

Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7, D-14195 Berlin

W. Dalitz, M. Grötschel, G. Heyer, J. Lügger, W. Sperber

Informationsdienste für die Mathematik im Internet (Math-Net)

DFN-Vorhaben DT11

1.1.1997 - 31.12.1998

Vorhabenbeschreibung eines DFN-Projektes

Informationendienste für die Mathematik im Internet (Math-Net)

6. Dezember 1996

Redaktion: W. Dalitz, M. Grötschel, G. Heyer, J. Lügger, W. Sperber

Antragsteller: Konrad-Zuse-Zentrum für Informationstechnik (ZIB)
Projektleiter: Prof. Dr. M. Grötschel
Adresse: Konrad-Zuse-Zentrum für Informationstechnik
Takustraße 7
D-14195 Berlin-Dahlem
URL: <http://www.zib.de>
Telefon: 030/84185-210
Fax: 030/84185-269
E-Mail: groetschel@zib.de

Teilnehmende

Einrichtungen:

1. Technische Univ. Chemnitz-Zwickau, Dr. R. Haftmann
(für die Region Sachsen)
2. Technische Univ. München, Dr. M. Kaplan
(für Bayern)
3. Univ. Halle, Dr. L. Boltze
(für Sachsen-Anhalt, Thüringen, Hessen)
4. Univ. Kaiserslautern, G. Schmitt
(für Rheinland-Pfalz, Saarland, Baden-Württemberg)
5. Univ. Köln, Dr. J. Kallies,
(für Nordrhein-Westfalen I)
6. Univ. Osnabrück, Dr. R. Schwänzl,
(für Niedersachsen, Bremen)
7. Univ. Paderborn, Dr. W. Werner,
(für Nordrhein-Westfalen II)
8. Univ. Rostock, Dr. A. Straßburg,
(für Mecklenburg-Vorpommern, Hamburg, Schleswig-Holstein)
9. ZIB Berlin, W. Dalitz
(für Berlin, Brandenburg)

Förderung: Deutsche Telekom AG und DFN-Verein
Projektlaufzeit: 1. Januar 1997 – 31. Dezember 1998

1 Projektbeschreibung

1.1 Ausgangssituation

Die derzeitige Situation des Informationsaustausches in den Wissenschaften und damit auch in der Mathematik ist von zwei Haupttendenzen geprägt [GrötschelL 95a]:

Einerseits ist das traditionelle Publikationswesen durch die ständig steigende Anzahl wissenschaftlicher Journale bei gleichbleibenden bis sinkenden Bibliotheksetats an seiner Kapazitätsgrenze angelangt. Dadurch müssen die Bibliotheken immer mehr Journale abbestellen und jede einzelne Bibliothek kann nur einen immer kleineren Anteil an der Gesamtmenge wissenschaftlicher Journale vorhalten. Die Wissenschaftler sind dadurch immer mehr auf aufwendige und umständliche Fernleihen angewiesen. Außerdem werden die Auflagen der Journale immer kleiner, was bei weitgehend auflageunabhängigen Erstellungskosten zu einer Verteuerung des einzelnen Abonnements führt. An dieser Stelle schließt sich der Teufelskreis: Noch weniger Bibliotheken werden ein bestimmtes Journal abonnieren, wodurch der Preis pro Abonnement weiter in die Höhe getrieben wird.

Auf der anderen Seite haben die Wissenschaftler die neuen, elektronischen Medien für den Informationsaustausch entdeckt: Auf elektronischem Wege, sei es durch Bereitstellen auf einem WWW-Server oder durch Versenden per E-Mail, lassen sich wissenschaftliche Erkenntnisse schneller, einfacher und preiswerter verbreiten als auf dem langwierigen Weg über die traditionellen Journale. Dies geschieht bisher fast ausschließlich in Selbstorganisation der Wissenschaftler nach dem Grundprinzip des gegenseitigen Gebens und Nehmens im Internet. Nach einer Modellrechnung hätte das ZIB bei der Verbreitung seiner ca. 50 Preprints jährlich an ca. 300 Projektpartner in aller Welt nur Kosten von etwa DM 500,- gegenüber bisher DM 125.000,- für den Druck und Versand in Papierform [GrötschelL 95b].

Bis neue Entwicklungen in der Mathematik auch in wissenschaftlichen Journalen zu finden sind, vergehen nicht selten zwei bis drei Jahre. Wer in seinem Fachgebiet auf dem neuesten Stand sein und den Anschluß nicht verlieren will, kommt an den neuen Medien nicht mehr vorbei. In den neuen Medien haben sich allerdings noch keine Mechanismen zur Qualitätssicherung etabliert, wie sie in den traditionellen durch die Herausbergremien und Begutachtungsverfahren gegeben sind.

Was den neuen Medien außerdem fehlt, ist die Unterstützung der Benutzer beim gezielten Suchen nach bestimmter Information, etwa nach neueren Artikeln zu einem gewissen Thema oder von einem speziellen Autor. Die Artikel in traditionellen Journalen werden von den Fachinformationszentren erfaßt und sind damit, wenn auch mit einer weiteren Zeitverzögerung von einigen Monaten, suchbar. Elektronische Informationsangebote im Internet sind dagegen nur schwer auffindbar, wenn der Leser nicht schon zuvor weiß, wo er suchen muß.

Schon das Auffinden eines elektronischen Preprints in einem WWW-Server ist oft nicht einfach, weil jeder Server anders strukturiert ist. Daher weiß der Leser nicht, welchem Verweis in der Einstiegsseite er folgen muß, um die gesuchte Information zu finden (von der er aber immerhin wissen muß, daß sie irgendwo auf diesem Server liegt!). Nur sehr wenige

WWW-Server bieten bisher die Möglichkeit, mit einer Anfragemaske nach Dokumenten mit bestimmten Merkmalen zu suchen, etwa mit einem bestimmten Suchbegriff im Titel. Zur übergreifenden Suche im Internet gibt es bisher nur globale Ansätze, bei denen versucht wird, alle Dokumente im Internet in einem zentralen Suchindex zu erfassen. Dies führt bei dem derzeitigen raschen Wachstum nicht nur zu Kapazitätsproblemen bei der Erfassung der Informationsmenge und der Bearbeitung der Suchanfrage, sondern auch zu einer Übersättigung der Benutzer mit irrelevanter Information, zu einem "Information Overload" von Hunderten oder Tausenden von Treffern bei einer Suchanfrage, aus der man sich dann erst mühsam die wirklich interessanten Dokumente herausuchen muß. Es gibt nicht die Möglichkeit, die Suche auf ein bestimmtes Fachgebiet einzuschränken, sondern es kann nur das Internet als Ganzes abgesucht werden.

