

---

Konrad-Zuse-Zentrum für Informationstechnik Berlin  
Heilbronner Str. 10, D-10711 Berlin - Wilmersdorf

**Joachim Lügger**

**Design-Überlegungen  
für ein Verteiltes Informationssystem  
für die Mathematik in Deutschland**

---

Technical Report TR 95-9 (August 1995)

**Design-Überlegungen  
für ein Verteiltes Informationssystem  
für die Mathematik in Deutschland <sup>1</sup>**

Joachim Lügger  
Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)  
Heilbronner Str. 10, 10711 Berlin-Wilmersdorf  
luegger@zib-berlin.de

Abstract

Dieser Artikel beschreibt die ersten Schritte auf dem Wege zu einem Verteilten Informationssystem für die Mathematik, das von der “Deutschen Mathematiker-Vereinigung” (DMV) vorbereitet wird. Die Darstellung erfolgt aus zwei unterschiedlichen Perspektiven. Der erste Teil soll vor allem Lesern, die an der Motivation des Vorhabens und den organisatorischen Rahmenbedingungen interessiert sind, einen Zugang zu den zugrundeliegenden technischen Ideen bieten. Er skizziert zunächst die wissenschaftliche Zielsetzung des geplanten Vorhabens. Dann diskutiert er das Verteilte Informationssystem als Kombination aus einem elektronischen Nachrichtensystem (auf der Basis von E-mail), einem globalen Hypertextsystem (das die Informationsangebote der Partner des Projekts miteinander vernetzt) und einer verteilten, objektorientierten Datenbank (mit Interfaces zu den bereits genannten Komponenten). Dieser Teil schließt mit einer knappen Betrachtung der Verhältnisse (auch der Chancen, Kosten einzusparen), die beim Übergang von traditionellen zu elektronischen Publikationsformen zu erwarten sind. Er führt insbesondere neue Möglichkeiten an, das geplante System auch für Zwecke des “Zentralblatt für Mathematik” einzusetzen.

Der zweite Teil richtet sich an den eher technisch interessierten Leser. Er diskutiert einen ersten Vorschlag zur Strukturierung der bei den Partnern verteilt vorliegenden Informationsangebote – vorwiegend aus Nutzersicht, d.h. möglichst unabhängig von den Idiosynkrasien der heute im Internet gegebenen Informationssysteme. Er enthält die technischen Kerngedanken zur Realisierung einer homogenen Informationsinfrastruktur bei den Partnern, die ein verteiltes und gleichzeitig hierarchisches Retrieval in den Informationsangeboten der Mathematik in Deutschland ermöglichen soll. Dieser Teil benennt erste Chancen zur Implementierung innerhalb der Hyper-G- und, als Fernziel, auch der WWW-Welt. Er zeigt aber, neben den neuen Möglichkeiten, auch Risiken und Grenzen der Machbarkeit mit heutigen Mitteln auf.

Die gesamte Darstellung orientiert sich an den Inhalten der bereits an anderer Stelle ausführlich vorgestellten Teilaufgaben und Teilprojekte und leitet daraus sowohl die vorgeschlagene Informationsstruktur als auch die technischen Mittel ab. Die wissenschaftlichen Inhalte und organisatorischen Ziele des Verteilten Informationssystems wurden in einer breiten (auch elektronisch geführten) Diskussion im Kreise der potentiellen Partner des Projekts ermittelt, insbesondere der Fachinformationsbeauftragten der Mathematik. Eine Anlage stellt eine vergleichende Betrachtung der Kosten auf, die einem mathematischen Fachbereich oder Forschungsinstitut bei der Weitergabe seiner wissenschaftlichen Reports auf Papier bzw. auf elektronischem Wege entstehen.

---

<sup>1</sup>Version vom 1. Juni 1995

## Teil I. Organisatorisch-technische Rahmenbedingungen

1. Wissenschaftliche Zielsetzung, Partner des Projekts und Aktivitäten
2. Math-Net: Eine integrierende Basis für das Verteilte Informationssystem
  - 2.1 Math-Net als elektronisches Nachrichtensystem, aktives Informationsangebot
  - 2.2 Math-Net als globales Hypertextsystem, örtliches Informationsangebot
  - 2.3 Math-Net als Datenbank, zentralisiertes Informationsangebot
  - 2.4 Zusammenwirken – dezentrale und zentralisierte Information
3. Übergang zur elektronischen Fachinformation – Math-Net und Zentralblatt
  - 3.1 Von traditionellen zu elektronischen Publikationsformen
  - 3.2 Zentralblatt und Math-Net
    - 3.2.1 Allgemeine Zielsetzungen in bezug auf Fachinformation
    - 3.2.2 Erweiterung des Math-Net auch für traditionelle Publikationen

## Teil II. Technische Struktur, Math-Net

4. Das Skelett der Informationsstruktur, Hierarchie der Kollektionen
  - 4.1 Teilaufgaben und Teilprojekte
  - 4.2 Verteilte Kollektionen von Informationen
  - 4.3 Ein Vorschlag zur Strukturierung des Informationsangebots
  - 4.4 Anmerkungen zur Implementierung mit WWW- oder Hyper-G-Servern
  - 4.5 Suchstrategien in Hypertextsystemen
    - 4.5.1 “Lost in Hypertext Syndrom” – erneut
    - 4.5.2 Hierarchische Suche in hierarchischen Kollektionen
5. Hierarchien und Klassifikationsschemata, Automatisierungsfragen
  - 5.1 Hierarchische Kollektionen und MSC-Index im Hypertext-Teil
  - 5.2 Hierarchische Kollektionen und MSC-Index im Nachrichtenteil
6. Architektur des Verteilten Informationssystems
  - 6.1 Zentralisierter Teil der Math-Net-Architektur
  - 6.2 Integration von Information in das Math-Net via E-mail
  - 6.3 Distribution des zentralisierten Teils des Math-Net
7. Zusammenfassung, Leitlinien und Designprinzipien
8. Ursprung und Dank

