburg und der Staats- und Universitätsbibliothek des Landes Niedersachen in Göttingen

---

[17] http://elib.zib.de/IMU/
[18] http://elib.zib.de/IMU/IMU_Committees/CEIC.html
[19] http://xxx.lanl.gov
[20] Neue Office-Versionen bereiten hierbei neue Probleme!
[21] http://www.ub.uni-duisburg.de/mathdiss/

(SUB)[22]. Letztere wird den Service nach Ablauf der Projektlaufzeit weiter betreuen und fortführen.

Zielsetzungen und Umsetzung des Projektes Math*Diss* International

Die Zielsetzungen des Projektes lassen sich in zwei Bereiche Aufteilen.

- **Suchoptionen**
  Die fachspezifische Suche in mathematischen Dokumenten unterliegt letztlich keinem Wandel. Die Wissenschaftler wenden bei Ihrer Literaturrecherche keine anderen Methoden an, nur weil sich das Suchmedium resp. die Publikationsform geändert hat. Angesichts der Funktionalitäten des Textsatzsystems LaTeX, von deren Gebrauch man ausgehen kann, erscheint eine adäquate Suche möglich. Der als selbstverständlich anzusehende Einsatz der Funktionalitäten, d.h. des automatischen Inhaltsverzeichnisses, der Bearbeitung der Bibliographie mit BibTeX, der Erstellung des Index mit MakeIndex u.a., erzeugt automatisch eigene Dateien, die für ein adäquates Retrieval herangezogen werden können. Insbesondere handelt es sich um folgende Informationen, die einen bedeutenden Mehrwert im Umfeld der Suchoptionen bilden:
  - Abstrakt,
  - Mathematical Subject Classification (MSC),
  - Inhaltsverzeichnisse,
  - Literaturverzeichnisse,
  - Index und
  - Schlag- und Schlüsselwörter.

  Ein Punkt sollte hier jedoch nicht verschwiegen werden, es ist derzeit noch nötig, die Autoren, also hier die Promovenden, zu einem hochwertigen Schreiben ihrer wissenschaftlichen Überlegungen zu erziehen. Schon zu Beginn einer Bearbeitung muss sich der Promovend darüber klar sein, welche Verwendung das spätere Machwerk finden soll. Bei einer elektronischen Veröffentlichung müssen schon von Anfang an die Voraussetzungen für ein **fachlich ansprechendes** und **qualitativ hochwertiges** Retrieval geschaffen werden. Diese Grundvoraussetzung, also die mit Informationen angereicherten Quellfiles, führen direkt zum zweiten zentralen Punkt:

- **Archivierung**
  Wie schon im allgemeinen Teil beschrieben, ist derzeit eine angemessene Konvertierung in entsprechende Archivierungsformate nicht ohne weitere durchführbar. Aus diesem Grunde müssen wir alle uns zur Verfügung stehenden Informationen bewahren. Dieses ist nur durch die Speicherung der Quellfile und aller benötigten Style- bzw. Usepackagesfile durchführbar.

Man erkennt nun, dass eine ‚Teilnahme' an Math*Diss* International (als Bibliothek, Institution oder als Einzelperson) mit einigen Vorleistungen verbunden ist. Aber welche Vorteile können dadurch gewonnen werden.
Die eingehenden Dissertationen werden in der angeschlossenen Allegro-Datenbank gespeichert, die mit den erzeugten Metadaten gefüttert wird. Diese sind auf der einen Seite die Metadaten, die von Der Deutschen Bibliothek (DDB) für die Abgabe einer elektronischen Dissertation gefordert werden, auf der anderen Seite werden auch automatisch die Metadaten nach Absprache mit der NDLTD-Initiative[23] erzeugt. Somit kann die Dissertation in einer internationalen und interdis-

[22] http://www.sub.uni-goettingen.de/
[23] http://www.ndltd.org

ziplinären Datenbank gefunden werden. Dieses ist von besonderer Bedeutung, denn wie aktuelle Meldungen besagen, besteht zwar die Möglichkeit auf eine elektronische Dissertation zeit- und ortsunabhängig zuzugreifen, jedoch leidet die Suche häufig an der Qualität des Retrievals, sei es bedingt durch die unspezifischen Suchoptionen (z.\,B.\ nur nach Autor und Titel) oder sei es durch verteilte Suchen auf unterschiedlichen Servern. Aufgrund der internationalen Anbindung und der fachspezifischen Auslegung des Services Math*Diss* International ist dieses Problem handhabbar.