1.2 Ziele

Ziel dieses Projektes ist der Aufbau von kostengünstigen und nutzergesteuerten elektronischen Informationsdiensten, die den Wissenschaftlern vom Arbeitsplatzrechner aus den effizienten Zugang zu aktueller, qualitativ hochwertiger Volltext-, Fakten-, Literaturhinweis- und Software-Information im Bereich der Mathematik ermöglichen [GrötschelL 96]. Hierzu ist der Aufbau einer angemessenen organisatorischen und technischen Infrastruktur erforderlich, die darüberhinaus einen Modellcharakter sowohl für andere mathematische Institute und Fachbereiche als auch für andere Wissenschaftsgebiete haben soll.

Dazu sollen die bisherigen (relativ unorganisierten) Formen elektronischer Informationsversorgung und -bereitstellung in den Fachbereichen in flächendeckende Informations- und Kommunikationsdienste für die mathematische Wissenschaft eingebettet werden. Hierfür sollen, so weit möglich, die aktuellen Standardtechniken und -werkzeuge im Internet angewendet bzw. auf diese aufgebaut werden.

Die Ziele des Vorhabens sind im einzelnen:

- Vervollständigung der lokalen Informationsserver, so daß sie alle relevanten Informationen der jeweiligen Institution enthalten,
- Inhaltliche Strukturierung, Erschließung durch Metadaten und effiziente Organisation der lokalen Informationsangebote.
- Indexierung der lokalen Informationsserver zur Suche im gesamten lokalen Server, aber auch in bestimmten Teilmengen davon (z.B. nur Preprints).
- Automatisches Sammeln der lokalen Suchindexe in zentralen (regional und bundesweit) Komponenten, die eine verteilte Suche über alle beteiligten Institutionen erlaubt. Auch hier soll eine Auswahl bestimmter Teilmengen (z.B. nur die Preprints aller beteiligten Institutionen) möglich sein.
- Profildienste, die die Teilnehmer automatisch über neue Dokumente in den Fachgebieten informieren, für die sie sich subscribieren können.
- Aufbau eines Meta-Indexes mit Verweisen auf elektronische Informationsangeboten der Mathematik in aller Welt, der Server von mathematischen Institutionen,

Preprint-Archive, Software-Sammlungen, Such-Indexe etc. nachweist und als Einstiegspunkt für die Mathematik in aller Welt dient.

Am Ende des Projektes soll bei allen beteiligten Institutionen die Erschließung und der Zugriff auf sowohl die lokalen wie auf die globalen mathematischen Ressourcen in effektiver Weise möglich sein.

Das langfristige Ziel des Projektes ist es, im Bereich der Mathematik in Deutschland ein flächendeckendes, möglichst alle mathematisch relevanten Ressourcen erschließendes Informationssystem zu etablieren. Innerhalb dieses Projektes soll anhand ausgewählter mathematischer Fachbereiche in Deutschland beispielhaft ein solches Kernsystem aufgebaut werden, an dessen inhaltlichen und organisatorischen Strukturen sich andere Institutionen und Fachbereiche, auch in anderen Fachrichtungen, orientieren können und sollen. Deshalb erfolgt die Auswahl der vorgesehenen teilnehmenden Einrichtungen auch bewußt unter geographischen Gesichtspunkten, um innerhalb einer Region, bzw. eines regionalen Hochleistungsnetzes eine technische und organisatorische Anlaufstelle zu errichten.

Der Nutzen eines solchen inhaltlich und organisatorisch abgestimmten Vorgehens ergibt sich u.a. aus den folgenden Punkten:

- Bessere, d.h. insbesondere umfassendere, schnellere und kostengünstigere Informationsversorgung,
- Intensivierung des Informationsaustausches,
- Schnellere Umsetzung von Forschungsergebnisse innerhalb und außerhalb der Mathematik,
- Qualitätsverbesserung der Lehrangebote, Steigerung der Transparenz der Mathematikausbildung für Studenten.

1.3 Inhaltlicher Hintergrund

Die Projektteilnehmer haben alle an der Konzeption der geplanten Informationsdienste mitgewirkt und sind durch die Teilnahme am Projekt "Fachinformation"¹ der Deutschen Mathematiker-Vereinigung (DMV), das vom BMBF in den Jahren 1992-1995 gefördert wurde, für die Probleme der Informationsversorgung der Wissenschaft sensibilisiert. Das Ziel des Projektes der DMV war die flächendeckende Integration elektronischer Informationsdienste – insbesondere der fachrelevanten Datenbanken – in die Lehre und Forschung mathematischer Institutionen in Deutschland.

Im Verlauf des Projektes wurde jedoch deutlich, daß fachspezifische Information in einem größerem Zusammenhang betrachtet werden muß und heute mehr umfaßt als den Zugriff auf spezielle Datenbanken. Dies führte zu intensiven Diskussionen über die grundsätzlichen Probleme des wissenschaftlichen Informationsaustausches im Zeitalter der elektronischen Medien, in denen sich auch die Konzepte für neuartige Informationsdienste abzeichneten. Die Ergebnisse dieser Diskussionen sind in [DalitzGLS 94] zu einer Reihe

¹genauer Titel: "Verbesserung des benutzerorientierten Zugriffs auf fachspezifische Online-Datenbanken und CD-ROM für Mathematische Institute in der Bundesrepublik"

von Arbeitsansätzen zusammengefaßt. Das hier beantragte Projekt umfaßt den kommunikationstechnischen Teil dieser Aktivitäten.