## Literatur und elektronische Referenzen

## Anhang

- A. Vergleichende Kostenbetrachtung für die Weitergabe von Reports durch mathematische Institutionen

## 1 Wissenschaftliche Zielsetzung, Partner des Projekts und Aktivitäten

Dieser Artikel beschreibt das technisch-organisatorische Konzept des Verteilten Informationssystems für die Mathematik in Deutschland. Die wissenschaftlich-technische Zielsetzung ist an anderer Stelle bereits ausführlich dargelegt [DalitzGLS 94] und soll hier nur noch auszugsweise skizziert werden (siehe auch [Grötschel 94], [DalitzGLS 95]).

Auf der technischen Grundlage des Internet und seiner weltweit verbreiteten Informationsdienste (WorldWideWeb, Hyper-G, Gopher, ftp und E-mail; siehe z.B. [BernersCGP92], [Kappe93b], [Krol94]) soll ein verteiltes, mathematisches Informationssystem geschaffen werden, bei dem die Partner des gemeinsamen Vorhabens ihre örtlichen Ressourcen untereinander und zugleich weltweit verfügbar machen. Offenheit im Internet ist einer der obersten Leitgedanken des Verteilten Informationssystems. Sie ermöglicht die Nutzung der umfangreichen, bereits im Internet vorhandenen und zumeist kostenfrei bzw. kostengünstig verfügbaren elektronischen Ressourcen.

Die “Deutsche Mathematiker-Vereinigung” (DMV) plant die Realisierung des *Verteilten Informationssystems* (wie es ab jetzt kurz bezeichnet werden soll) als Querschnittsaktivität für die Mathematik, die nicht allein auf Deutschland beschränkt ist. Gleichwohl strebt die DMV für die kommenden Jahre eine Förderung des Projekts durch den Bundesminister für Bildung, Wissenschaft, Forschung und Technologie (BMBF) und Organisationen der Forschungsförderung an [DMV 95].

Als Partner kommen nicht nur die mathematischen Fachbereiche und Forschungsinstitute in Betracht, sondern auch mathematische Forschungslabors aus der Industrie und die Informationsversorger der Mathematik, insbesondere die wissenschaftlichen Verlage, die “Technische Informationsbibliothek Hannover” (TIB Hannover), die “Staats- und Universitätsbibliothek Göttingen” (SUB Göttingen) mit ihrem mathematischen Archiv, die Bibliotheken der Universitäten und das “Fachinformationszentrum Karlsruhe” (FIZ Karlsruhe) mit dem “Zentralblatt für Mathematik und ihre Grenzgebiete” (Zentralblatt).

Darüber hinaus können auch Museen mit naturwissenschaftlich-technischen Abteilungen, internationale Anbieter von Forschungssoftware und fachspezifische Kompetenzzentren, wie z.B. das Euromath Center in Kopenhagen und “Computer Algebra Netherlands” (CAN) und sogar einzelne Mathematiker mit besonderem Interesse an Teilprojekten (wie z.B. am Lebenden Mathematischen Museum) an dem Projekt teilnehmen. Nicht zuletzt werden sich auch die DMV selbst und ihre Fachgruppen am Verteilten Informationssystem beteiligen.

Die DMV strebt mit wissenschaftlichen Gesellschaften verwandter Ausrichtung bzw. Zielsetzung, wie z.B. mit der “Deutschen Physikalischen Gesellschaft” (DPG) [DPG 94], der “Gesellschaft Deutscher Chemiker” (GDCh), der “Gesellschaft für Informatik” (GI) [GI 94], und mit der “Gesellschaft für Angewandte Mathematik und Mechanik” (GAMM) eine enge Kooperation an. Weitere – insbesondere internationale – Forschungsorganisationen sowie Träger der deutschen Forschungsförderung, wie z.B. die “Deutsche Forschungsgemeinschaft” (DFG) und der BMBF sind ebenfalls dazu eingeladen, sich an der Ausgestaltung des Projekts zu beteiligen.

Als Aktivitäten – mit Focus auf mathematische Fachinformation – sind folgende Teilaufgaben und Pilotvorhaben bei allen Partnern geplant:

- \* Einrichtung von *elektronischen Informationsstationen* (Server & Clients)
- \* *Verteiltes* elektronisches Angebot von *Preprints* und *Skripten* (Volltexte)
- \* *Verteiltes* elektronisches Angebot von *Software- und Datensammlungen*
- \* Schaffung einer Organisationsform für *elektronisches Reviewing*
- \* Entwicklung eines organisatorischen Konzepts für *elektronische mathematische Journale*
- \* Zugriff auf globale *Informationssysteme in der Mathematik*
- \* Zugriff auf *elektronische Kataloge von Bibliotheken*
- \* Erprobung von Verfahren zur *elektronischen Lieferung von Dokumenten*
- \* Elektronisches Angebot von eingescannten *historischen Dokumenten und Büchern*
- \* Organisation *Lebenden Mathematischen Museums*
- \* *Elektronische Projektorganisation*

Der Zugriff auf diese Informationen – einschließlich aller Volltexte – soll kostenfrei bzw. kostengünstig von jeder Workstation aller Partner bzw. von allen Workstations im Internet möglich sein. Diese Ziele sind durch konsequenten Einsatz von Software erreichbar, die nach dem Client-Server-Modell strukturiert und die im Internet verbreitet und bekannt ist.