Ein weiteres Bonbon für die Promovenden kann jedoch auch noch geboten werden. Aufgrund eines Abkommens mit dem Zentralblatt MATH[24] der European Mathematical Society (EMS)[25] werden alle Dissertationen, die den Anforderungen des Projektes Math*Diss* International resp. des Services genügen im Zentralblatt MATH erfasst und reviewed. Auch hiermit wird die internationale Kenntnisnahme der Arbeit ermöglicht.

Internationale Kooperation

Ein weiterer Arbeitsaspekt des Projektes ergibt sich schon durch die Namensgebung. Der internationale Anspruch soll erfüllt werden. Zu diesem Zweck wurden und werden internationale Kooperationen angestrebt. Neben den schon erwähnten Kooperationen mit NDLTD und somit auch mit der Open Archive Initiative (OAI)[26] sind auch bereits weitere Kontakte geschlossen worden. Hierzu zählt insbesondere das deutschsprachige Ausland.
Um eine Zusammenarbeit mit Österreich zu fördern wurden schon entsprechende Schritte eingeleitet, aber derzeit steht eine offizielle Entscheidung des Österreichischen Bundesministeriums noch aus. Auch in der Schweiz konnte mit der ETH Zürich ein kompetenter und hochrangiger Partner gefunden werden. Neben anderen Kontakten in nord- und osteuropäischen Staaten möchten wir hier die schon weit fortgeschrittene Zusammenarbeit mit Cellule MathDoc der Universität Grenoble vermerken, die eine jährliche Liste der mathematischen Dissertationen in Frankreich veröffentlichen[27].
Eine weitere Zusammenarbeit wurde bereits mit der Universität Uppsala vereinbart. Diese und weitere entsprechende Gespräche werden in Kürze geführt.

Fazit

Allgemein kann man sagen, dass mit Math*Diss* International ein Instrument für die Mathematik von besonderer Bedeutung für die internationale mathematische Forschung geschaffen wird. Im Vordergrund dieser Initiative steht der Benutzeraspekt. Die zentrale Organisationsstruktur dieses Services ermöglicht einen raschen Überblick über die aktuellen Forschungsergebnisse. Dieses Projekt ist jedoch nicht nur als mathematikspezifisches Forschungsvorhaben gedacht, sondern soll als Modell für den Aufbau eines fach- und formathomogenen Dokumentenservers auf internationaler Ebene dienen.

Wir laden alle interessierten Personen ein, die Entwicklung und den Aufbau diesen Service kritisch zu begleiten und möchten allen Promovenden der Mathematik, wo auch immer in der Welt, die Möglichkeit eröffnen, ihre Dissertation über den Math*Diss* International Server zu publizieren, die automatische Referenzierung im Zentralblatt MATH inbegriffen.

---

[24] http://www.zblmath.fiz-karlsruhe.de/MATH/home
[25] http://www.emis.de/
[26] http://www.openarchives.org
[27] http://www-mathdoc.ujf-grenoble.fr/these.html

Ansprechpartner

Prof. Dr. Günter Törner
toerner@math.uni-duisburg.de

Dipl.-Math. Thorsten Bahne
bahne@math.uni-duisburg.de

Gerhard-Mercator-Universität Duisburg
Fakultät 4 – Naturwissenschaften; Institut für Mathematik

BD Hans-J. Becker
becker@mail.sub.uni-goettingen.de

Dr. Thomas Fischer
fischer@mail.sub.uni-goettingen.de

SUB Göttingen

URL von Math*Diss* International: http://www.ub.uni-duisburg.de/mathdiss/