Das Projekt "Fachinformation" hat wesentlich dazu beigetragen, daß in den mathematischen Fachbereichen mit dem Aufbau von eigenen WWW-Servern begonnen wurde. Der heutige Stand ist aber bei weitem (sowohl technisch als auch inhaltlich) nicht ausreichend. Das betrifft vor allem Aspekte der Orientierung in den einzelnen Systemen, die Suche und das Auffinden (auch über Systemgrenzen hinaus) von Informationen, die Konsistenz, die Strukturierung und die Menge des Informationsangebots. Etwa die Hälfte der 60 Mathematik-Server in Deutschland enthält bisher nur organisatorische Informationen. Es wurden erste Überlegungen und Versuche gemacht, wie diese Informationsangebote mit technischen Mitteln zusammengeführt und absuchbar gemacht werden können.

1.4 Informations- und kommunikationstechnische Beschreibung

Die neuen, übergreifenden Informationsdienste sollen auf der Basis des Internet und seiner heute weltweit verbreiteten und anerkannten Informationsdienste (World Wide Web, Hyper-G, WAIS, ftp und e-mail) aufgebaut werden. Sie basieren auf den örtlichen Informationsstationen, den *lokalen Komponenten*, die ein strukturiertes, aufeinander abgestimmtes und qualitativ hochwertiges Informationsangebot bereitstellen. Die Teilnehmer des Projektes stellen dafür in koordinierter Weise auf eigenen Informationsservern die folgenden Informationen bereit und machen sie dadurch untereinander und allgemein verfügbar:

- Preprints, Dissertationen, Habilitationen, ausgewählte Studien- und Diplomarbeiten im Volltext,
- wissenschaftliche Software- und Datensammlungen,
- Materialien zur Lehre, Vorlesungsskripte, Studienpläne und Prüfungsordnungen,
- organisatorische Informationen, wie Mitarbeiterverzeichnisse, Selbstdarstellungen der Institutionen, Veranstaltungsankündigungen, Kontaktadressen.

Diese lokalen Informationsangebote sollen durch geeignete Indexierungs- und Suchwerkzeuge absuchbar gemacht werden. Dabei soll nicht nur eine Suche in der gesamten lokalen Information möglich sein, sondern auch in deren Rubriken, wie Preprints, zumindest auf der obersten Ebene. Zur gezielten Suche müssen die in den Server eingebrachten Dokumente von den Partnern mit der zugehörigen Metainformation versehen werden, wie Autor, Titel, Datum, Stichworte und Klassifikation nach dem in der Mathematik verbreiteten MSC-Index. Dabei sollen bezüglich mathematischer Publikationen die Vorschläge zur Normierung der Metainformationen des Dublin Core [WeibelGMD 94] und des Warwick Framework [LagozeLD 95] berücksichtigt werden.

Es hat sich gezeigt, daß die Metainformation nicht automatisch erschlossen werden kann, sondern von den Informationsanbietern hinzugefügt werden muß. Die Eingabe der Metainformationen kann aber durch entsprechende Formularechnittstellen erleichtert werden.

Gleichzeitig sind die teilnehmenden Einrichtungen für die lokale institutionelle Koordination verantwortlich (z.B. durch Motivation der Mitarbeiter in den Fachbereichen und Einbindung der jeweiligen Bibliotheken).

Zusätzlich sollen *regionale und zentrale Komponente* aufgebaut werden, die für die übergeordneten Navigations- und Retrievalmöglichkeiten innerhalb dieser verteilt bereitgestellten Information und für die dafür notwendige Koordination sorgen. Eine allem übergeordnete zentrale Komponente vernetzt die lokalen und regionalen Informationssysteme zu einem Gesamtsystem, welches die lokalen Strukturen auf globaler Ebene widerspiegelt. Dadurch wird sowohl eine Navigation als auch eine verteilte Suche nicht nur über die Gesamtmenge, sondern auch global über gewisse Rubriken des Informationsangebotes (z.B. Preprints) möglich.

Mit anderen Worten: die globale Komponente bietet dem Benutzer nicht nur eine Übersicht über alle Informationssysteme der beteiligten Einrichtungen (in Form von Listen von Verweisen), sondern auch eine Übersicht über alle Preprintsammlungen dieser Einrichtungen, die auch absuchbar sein soll, etc.

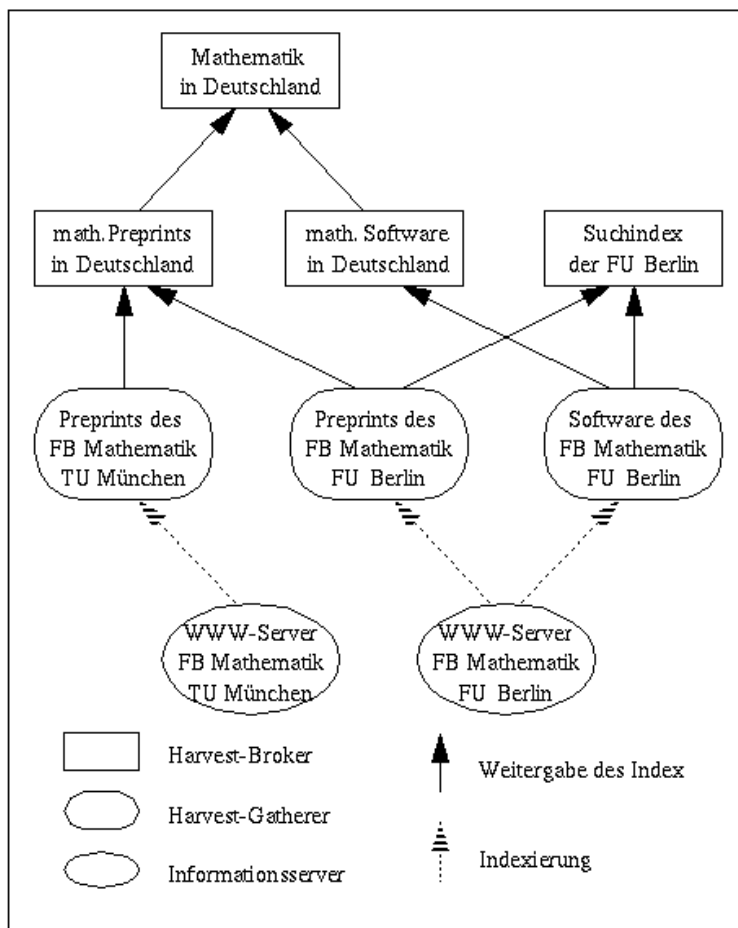


Abb. 1: Ausschnitt des geplanten Netzwerkes von Suchindexen

An dieser Stelle wird die Notwendigkeit der Koordination der lokalen Server deutlich: Eine derartige Vernetzung ist nur möglich, wenn es in jedem lokalen Server zu jeder definierten Rubrik einen Teilbaum gibt, unter dem alle Informationen zu dieser Rubrik, aber auch nur solche, zu finden sind.