Die Bereitstellung von Technik reicht keineswegs aus. Ohne die Schaffung einer personellen, organisatorischen und technischen Infrastruktur ist die Durchführung eines solchen Projekts nicht möglich. Zentrale Konzepte der Informationsversorgung sollen dabei durch dezentrale ergänzt werden. Als Maßnahmen sind – für alle Partner – geplant:

- \* Schaffung der *Funktion des Informationsbeauftragten*
- \* Einrichtung eines gemeinsamen *Forums für mathematische Information*
- \* Teilnahme an nationalen und internationalen *Standardisierungsmaßnahmen*
- \* *Workshops* und regelmäßige Schulungen, *Öffentlichkeitsarbeit*

Das Forum für mathematische Information besteht aus der Versammlung der Informationsbeauftragten, einer zentralen Projektleitung und einem *wissenschaftlich-technischen Ausschuß*. Ihm ist eine *wissenschaftlich-technische Arbeitsgruppe* zugeordnet.

Die langfristigen Ziele dieses Vorhabens der DMV sind:

- \* Ergänzung der bisherigen zentralen Informationsversorgung durch ein dezentrales, flächendeckendes und interaktives Informations- und Kommunikationssystem für Wissenschaft und Industrie.
- \* Erschließung der gesamten elektronischen Fachinformation (Volltexte, Software, Datensammlungen etc., Kontaktangebote), auch durch internationale Kooperation.
- \* Aktive Beteiligung von Mathematikern aus Wissenschaft und Industrie an der elektronischen Versorgung mit Fachinformation.
- \* Effizienzsteigerung der aktuellen Forschung durch Intensivierung des elektronischen Informationsaustausches.
- \* Verbesserung der Lehrangebote, Steigerung der Transparenz für Studenten, Studienzeitverkürzung.

- \* Beschleunigung in der allgemeinen Informationsversorgung, Kostendämpfung.
- \* Erprobung neuer elektronischer Vertriebswege.

Das Vorhaben zielt ab auf die Schaffung einer informationstechnischen Infrastruktur für Datenbank-Retrieval, E-mail, “electronic conferencing” und fachspezifische Informationsnetze. Der Schwerpunkt liegt auf der Mathematik in Deutschland, der allgemeine Rahmen ist jedoch weiter gesteckt. Das Projekt ist offen gegenüber Diskussion, Koordination und Kooperation mit Partnern in anderen Wissenschaftsbereichen, in der Industrie bzw. in anderen Ländern. Dieses Projekt soll insbesondere auch als Beispiel für die Entwicklung anderer Modelle elektronischer Information und Kommunikation dienen. Darüber hinaus soll das Projekt den Beitrag der deutschen Mathematik zum weltweiten “give and take” in der Gemeinschaft des Internets verbessern helfen.

## **2 Math-Net: Eine integrierende Basis für das Verteilte Informationssystem**

Ein Kerngedanke des Verteilten Informationssystems ist, daß die Partner des Projekts ihre eigenen Informationsangebote – jeweils vor Ort und auf eigenen Servern – selbst aufbauen und im Internet anbieten; zur Thematik des Publizierens mittels elektronischer Netze siehe auch [Ginsparg 94], [Stix 95] und [Grötschel 95a]. Die schiere Menge an schon heute auf die eine oder andere Weise elektronisch vorliegendem, mathematischen Material verbietet eine zentralistische Lösung. Die Information muß dort bereitgehalten und aktualisiert werden, wo sie entstanden ist.

Die mathematische Information ist also über viele Institutionen “verteilt”. Wie soll sich aber der Wissenschaftler und “Leser” solcher Information nun in der verwirrenden Welt des Internet orientieren und insbesondere die für ihn interessante, aber über viele Orte verstreute Information herausfinden? In aller Regel sucht er Informationen zu einem bestimmten Thema und will dafür nicht erst eine große Zahl von Servern durchsuchen.

Die örtliche Information soll deshalb zur leichteren Auffindbarkeit in Kollektionen desselben Typs, die mit Datenbankmitteln durchsuchbar sind, vielstufig hierarchisch gegliedert und systematisch klassifiziert sein. Für die Klassifizierung gibt es zwei in der Mathematik weltweit akzeptierte Schemata, das “Mathematics Subject Classification Scheme” (MSC-Index) für Publikationen [FIZ-Ka 94] und den “Guide to Available Mathematical Software” (GAMS-Index) des amerikanischen Normeninstituts; siehe [NIST]. Diese “vertikalen” Informationshierarchien sollen “horizontal” miteinander vernetzt sein, indem alle Kollektionen desselben Typs, z.B. alle Kollektionen mit Preprints, (auf jeder Hierarchieebene) durch elektronische Verweise (Hypertext-Links) miteinander verbunden sind. Mit ihrer Hilfe sollen Nutzer “quer durch Deutschland” nach speziellen Inhalten suchen können.

Wie kann nun aber der Wissenschaftler erfahren, ob das System für ihn neue Informationen enthält? Er kann ja in der Regel nicht wissen, wann eine ihm fremde Institution neue Dokumente einbringt, und er will ja auch nicht immer die gesamte (und wohl bald sehr umfangreiche) Information durchsuchen müssen, was ihm eine regelmäßige Initiative abverlangt. Das Verteilte Informationssystem soll deshalb mit einer Art Nachrichtenagentur gekoppelt werden, die in gleicher Weise gegliedert ist. Ein Leser grenzt dann mit Hilfe der Hierarchie- und Klassifikationsmerkmale sein Interessengebiet präzise ein, schon um nicht mit irrelevanten Informationen überschwemmt zu werden. Er soll dann elektronische

Nachrichten über diejenigen neuen Informationen erhalten, zu denen er auf diese Weise “subskribiert” hat.

Dieser Artikel beinhaltet – technisch ausgedrückt – den Vorschlag, ein die “dezentralen” Informationsangebote integrierendes, elektronisches System aus

- \* einem elektronischen Nachrichtensystem auf E-mail-Basis,
- \* einem globalen Hypertextsystem im World Wide Web und
- \* einer verteilten (objektorientierten) Datenbank

aufzubauen, das Math-Net. Es soll in erster Linie mathematische Informationen erschließen und verbreiten. Wir werden aber sehen, daß sich das System in gleicher Weise auch für die elektronische Vermittlung von Ankündigungen, Nachrichten, Bewertungen und Meinungen über traditionelle Veröffentlichungen in der Mathematik einsetzen läßt.

Das Verteilte Informationssystem soll sich insbesondere dafür eignen, die Gemeinschaft der Mathematiker in den Prozeß der Informationsvermittlung stärker einzubeziehen. Der Begriff “verteilt” beinhaltet nicht nur eine dezentrale Organisationsform, sondern auch dezentrale Verantwortung. Welche Information in das Netz eingebracht wird, entscheidet die jeweilige Institution und nicht allein der Autor eines Dokuments. Das hat viele Vorteile, auch rechtlicher Natur. Das klärt z.B. zugleich auch die Frage, ob ein Dokument im Verteilten Informationssystem “offiziell” bzw. (zumindest) den qualitativen Ansprüchen der bereitstellenden Institution genügt.

Die elektronische Publikation von personenbezogener Information, wie z.B. die Bekanntgabe von Forschungsprojekten und deren Teilnehmern (Kontaktadressen), hat ferner – nicht nur in Deutschland – eine datenschutzrechtliche Komponente, die nur im Binnenverhältnis zwischen der jeweiligen Institution und der betroffenen Person eindeutig zu klären ist. Die meisten Entscheidungen darüber, welche Information im Netz bereitgestellt wird, in welcher Art und zu welchem Zeitpunkt (Sicherung der Aktualität), werden i.d.R. nur vor Ort getroffen und koordiniert werden können. Nicht zuletzt deshalb bedarf das gesamte System einer personellen und organisatorischen Infrastruktur.

Dezentralisierung bedeutet aber nicht, daß auf zentrale Komponenten ganz verzichtet werden soll. Der Kreis der am Verteilten Informationssystem teilnehmenden Institutionen wird zunächst schon aus pragmatischen Gründen recht begrenzt sein. Dennoch sollen auch “externe Partner” die Möglichkeit erhalten, wichtige Informationsangebote, wie z.B. Adressen ihrer ftp-Archive oder Verweise auf ihre “Home Page” im World Wide Web oder auch auf Teilbereiche ihres lokalen Webs, im Math-Net zu “registrieren” und damit in das Verteilte Informationssystem einzubringen, ohne daß sie gezwungen sind, ihr gesamtes Angebot nach den Regeln des Verteilten Informationssystems durchzustrukturieren. Es soll genügen, wenn sie ihre Informationsangebote durch Referenzdokument vertreten lassen.

Diese Art der “Essenz” mathematischer Information (Abstracts, Referate, Verweise auf elektronische “Standorte” von Dokumenten, Software und Bilder etc.) soll im Math-Net zentralisiert behandelt, also zusammengeführt und ebenfalls absuchbar erschlossen werden. Insofern wird das Math-Net eine zentralisierte Komponente enthalten, den Math-Net-Kern. Der Begriff “zentralisiert” bringt zum Ausdruck, daß es sich hierbei nicht um eine zentrale Komponente handeln soll, etwa im Sinne einer zentralen Datenbank. Die auf diese Weise zusammengeführte Information soll nämlich selbst bei allen Partnern des Projekts direkt verfügbar sein, d.h. sie soll selbst wieder verteilt organisiert werden.

Der Math-Net-Kern wird damit ein integraler Bestandteil des Math-Net, dem technischen Teil des Verteilten Informationssystems. Der Unterschied zwischen dem Math-Net und

dem Math-Net-Kern liegt – für die Partner des Projekts – vorwiegend in der technischen Art der Einbindung der mathematischen Information bzw. in den technischen Mitteln der Realisierung des Verteilten Informationssystems. Er soll für sie idealerweise möglichst verschwinden. Der Artikel verwendet deshalb im folgenden nur noch den Ausdruck Math-Net, auch für den Math-Net-Kern, – wenn aus dem Kontext klar wird, daß der Math-Net-Kern gemeint ist.

Der Begriff “mathematische Information” ist hier nicht im begrenzten Sinne traditioneller Fachinformation zu verstehen. Er umfaßt in dieser Arbeit aber auch Information “auf Papier”. Denn das Math-Net soll sich auch für Aufgaben und Ziele des Zentralblatt eignen, zunächst bzgl. elektronischer Information und – in einer zweiten Stufe – auch für Referate zu traditionellen Publikationsformen.