In diese zentralen Komponenten sollen dabei alle potentiellen Informationsanbieter in der Mathematik wie Universitäten, außeruniversitäre Forschungseinrichtungen, mathematische wissenschaftliche Fachgesellschaften, Bibliotheken, Fachinformationszentren, wissenschaftliche Verlage, Softwarefirmen, etc. eingebunden werden können.

Eine weitere Aufgabe der zentralen Komponenten sind die Profildienste zur aktiven elektronischen Informationsverteilung im Sinne eines verallgemeinerten Konferenz- und Nachrichtensystems. Diese Profildienste ermöglichen die Subskription in den Rubriken des Gesamtsystems, um über Veränderungen in diesen benachrichtigt zu werden. Anstatt jeden Monat die gleiche Anfrage zu stellen, ob es neue Dokumenten in einem Fachgebiet gibt, können sich die Benutzer für ihre Interessengebiete subskribieren und werden dann automatisch über alle neuen Dokumente in diesen Gebieten informiert.

Als Einstiegspunkt zu mathematischen Informationsangeboten in aller Welt soll ein Meta-Index mit Verweisen auf die jeweiligen Informationsdienste geschaffen werden. Die Verwaltung dieses Meta-Indexes soll so weit wie möglich automatisiert werden, indem die Informationsanbieter ihre Informationsangebote selbst eintragen können und die URLs der eingetragenen Dokumente regelmäßig automatisch verifiziert werden. Informationsanbieter aus aller Welt sollen ihre Angebote selbst in den Meta-Index eintragen können, diese Eintragungen sollten aber von Moderatoren verifiziert werden.

Die Bereitstellung von Technik und die lokale Abstimmung reicht für das Projekt keineswegs aus. Um die Funktion des geplanten Systems langfristig zu sichern, ist der gleichzeitige Aufbau einer personellen und organisatorischen Infrastruktur für die beteiligten Einrichtungen erforderlich. Als Maßnahmen sind deshalb geplant:

- Schaffung der *Funktion des Informationsbeauftragten* bei den teilnehmenden Einrichtungen,
- regelmäßige *Workshops* (2 pro Jahr) und *Öffentlichkeitsarbeit*.

Die Informationsbeauftragten sollen möglichst aus dem akademischen Mittelbau kommen und unbefristete Stellen besetzen, so daß sie ihre Aufgabe dauerhaft als wissenschaftliche Dienstleistung in ihrer Institution wahrnehmen können. Langfristig kann diese Aufgabe auch von den Fachbibliotheken übernommen werden. Durch den verteilten Ansatz werden die hohen Folgekosten vermieden, die beim Aufbau zentraler Volltextspeicher entstehen. Nicht nur die lokalen Informationsserver sollen langfristig existieren. Die zentrale Komponente am ZIB wird zur Hälfte aus Eigenleistung entwickelt, so daß auch deren langfristige Betreuung sichergestellt werden kann.

Zur Koordination der Aktivitäten der Teilnehmer sollen mehrere Projekt-Workshops durchgeführt werden. Außerdem sollen die Ergebnisse des Projektes auf anderen Workshops und Tagungen vorgestellt werden und im Internet propagiert werden, um eine möglichst breite Öffentlichkeit zu informieren und ihr die Ergebnisse zugänglich zu machen. Mit der DMV soll ein Kooperationsvertrag geschlossen werden, um die Ergebnisse des Projektes auch anderen mathematischen Fachbereichen zur Verfügung zu stellen.

Das geplante Vorhaben soll mit parallelen Aktivitäten anderer Fachgesellschaften, wie der Deutschen Physikalischen Gesellschaft (DPG), der Gesellschaft Deutscher Chemiker (GDCh) und der Gesellschaft für Informatik (GI), abgestimmt werden [DMV+ 95] sowie auf der internationalen Ebene mit der International Mathematical Union (IMU), der European Mathematical Society (EMS) und der American Mathematical Society (AMS). Die DMV hat zum Ausbau der elektronischen Information und Kommunikation die Arbeitskreise "Inhalt" und "Technik" gebildet, an denen sich auch die Teilnehmer dieses Vorhabens beteiligen sollen.

Sowohl auf lokaler Ebene als auch bei den zentralen Komponenten wird eine Kooperation mit Bibliotheken, wie z.B der Universität Osnabrück (die gemeinsam mit dem dortigen FB Mathematik vorgehen will), Rechenzentren und dem Fachinformationszentrum Karlsruhe (FIZ) angestrebt. Insbesondere ist eine enge Zusammenarbeit mit der Staats- und Universitätsbibliothek Göttingen (SUB) geplant, deren Sammelschwerpunkt die Mathematik ist und die derzeit das DFG-Projekt "SSG-Fachinformationssystem" zur Erfassung von Sondersammelgebieten durchführt .

2 Projektplanung

2.1 Kommunikationstechnische Vorleistungen

Alle Teilnehmer verfügen über eine ausreichende Rechnerausstattung und Netzverbindungen und betreiben bereits WWW-Server, in denen ein mehr oder weniger großer Anteil der relevanten Information enthalten ist.

Einige Teilnehmer haben sich auch schon mit Indexierungswerkzeugen für das WWW beschäftigt und damit experimentiert. So hat z.B. die Universität Osnabrück *Harvest* installiert [PlümerS 96] und damit einen, allerdings noch unstrukturierten und unvollständigen, Preprint-Index für 38 Fachbereiche aufgebaut. Dieser Index wird jedoch zentral erzeugt und aktualisiert und belastet das Wissenschaftsnetz stärker als erforderlich, da zur Indexierung die vollständigen Dokumente über das Netz geholt werden müssen.

Am ZIB existieren Erfahrungen mit mathematische Software-Archiven, dem Aufbau von WWW-Servern und Meta-Indexen sowie mit automatischen E-Mail-Verteilern und einem objektorientierten Datenbanksystem (O2). So baut das ZIB zur Zeit den Informationsserver der *International Mathematical Union* (IMU) auf Basis von Hyper-G auf, der bereits einen Index der zugehörigen mathematischen Gesellschaften in aller Welt beinhaltet. Zur Überprüfung der URLs soll der WWW-Robot MomSpider [Fielding 94] eingesetzt werden.