Das Math-Net soll, neben der Vermittlung rein fachlicher Information auch organisatorische Inhalte (e.g. Kontaktadressen) zugänglich machen und die Partner insgesamt organisatorisch unterstützen. Die DMV verfolgt mit dem Math-Net insbesondere das Ziel, die Redaktion des Zentralblatt durch Einsatz elektronischer Mittel wirkungsvoll zu unterstützen.

Die technischen Eigenschaften des Math-Net sollen jedoch nicht durch die Gegebenheiten heutiger Hypertextsysteme eingeschränkt oder gar auf das Zentralblatt spezialisiert, sondern in gewisser Weise universell sein. Mit dem Math-Net soll ein integriertes Informations- und Kommunikationssystem entstehen, auf dessen technischer Grundlage im Prinzip alle Teilaufgaben und Teilprojekte der Partner des Projekts wirkungsvoll unterstützt werden. Das Design richtet sich nach den Anforderungen seiner Nutzer und ihrer Vorhaben aus und nicht nach den technischen Randbedingungen und Eigenarten der heute verbreiteten Informationssysteme.

Das Math-Net soll sich nicht nur für die Distribution von Nachrichten eignen, sondern auch für die Verbreitung mathematischer Texte und Paper (etwa in der Art des Ginsparg-Systems des “Los Alamos National Laboratory” [Ginsparg 94] [LANL] im Bereich der Hochenergiephysik) bzw. von Ausgaben elektronischer Journale in anspruchsvoller Form (e.g., als  $\text{\LaTeX}$  - oder Postscript-Files). Es soll auch für die unmittelbare, fachliche Kommunikation aller Partner des Projekts einsetzbar sein, d.h. für fachliche und organisatorische Fragen an die Community und Antworten dazu. Denkbar ist insbesondere, daß die Referenten und die Redaktion des Zentralblatt auf diese Weise mit der Mathematischen Community in engere Kommunikation eintreten und auch mit anderen Teilen ihrer Redaktion, sofern sich eine Arbeitsteilung auf europäischer Basis ergeben sollte.

Für das technische Design des Math-Net ist der Gedanke der Dezentralisierung wesentlich. Allein schon der Umfang und die Vielfalt der Informationen, die heute bei den mathematischen Fachbereichen elektronisch verfügbar sind (Volltexte), lassen eine zentrale Speicherung (in einer einzigen zentralen Datenbank) wenig sinnvoll erscheinen. Der Hauptgrund für die dezentrale Anlage liegt jedoch im organisatorischen Bereich. Man denke z.B. daran, daß, falls das Verteilte Informationssystem akzeptiert wird, sich ein großer Teil der Kommunikation vermutlich vor Ort (innerhalb der Universitäten) bzw. innerhalb der Region abspielen wird (Nutzung durch Hochschullehrer, wissenschaftliche Mitarbeiter, Studenten und Verwaltung).



## 2.1 Math-Net als elektronisches Nachrichtensystem, aktives Informationsangebot

Elektronische Post ist heute – immer noch – eine universelle Basis für den internationalen Austausch von Informationen unter Mathematikern. Das Internet ist heute zwar DAS weltumspannende Netz, doch verfügt nicht jeder Mathematiker über einen direkten Internet-Anschluß oder gar über die im Internet verbreiteten Werkzeuge zur Informationsvermittlung.

Man denke nur – insbesondere auch in Deutschland – an die “isolierten” Mathematiker, die vielleicht in einem Forschungslabor in der Industrie arbeiten, oder die in einer kommerziellen Umgebung leben, oder die in eine öffentliche Verwaltung eingebunden sind. Man denke auch an den “restlichen Teil” der Welt, an den östlichen Teil Europas oder an Entwicklungsländer. Viele Mathematiker haben ferner Zugang zum Bitnet oder zum UUCP (sie verfügen also über elektronische Post), aber nicht zum Internet. Mit Mathematikern in Schulen, oder mit historisch interessierten Mathematikern, die vielleicht mit einer Bibliothek oder einem Museum assoziiert sind, oder mit selbständigen, “technischen Schriftstellern” verhält es sich heute oft nicht anders.

Gerade in Deutschland sind auch die Mehrzahl der Verlage und die kleinen und mittleren Unternehmen vom Internet (und insbesondere vom deutschen Wissenschaftsnetz) ausgeschlossen. Bei vielen deutschen Softwarehäusern ist die Situation sehr ähnlich. Sie haben zwar Zugriff zu gewissen kommerziellen, elektronischen Dialogdiensten, jedoch heute noch selten einen Internet-Zugang.

Dieser Artikel beinhaltet deshalb den Vorschlag, zur allgemeinen Vermittlung von mathematischen Nachrichten (Informationen) ein elektronisches Nachrichtensystem auf der Basis von elektronischer Post aufzubauen, das z.B. Ankündigungen über neue Reports oder Preprints oder auch Preprint-Archive verteilt. Als erste Vorlage kann hier das Opt-Net dienen [LüggerD 93], das als internationales Kommunikationsforum für Wissenschaftler eingesetzt wird, die an Fragen der Optimierung interessiert sind, und das nach dem Vorbild des NA-Net aufgebaut wurde [DongarraR 91] (aber über dieses funktional hinausgeht).

Die Grundfunktionen für Nutzer bzw. Leser eines solchen Nachrichtensystems (auf der Basis von E-mail), lassen sich wie folgt zusammenfassen:

- |                    |  |
|--------------------|--|
| <b>help</b>        | Anfordern einer kurzen Einführung in den Gebrauch (Liste von E-mail-Kommandos).  |
| <b>info</b>        | Anfordern einer allgemeinen Einführung in Ziel, Zweck und in die Art und Weise, wie das Nachrichtensystem betreut wird; info liefert z.B. Angaben darüber, ob es durch einen Moderator betreut wird und wie er die Beiträge verteilt, oder, wer Beiträge leisten darf (und in welcher Form), falls das System ohne Moderation “gefahren” wird. |
| <b>subscribe</b>   | Subscribieren; anschließend erhält der Leser alle Nachrichten, die das System verteilt; welche Diskussionsbeiträge verteilt werden entscheidet i.d.R. der Moderator.   |
| <b>who</b>         | Herausfinden, “wer” (welche E-mail-Adressen) subscribiert hat.   |
| <b>unsubscribe</b> | Subscription auflösen.   |

<b>submit</b>	Diskussionsbeitrag einreichen; solche Beiträge können, wenn das Nachrichtensystem von einem Moderator betreut wird, auch von “externen” Teilnehmern (und ggf. auch anonym) vorgelegt werden; sofern es unmoderiert geführt wird, sollte der Beitragende namentlich als Mitglied bekannt sein.
<b>index</b>	Anfordern eines Index über bisherige Beiträge.
<b>join</b>	Als Mitglied namentlich in die Gemeinschaft der Teilnehmer am Nachrichtensystem eintreten. Aus technischen Gründen erhält das neue Mitglied eine eindeutige Kurzbezeichnung (nickname) und ein Paßwort, mit deren Hilfe er weitere Angaben zu seiner Person in einer “weißen Seite” machen (und diese auch ändern) kann, wie z.B. nähere Adressenangaben. Seine Whitepage ist anfangs fast leer, d.h. nur mit einem Namen versehen, dessen Korrektheit aber nicht (postalisch) verifiziert wird.
<b>memberlist</b>	Anfordern der aktuellen Liste der Mitglieder (incl. der nur per E-mail-Adressen und nicht namentlich bekannten Subscribenten).
<b>leave</b>	Als namentliches Mitglied austreten; eine eventuelle Subscription bleibt erhalten, aber alle näheren Angaben zur Person (und die Whitepage) werden gelöscht.
<b>whitepage</b>	Eintragen von Adressenangaben in die Whitepage; diese Angaben werden, nicht nur aus Datenschutzgründen, auf postalischem Wege validiert. Solange die Verifikation dieser Angaben nicht erfolgt ist, bleibt die Whitepage für andere Teilnehmer nicht lesbar (unpublished).
<b>publish</b>	Veröffentlichen einer persönliche Whitepage.
<b>unpublish</b>	Zurückziehen der Veröffentlichung.
<b>whois</b>	Abfragen von Whitepage-Einträgen anderer Teilnehmer, deren Whitepage sich im Zustand “publish” befindet.
<b>update</b>	Ändern von Einträgen in der Whitepage; solche Änderungen werden nicht (postalisch) verifiziert.

Neben diesen Grundfunktionen gibt es eine Reihe von weiteren Funktionen, z.B. zur Änderung des Paßwortes, zur detaillierteren Abfrage von Information etc., die in einer anderen Arbeit erläutert werden sollen. Auch dem Moderator (bzw. dem Verwalter) des Nachrichtensystems stehen eine Reihe von (administrativen) Kommandos zur Verfügung, die hier nicht näher ausgeführt sind.

Das Opt-Net kann nur als Modellvorstellung für ein einzelnes Nachrichtensystem dienen. Das Math-Net soll jedoch insofern über das Opt-Net hinausgehen, als es inhaltlich die ganze Mathematik umfassen soll und nicht nur das Teilgebiet Optimierung. Insofern sollen Nachrichtensysteme zu weiteren Gebieten, etwa für weitere Fachgruppen der DMV, im Math-Net eingerichtet werden können. Das Math-Net soll also eine Sammlung bzw. ein System von Nachrichtensystemen (der Opt-Net-Art) beinhalten, deren Diskussionsbeiträge nicht nur über das Math-Net verteilt, sondern auch im Math-Net langfristig archiviert werden, damit sie dort mittels Hypertext- und/oder Datenbank-Retrieval-Techniken suchbar sind (s.u.).

Wir sprechen in diesem Zusammenhang auch von Kollektionen und kommen – für eine genauere, technische Definition – später darauf zurück. Insbesondere diskutiert diese Arbeit in den Kapiteln 4 und 5, daß und wie die Kollektionen hierarchisch strukturiert werden können, damit sie auch hierarchisch (separierbar) absuchbar sind. Weitere Kollektionen des Math-Net werden, neben Nachrichten über neue, elektronisch verfügbare Reports, etwa Ankündigungen neuer Software- und Datensammlungen, oder Nachrichten über neue mathematische Informationssysteme (oder Teilen davon), oder auch Mitteilungen über neuere Beiträge zum Lebenden Mathematischen Museum beinhalten.