2.2 Kommunikationssoftware

Für die lokalen Informationsangebote können die Teilnehmer jeden WWW-Server einsetzen, der mit dem Internetprotokoll HTTP arbeitet. Viele Institutionen, so auch einige Teilnehmer dieses Vorhabens, bauen bereits ihre Informationsangebote mit den WWW-Servern von NCSA oder CERN, etc. auf. Einige Teilnehmer, darunter das ZIB, haben beim Aufbau umfangreicher Informationsserver gute Erfahrungen mit dem Hyper-G-Server gemacht, der alle eingebrachten Dokumente automatisch in einer eingebauten Datenbank

indexiert und auch eine verteilte Suche über mehrere Hyper-G-Server [Maurer 96] erlaubt. Für die globale Indexierung soll daher eine Lösung gefunden werden, die alle Arten von WWW-Servern einbezieht und die auch für mögliche künftige Neuentwicklungen offen ist. Zur Erzeugung der lokalen Indexe soll das System *Harvest* [BowmanD 95] verwendet werden, das aus mehreren Komponenten (Gatherer, Broker, Replicator, Object Cache) besteht. Die Gatherer sammeln die Informationen von WWW-, Hyper-G- und FTP-Servern lokal oder über das Internet, indem sie von allen Dokumenten bestimmte Bestandteile extrahieren. Welche dies sind, kann durch die Konfiguration des Gatherers für jeden Dokumenttyp (HTML, Postscript) genau gesteuert werden. Dadurch kann Harvest sehr gut an die Erfordernisse der Mathematik angepaßt werden.

Die von den Gatherern gebildeten Suchindexe können von Brokern lokal oder auch überregional (zentral) zu umfassenderen Indexen zusammengefaßt werden. Auch können die Broker ihre Indexe wiederum an andere Broker weitergeben (s. Abb. 1 auf S. 4). Suchinterfaces erlauben dem Benutzer die Abfrage der von Gatherern und Brokern gebildeten Indexe und präsentieren die Ergebnisse.

Die Suche ist mit WWW-Clients über das HTTP-Protokoll möglich. Bezüglich der Suche in zentralisierten (regionalen und bundesweiten) Brokern soll die Möglichkeit des Einsatzes von Object Caches geprüft werden. Für den zentralen Gesamtindex soll am ZIB ein objektorientiertes Datenbanksystem (O2) erprobt werden, dessen WWW-Schnittstelle vielfältigere Möglichkeiten zur Anfrageformulierung und zur Präsentation der Suchergebnisse bietet als Harvest. Auch die Verwaltung der Teilnehmer am Profildienst soll experimentell damit erfolgen. Durch die Definition technischer Schnittstellen soll dafür gesorgt werden, daß bzgl. O2 keine langfristigen Abhängigkeiten entstehen.

2.3 Beschreibung des Arbeitsverlaufs

An jeder Einrichtung (einschließlich des ZIB) stehen 1/2 Wissenschaftliche Mitarbeiter-Stelle aus Projektmitteln und eine 1/4 Stelle eines wissenschaftlichen Mitarbeiters (Informationsbeauftragter) als Eigenleistung zur Verfügung. Am ZIB stehen zusätzlich 2 Wissenschaftliche Mitarbeiter zur Verfügung, davon 1 für das Projektmanagement aus Projektmitteln (in der folgenden Auflistung nicht aufgeführt) und 1 (je 1/2 aus Projekt- und Eigenmitteln) für den Aufbau der zentralen Komponenten.

Der folgende Arbeitsplan beschreibt alle Aufgaben, die nicht dem Projektmanagement zuzuordnen sind, d.h. alle Aufgaben der Fachbereiche und die technischen Aufgaben des ZIB.

Die in dem Projekt auszuführenden Arbeiten lassen sich in 7 Arbeitsfelder einteilen:

1. Informationsangebot (Relevanz, Vollständigkeit)
2. Informationsstruktur und Metadaten (Qualität der Erschließung)
3. Informationsvermittlung (Indexierer, Broker und deren Vernetzung)
4. Organisation der Informationsbereitstellung (Aktualität)
5. Zentrale Verweisstrukturen und Moderation (Qualitätssicherung)
6. Zentrales Archiv und Profildienst

7. Kooperationen und Öffentlichkeitsarbeit

Für die sechs geplanten Meilensteine ergeben sich daraus jeweils sechs Arbeitspakete für alle Projektpartner und ein für das ZIB spezifisches (jeweils die Nummer 6).

1. Meilenstein (4. Monat): konzeptionelle Vorbereitung

1. Erstellen eines Kataloges der für das Projekt relevanten Daten und zugehörigen Metadaten in Abstimmung mit den Fachbereichen.
2. Entwurf einer verteilten Informationsstruktur und von Formaten von Metadaten.
3. Installation, Erprobung und Anpassung der Harvest-Gatherer und -Broker in den Fachbereichen; Entwurf einer allgemeinen Konfiguration zur Indexierung und Suche; Konzeption eines Netzes von Gatherern und Brokern.
4. Erprobung und Anpassung von Werkzeugen zur Erfassung der Metadaten (z.B. Meta-Maker); Piloterfassung von Metadaten für Preprints.
5. Entwurf zentraler Verweisstrukturen zum Browsen und Suchen für die verschiedenen Informationstypen (Preprints, Software, etc.).
6. Erstellen eines Pflichtenheftes für das zentrale Archiv und das zentrale Nachrichtensystem (Profildienst) über Neueinträge.
7. Informationsveranstaltungen in den Fachbereichen, im DFN (Betriebstagung) und den kooperierenden Fachgesellschaften (Frühjahrs-Workshop).