Leser des Math-Net werden sich i.d.R. nicht immer für alle Nachrichten zu allen Kollektionen interessieren. Sie sollten deshalb die Möglichkeit erhalten, die für sie nicht relevanten Informationen “auszublenden”, etwa dadurch, daß sie nur bei bestimmten Nachrichtensystemen des Math-Net subscribieren. Eine Unterteilung von Math-Net-Nachrichten nach Kollektionen, wie Reports, Software, Datensammlungen, Beiträgen zum Museum etc. erscheint noch relativ grob zu sein, auch wenn man diese noch weiter (hierarchisch) unterteilt (was dieser Artikel weiter unten diskutiert).

Eine Klassifikation von Nachrichten nach dem in der Mathematik international anerkannten “Mathematics Subject Classification” (MSC) Schema [FIZ-KA 94] oder nach dem in Bezug auf mathematische Software verbreiteten GAMS-Schema [NIST] erscheint deshalb unumgänglich. Mit Hilfe solcher Klassifikationen kann ein Leser von Math-Net-Nachrichten dann weitere, für ihn nicht relevante Informationen (ebenfalls) adressieren bzw. ausblenden (Vermeidung von “Information Overload”). Man kann sich z. B. vorstellen, daß Leser nur zu einer ganz bestimmten MSC-Klassifikation subscribieren (oder zu einer Obergruppe oder auch zu einer Kombination von solchen Gruppen) und auf diese Weise genau diejenigen Nachrichten erhalten, die sie fachlich interessieren.

Das Math-Net distribuiert also Informationen, die an es geschickt werden, aktiv und zielgenau an alle (fachlich) Interessierten. Wir werden später noch sehen, daß und wie sich der Nachrichtenteil des Math-Net auch im Zusammenhang mit dem traditionellen Publikations- und Referatewesen einsetzen läßt. Wir werden ebenfalls sehen, daß die über das Math-Net verteilten Informationen nicht nur langfristig im Verteilten Informationssystem gespeichert, sondern von dort wieder auf komfortable Art und Weise (d.h. ohne großen Lernaufwand seitens der Nutzer) abgefragt und abgerufen werden können – einschließlich der originalen, elektronischen Dokumente. Das Mittel der Wahl sind heute hierfür globale Hypertextsysteme.

## **2.2 Math-Net als globales Hypertextsystem, örtliches Informationsangebot**

Elektronische, fachspezifische Information hat – nicht nur aus technischen Gründen – einen völlig anderen Charakter als traditionelle Fachinformation. Sie steht, sei es nun eine Veröffentlichung (als Preprint, Buch oder Skript) oder eine Softwarebibliothek oder eine strukturierte Datensammlung oder ein elektronisches Journal, auf Servern als Original zur Verfügung und kann i.d.R. ohne besondere Kosten elektronisch sofort abgerufen werden. Im Internet gibt es inzwischen eine Fülle von solchen Servern, auf denen das Informationsangebot in sogenannten ftp-, Gopher- oder WWW-Archiven bereitsteht; siehe z.B. [Krol 94].

Elektronische Information hat heute – im Internet – einen solchen Umfang angenommen, daß an eine zentrale Speicherung nicht zu denken ist. Qualitativ hochwertige elektronische Information steht, durch freiwillige Beiträge vieler Institutionen, über die ganze

Welt verteilt zum Abruf bereit. Um sie nutzen zu können benötigt man, neben dem Anschluß an das Internet, nur noch einen sogenannten Client oder "Browser", wie z.B. Mosaic [AndreessenB 94] oder Netscape. Mosaic und Netscape sind darüber hinaus WWW-kompatible, globale Hypertextsysteme, die den einfachen und direkten Zugriff auf eine Fülle von Informationsarchiven gestatten. Globale Hypertextsysteme, wie WWW, Mosaic, Netscape und Hyper-G sind heute die Methode der Wahl, elektronische Informationen zu organisieren bzw. abzurufen [BernersCGP 92] [KappeM 93b] [Krol 94], wenn man weiß, wo sie zu finden ist .

Das Verteilte Informationssystem soll deshalb – langfristig – auch als globales Hypertextsystem organisiert werden. Die Partner stellen ihr örtliches Informationsangebot als Netz von Hypertext-Dokumenten auf lokalen Servern bereit, auf die sie dann mit Internet-kompatiblen "Browsern" dezentral und global zugreifen. Auf diese Weise tauschen die Partner des geplanten Projekts fachspezifische Information nicht nur untereinander aus, sondern – weltweit – mit ihren Kooperationspartnern aus anderen Ländern, aus anderen Fachrichtungen und aus anderen beruflichen Sparten (Industrie, Wirtschaft etc.).

Globale Hypertextsysteme, wie WWW, Mosaic, Netscape und Hyper-G, haben heute aber nicht nur weltweite Akzeptanz und Verbreitung gefunden, sie sind – gleichzeitig – noch Gegenstand intensiver Forschung und Entwicklung, insbesondere, wenn mit ihrer Hilfe umfangreiche Informationsangebote gemacht werden sollen. Ein besonderes, für Hypertexte (und für die Informationsangebote im Internet) spezifisches Problem ist die Erschließung der vorhandenen Information ("wo finde ich was?"). Vielfach fehlt den Informationsangeboten eine geeignete Strukturierung, die ein systematisches, globales Retrieval ermöglichen würde.

Mit dem Verteilten Informationssystem streben wir die Einführung einer (möglichst) homogenen, fachbezogenen bzw. "problemorientierten" Informationsstruktur (bzgl. der Mathematik in Deutschland) an, die die Orientierung in den vorhandenen Informationsangeboten erleichtern und eine simultane Suche nach relevanter Information in mehreren Servern gestatten soll.