2. Meilenstein (8. Monat): Implementierung eines Prototyps

1. Vervollständigen der lokalen Informationsserver um die projektrelevanten Informationen in Zusammenarbeit mit den Fachbereichen.
2. (Re-)Strukturierung der lokalen Informationsserver in Abstimmung mit den Fachbereichen; Ergänzen der Metadaten durch Fachinformationsbeauftragte.
3. Lokale Indexierung der Informationsserver mit Harvest; Aufbau lokaler Broker für verschiedene Arten von Informationen und eines lokalen Gesamtindex; Aufbau spezieller regionaler und bundesweiter Broker, z.B. für Preprints.
4. Vermittlung von Kenntnissen über Metadaten in den Fachbereichen mit dem Ziel der Erfassung der Metadaten durch die Autoren der Dokumente.
5. Aufbau zentraler Verweisstrukturen für alle projektrelevanten Informationstypen; Benennung von Moderatoren in Kooperation mit DMV und ggf. internationalen mathematischen Fachgesellschaften.
6. Entwurf von Datenbankschema und Benutzerinterface (WWW und E-Mail) für das zentrale Archiv und den Profildienst mit einem objektorientierten Datenbanksystem. (O2)

7. Präsentation der Konzepte auf der gemeinsamen Tagung der DMV und der Österreichischen Mathematischen Gesellschaft (ÖMG).

3. Meilenstein (12. Monat): Aufnahme des Probetriebs

1. Unterstützen anderer mathematischer Fachbereiche in den jeweiligen Regionen bei der Anpassung (Vervollständigung, Restrukturierung) ihrer Informationssysteme; Integration externer Informationen und Verweise vor Ort (fortlaufend bis zum Ende des Projektes).
2. Entwurf und Erprobung von Verfahren zur Konsistenzprüfung eingegebener Metadaten; Erfassung und Fortschreibung der Metadaten durch Autoren und Bibliothekskräfte.
3. Automatische Aktualisierung der Indexe (lokal, regional und bundesweit), d.h. der entsprechenden Broker-Hierarchien.
4. Vereinbaren organisatorischer Regelungen in den Fachbereichen, insbesondere mit Dekanaten und Bibliotheken, zur Bereitstellung hochschulzentraler Informationen (z.B. Personal- und bibliographische Informationen).
5. Integration erster Informationen aus internationalen Informationssystemen, z.B. Forschungssoftware aus der NetLib mit den entsprechenden Metadaten.
6. Implementierung des zentralen Archives und des Profildienstes auf Basis von O2.
7. Präsentation der Konzepte auf internationaler Ebene, z.B. bei der "European Mathematical Society" (EMS) und der "International Mathematical Union" (IMU).

4. Meilenstein (16. Monat): Evaluation und fachübergreifende Kooperation

1. Evaluation der Informationsstrukturen und Metadaten vor dem Hintergrund universitätsweiter Kooperation mit anderen Fachbereichen.
2. Entwurf von Verfahren zur Anpassung an fachübergreifende Informationsstrukturen und zur Umsetzung von Metadaten bzgl. der Kataloge zentraler Bibliotheken und Servern anderer Fachbereiche.
3. Aufbau und Erprobung fachübergreifender Harvest-Broker in Kooperation mit Universitäts-Rechenzentren, -Bibliotheken und -Verwaltungen.
4. Vorbereiten bzw. Vereinbaren spezieller, aber universitätsweiter Regelungen mit ausgewählten anderen Fachbereichen.
5. Entwurf und Erprobung von Verfahren zur Umsetzung in andere, internationale Informationssysteme, z.B. bibliographischer Information in andere Server wie DanEmir oder WWW-Robots.
6. Aufbau der Datenbasis für das zentrale Archiv; Integration mathematischer Informationsangebote aus aller Welt.

7. Präsentation der Konzepte als mögliche Basis für hochschulweite Informationssysteme; Kooperation mit vergleichbaren Nutzergruppen im DFN und im Hochschulbereich (ZKI-Verein).

5. Meilenstein (20. Monat): Optimierung und Konsistenzerhaltung

1. Erprobung von Werkzeugen und Verfahren zur Verbesserung des Laufzeitverhaltens, wie Caches im Harvest/Squid-Cache-Verbund, Proxy-Server und lokales Replizieren zentraler Broker.
2. Sicherung der Konsistenz der Informationen und ihrer Vernetzung durch Einsatz von Werkzeugen zur Überprüfung der URLs, wie z.B. Momspider.
3. Organisation der Überwachung der Betriebssicherheit des verteilten Netzes von Informations-Gatherern und -Brokern (Monitoring).
4. Einbeziehen weiterer Moderatoren zur Begutachtung von Informationen aus speziellen mathematischen Teilgebieten.
5. Erweitern des Spektrums externer Informationen und Anbieter.
6. Aufbau von Schnittstellen zur moderierten Eingabe von Ressourcen durch externe Benutzer.
7. Präsentation des verteilten Informationssystems der (deutschen) Mathematik und seiner zentralisierten Teile auf dem "International Congress of Mathematicians 1998" (ICM'98) in Berlin.

6. Meilenstein (24. Monat): Konsolidierung, Schlußbericht

1. Nutzungstudien und Statistik über den Gebrauch des Informationsangebotes.
2. Kritische Bewertung der Informationsstrukturen der Informationsserver der beteiligten Fachbereiche.
3. Technische Abschlußberichte (Systemdokumentation) für die Server (Broker) mit zentralisierten Aufgaben, wie z.B. regionale und bundesweite Preprint- und Dissertations-Server).
4. Erfahrungsberichte ausgewählter Moderatoren und anderer Personen, die an der zentralen Organisation beteiligt sind; Einschalten der Universitätsbibliotheken und des FIZ Karlsruhe.
5. Untersuchung der Präsenz des verteilten Informationssystems auf internationaler Ebene; ggf. Evaluation von Referenz-Counts bei globalen Systemen wie AltaVista.
6. Technischer Schlußbericht für die zentralisierten Teile (zentrales Archiv und Profildienst).
7. Wissenschaftliche Veröffentlichungen und Präsentation.

3 Kosten des Projektes

3.1 Aufbau der lokalen Informationsstationen

Bei den nachfolgende Berechnungen gehen wir davon aus, daß sich insgesamt 9 Institutionen über einen Zeitraum von 2 Jahren an den Teilaufgaben beteiligen, das ZIB sowohl als Partner in der Informationsbereitstellung als auch als Sitz der Projektleitung. Die am Vorhaben teilnehmenden mathematischen Institutionen schaffen die Funktion eines Informationsbeauftragten. Der Informationsbeauftragte trägt die Verantwortung für den Aufbau der lokalen Informationssysteme. Zu seinen Aufgaben zählen weiterhin die Abstimmung und Kooperation mit der Projektleitung zur Koordinierung der Informationen. Die Personalkosten für den Informationsbeauftragten für den Aufbau der lokalen Informationssysteme werden aus Eigenmitteln finanziert. Der Zeitfonds des Informationsbeauftragten für die im Projekt anfallenden Arbeiten wird mit 25% einer BAT IIa/Ib-Stelle angesetzt. Der Informationsbeauftragte wird während der Laufzeit des Projekts durch einen wissenschaftlichen Mitarbeiter unterstützt und entlastet. Die wissenschaftlichen Mitarbeiter (50% BAT IIa/Ib) werden aus Projektmitteln bezahlt. Die Fachbereiche können dazu alternativ im Umfang der dafür bereitgestellten Mittel auch studentische Hilfskräfte zu den Arbeiten einsetzen.

Die Bereitstellung einer Workstation für elektronische Information und Kommunikation durch die teilnehmenden Institutionen wird für die Teilnehmer am Projekt als Eigenleistung vorausgesetzt. Ebenso sind die Anbindung der Informationsstationen an das WIN-Netz mit einer akzeptablen Übertragungsgeschwindigkeit und die interne Vernetzung innerhalb der Institutionen Voraussetzungen für die Teilnahme am Projekt. Für zusätzliche Investitionen an den Informationsstationen werden den Institutionen in geringem Umfang Mittel zur Verfügung gestellt (3.000 DM zu Beginn des Projektes), z.B. für die Anschaffung zusätzlicher Speicherkapazitäten.

Es sind im Projekt zwei mehrtägige Workshops/Tutorials p.a. geplant, die der Diskussion und dem Erfahrungsaustausch dienen sollen. Den Institutionen sollten zum Besuch dieser Workshops sowie weiterer wichtiger Veranstaltungen zur elektronischen Information und Kommunikation Reisemittel zur Verfügung gestellt werden (3.500 DM p.a. pro Institution). Die Projektleitung als Organisatorin der Workshops soll die Kosten für die Vorbereitung und Durchführung der Workshops (z.B. Reisekosten für eingeladene Gäste) übernehmen. Dafür sind 5.000 DM p.a. vorgesehen.

Für die Berechnung der Personalausgaben wurden für die Jahre 1997 und 1998 die von der Berliner Senatsverwaltung für Finanzen festgelegten Durchschnittssätze angewandt:

1997: 108.740 DM und

1998: 111.460 DM, jeweils für eine BAT IIa/Ib-Stelle

Zur Vereinfachung wurden Mittelwerte von

110.100 DM für eine ganze Stelle

55.050 DM für eine halbe Stelle

27.525 DM für eine viertel Stelle

angesetzt.

Kosten pro Institution, einschließlich der lokalen Informationsstation am ZIB

Eigenmittel der Einrichtung		
1.) 1/4 wissenschaftlicher Mitarbeiter (Informationsbeauftragter)		
27.525 DM p.a.		55.050 DM
2.) Technische Ausstattung:		
Bereitstellung einer Workstation als Informationsstation		13.000 DM
3.) Reisekosten und Geschäftsbedarf 3.500 DM p.a.		7.000 DM
Summe		75.050 DM

DFN-Mittel für Unteraufträge an jede Einrichtung		
1.) 1/2 wissenschaftlicher Mitarbeiter 55.050 DM p.a.		110.100 DM
2.) Technische Ausstattung:		
Ausbau einer Workstation als Informationsstation		3.000 DM
3.) Reisekosten 3.500 DM p.a.		7.000 DM
Summe		120.100 DM

Gesamtkosten für den Aufbau der lokalen Informationsstationen

Eigenmittel der Einrichtungen		
1.) je 1/4 wissenschaftlicher Mitarbeiter (Informationsbeauftragter)		
9 x 27.525 DM p.a.		495.450 DM
2.) Bereitstellung der Informationsstationen		117.000 DM
3.) Reisekosten und Geschäftsbedarf 9 x 3.500 DM p.a.		63.000 DM
Summe		675.450 DM

DFN-Mittel		
1.) je 1/2 wissenschaftlicher Mitarbeiter 9 x 55.050 DM p.a		990.900 DM
2.) Ausbau der Informationsstationen		27.000 DM
3.) Reisekosten 9 x 3.500 DM p.a.		63.000 DM
Summe		1.080.900 DM
davon für Unteraufträge des ZIB an die acht anderen Institutionen		960.800 DM

Gesamtkosten lokale Informationsstationen

Summe Eigenmittel und DFN-Mittel		1.756.350 DM
----------------------------------	--	---------------------

3.2 Projektleitung und zentrale Komponente am ZIB

Als Leiter des Projekts ist Prof. Dr. M. Grötschel, Vizepräsident des ZIB und C4-Professor für Mathematik an der TU Berlin sowie Informationsbeauftragter der DMV, vorgesehen. Für seine Tätigkeit gehen wir von einem Zeitaufwand im Umfang von 10% seiner Arbeitszeit aus. Das ZIB sichert die Finanzierung dieser Funktion aus Eigenmitteln ab.

Die Aufgaben des Projektmanagers (1 Stelle BAT IIa/Ib) umfassen insbesondere alle organisatorischen und administrativen Aufgaben im Projekt, das Berichtswesen, die Ab-

stimmung und die Kommunikation mit den Fachbereichen und die Vorbereitung und Organisation der Workshops.

Für den Aufbau der zentralen Komponente des verteilten Informations- und Kommunikationssystems beantragen wir eine halbe Stelle eines wissenschaftlichen Mitarbeiters (BAT IIa/Ib). Das ZIB stellt hierfür Arbeitskraft im Umfang einer halben wissenschaftlichen Mitarbeiterstelle (BAT IIa/Ib) aus Eigenmitteln bereit.

Für Reisekosten der Projektleitung (u.a. zur Finanzierung der Workshops) werden 5.000 DM p.a. angesetzt.

Kosten für Projektleitung und zentrale Komponente am ZIB

Eigenmittel

1.) Projektleiter 15.153 DM p.a.	30.306 DM
2.) 1/2 wissenschaftlicher Mitarbeiter 55.050 DM p.a.	110.100 DM
3.) Technische Ausstattung: Bereitstellung von Workstations	24.000 DM
4.) Geschäftsbedarf, etc. 5.000 DM p.a.	10.000 DM
Summe	174.406 DM

DFN-Mittel

1.) Projektmanager 110.100 DM p.a.	220.200 DM
2.) 1/2 wissenschaftlicher Mitarbeiter 55.050 DM p.a.	110.100 DM
3.) Reisemittel (u.a. Workshops) 5.000 DM p.a.	10.000 DM
Summe	340.300 DM

Gesamtkosten Projektleitung und zentrale Komponente

Summe Eigenmittel und DFN-Mittel	514.706 DM
----------------------------------	-------------------

3.3 Übersicht über die Gesamtkosten

Eigenmittel

1.) für die lokalen Informationsstationen	675.450 DM
2.) für Projektleitung und zentrale Komponente	174.406 DM
Summe	849.856 DM

DFN-Mittel

1.) für Unteraufträge des ZIB an die acht anderen Institutionen	960.800 DM
2.) für die lokale Informationsstation am ZIB	120.100 DM
3.) für Projektleitung und zentrale Komponente	340.300 DM
Summe	1.421.200 DM

Gesamtumfang des Projektes

1. Deckung aus Eigenmitteln	849.856 DM
2. Deckung aus DFN-Mitteln	1.421.200 DM
Summe	2.271.056 DM

Alle genannten Beträge verstehen sich jeweils zuzüglich der geltenden gesetzlichen Mehrwertsteuer.

3.4 Gliederung der Mittel nach Verwendungszweck

Personalkosten

1.) je 1/4 wissenschaftlicher Mitarbeiter (Informationsbeauftragter) in den Institutionen 9 x 27.525 DM p.a.	495.450 DM
2.) je 1/2 wissenschaftlicher Mitarbeiter in den Institutionen 9 x 55.050 DM p.a.	990.900 DM
3.) Projektleiter 15.153 DM p.a.	30.306 DM
4.) Projektmanager 110.100 DM p.a.	220.200 DM
5.) 1/2 wissenschaftlicher Mitarbeiter (zentrale Komponenten, Eigenmittel) 55.050 DM p.a.	110.100 DM
6.) 1/2 wissenschaftlicher Mitarbeiter (zentrale Komponenten, DFN-Mittel) 55.050 DM p.a.	110.100 DM
Summe	1.957.056 DM

Investitionsmittel

1.) Bereitstellung je einer Workstation als Informationsstation (Eigenmittel)	117.000 DM
2.) Ausbau je einer Workstation als Informationsstation (DFN-Mittel)	27.000 DM
3.) Bereitstellung von Workstations für zentrale Komponenten (Eigenmittel)	24.000 DM
Summe	168.000 DM

Reisemittel und Geschäftsbedarf

1.) Reisekosten der Institutionen 9 x 3.500 DM p.a. (DFN-Mittel)	63.000 DM
2.) Reisekosten und Geschäftsbedarf der Institutionen 9 x 3.500 DM p.a. (Eigenmittel)	63.000 DM
3.) Reisekosten der Projektleitung 5.000 DM p.a.	10.000 DM
3.) Geschäftsbedarf der Projektleitung 5.000 DM p.a.	10.000 DM
Summe	146.000 DM

Literatur

[BowmanD 95] C.M. Bowman, P. Danzig et al.; Harvest: A Scalable, Customizable Discovery and Access System; Techn. Report CU-CS 734-94, Dept. of Computer Science,

- [DalitzGLS 94] W. Dalitz, M. Grötschel, J. Lügger, W. Sperber; Neue Perspektiven eines Verteilten Informationssystems für die Mathematik; Rundbrief der GAMM; Sept. 1994; 8 - 26
- [DMV+ 95] Kooperationsvereinbarung
zwischen der Deutschen Mathematiker-Vereinigung (DMV), der Deutschen Physikalischen Gesellschaft (DPG), der Gesellschaft Deutscher Chemiker (GDCh) und der Gesellschaft für Informatik (GI) zur elektronischen Information und Kommunikation; DMV-Mitteilungen, Heft 3, 1995, S. 6-11
- [Fielding 94] R. Fielding; Maintaining Distributed Hypertext Infostructures: Welcome to MOMSpider's Web; 1. Int. WWW Conf., Genf, 1994
- [GrötschelL 95a] M. Grötschel, J. Lügger; Wissenschaftliche Kommunikation am Wendepunkt – Bibliotheken im Zeitalter globaler elektronischer Netze; Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) 42, Heft 3, Mai/Juni 1995, 287-312
- [GrötschelL 95b] M. Grötschel, J. Lügger; Aufbau elektronischer Informations- und Kommunikationsstrukturen; in Wolfram Neubauer (Hrsg.) Deutscher Dokumentartag 1995, Proceedings einer Konferenz an der Fachhochschule Potsdam, 26.–28. September 1995, Deutsche Gesellschaft für Dokumentation, Frankfurt, 1995, 13-58
- [GrötschelL 96] M. Grötschel, J. Lügger; Neue Produkte für die digitale Bibliothek: die Rolle der Wissenschaften; in Die Deutsche Bibliothek, Bundesvereinigung Deutscher Bibliotheksverbände und Börsenverein des Deutschen Buchhandels (Hrsg.): Die unendliche Bibliothek. Digitale Information in Wissenschaft, Verlag und Bibliothek; Harrassowitz Verlag, Wiesbaden, 1996, 38–67
- [LagozeLD 96] C. Lagoze, C.A. Lynch, R. Daniel jr.; The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata; <http://cs-tr.cs.cornell.edu/Dienst/UI/2.0/Describe/ncstrl.cornell%2fTR96-1593>
- [Maurer 96] H. Maurer (Hrsg.); HyperWave: The Next Generation Web Solution; Addison-Wesley, Reading (Mass.), 1996
- [PlümerS 96] J. Plümer, R. Schwänzl; Harvesting Mathematics; <http://www.mathematik.uni-osnabrueck.de/projects/harvest/harvest.html>
- [WeibelGMD 95] S. Weibel, J. Godby, E. Miller, R. Daniel jr.; OCLC/NCSA Metadata Workshop Report, Dublin, Ohio, 1995 http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html