Das geplante, mathematische Informationsangebot läßt sich etwa wie folgt charakterisieren:

- \* Umfangreiche Menge von Dokumenten
  - lokal: ca. 100 - 1000
  - national: über 100.000 (in ein bis zwei Jahren)
  - global: ca. 1.000.000
- \* Komplexe Struktur der Information
  - bzgl. des Informationsnetzes (Struktur der "Links and Nodes")
  - bzgl. der Dokumente ("innere Struktur" sehr unterschiedlich)
  - bzgl. ihrer Anordnung und Klassifikation
- \* aktuell, hochgradig dynamisch (laufend zu aktualisieren)
  - mit Beiträgen von vielen Seiten bzw. Autoren, die aber nur
  - über sehr unterschiedliche Grade von Erfahrung verfügen und die
  - nur relativ schwach koordiniert werden können.

Die Frage nach der "besten Struktur" oder einer "vereinheitlichten Struktur" oder nach der "besten Anordnung" des Informationsangebots läßt sich dabei leider prinzipiell nur

sehr unvollkommen beantworten. Was “das Beste” ist, hängt eher von den Bedürfnissen eines Lesers ab und den Themen, die er im Sinn hat und zu denen er hilfreiche Information sucht. Für Autoren und “Betreiber” von Informationssystemen hängt diese Frage auch von den speziellen Erfahrungen und Fähigkeiten ab, die sie erwarten können bzw. die sie selbst einzubringen in der Lage sind. All dieses ändert sich aber von Zeit zu Zeit.

Außerdem kommt – bei einem weltweiten Angebot – eine Schulung nicht mehr in Frage. Die innere Struktur und die Benutzeroberfläche des Math-Net soll intuitiv sofort verständlich oder – aufgrund der Verbreitung von globalen Hypertexten im Internet – allgemein bekannt sein.

Obendrein entwickeln sich Informationssysteme im Internet heute technisch mit großer Geschwindigkeit weiter. Dieser Artikel unterbreitet deshalb den Vorschlag, die Gestaltung des Verteilten Informationssystems nicht so sehr an den Gegebenheiten und Randbedingungen der heute vorhandenen, globalen Hypertextsysteme zu orientieren, sondern in erster Linie an den Bedürfnissen, die die Nutzer solcher Systeme und die Anwender der Mathematik haben.

Der Vorschlag ist, das Augenmerk auf die für die Mathematik notwendigen Informationsstrukturen zu richten. Die im Internet verfügbaren Tools unterliegen einem ständigen Wandel. Neue Werkzeuge werden erscheinen, die die alten ablösen. Die Informationsstrukturen der Mathematik aber sollen langfristig Bestand haben. Auch nach dem Projekt soll das geplante Verteilte Informationssystem lebensfähig sein.

Unter diesem Gesichtspunkt erhält der Punkt “Offenheit bzgl. des Internet” ein vorrangiges Gewicht. Das heißt insbesondere auch, daß sich die Partner des Projekts mit ihren Informationsstrukturen nicht an spezifische Eigenheiten der heute verfügbaren Systeme derart binden sollten, daß sie langfristig darauf “festgenagelt” sind. Welche Informationsangebote die Partner machen und welche Tools sie dabei verwenden, liegt, um es nochmals zu betonen, in ihrer eigenen Verantwortung. “Verteilt” bedeutet – auch organisatorisch – “dezentral”, und damit auch eine “verteilte Verantwortung”.

Zur Bewertung von Informationssystemen der Mathematik und insbesondere zur Gestaltung des Verteilten Informationssystems (auf der Basis von globalen Hypertextsystemen) aus Nutzersicht wurde ein Katalog von Kriterien entwickelt, der auch – bzgl. des Hypertext-Anteils – als Anforderungskatalog für das Verteilte Informationssystem gelesen werden kann [Lügger 94]. Dieser Kriterien-Artikel diskutiert, neben der spezifischen Problematik des Einsatzes globaler Hypertextsysteme, vor allem die Bedürfnisse und Anforderungen der Leser, der Autoren und der Systemadministratoren von Hypertexten und setzt dabei voraus, daß Leser i.d.R. ein spezifisches Thema verfolgen.

Ein Wissenschaftler, der z.B. an Fragen der Optimierung interessiert ist, wird nicht nach Servern oder Archiven fragen, sondern eher nach Verweisen (Hypertext-Links oder ftp-Adressen) auf Information zur Optimierung, z.B. nach

- \* Forschungsprojekten zur Optimierung: an welchen Fachbereichen?
- \* Elektronischen Reports der Kombinatorischen Optimierung: gibt es neue?
- \* diesbezüglichen Vorlesungsskripten: sind auch die  $\text{\TeX}$  -Quellen dazu verfügbar?
- \* Software und Testdaten zum “Matching Problem”: was ist verfügbar und wie greift man darauf zu?
- \* Kontaktadressen von Entwicklern dieser Software: wie lautet ihre E-mail-Adresse?

- \* Problemstellungen und Antworten dazu, etwa im Opt-Net: wie behandelt man z.B. “parametric submodular maximization”?
- \* Vorträgen und Konferenzen: welche finden in meiner Nähe statt (wo und wann)? Oder, wie kann man sich anmelden?

Fragen dieser Art sind mit üblichen WWW-Servern kaum zu behandeln, vor allem wenn eine deutschlandweite (oder gar weltweite) Suche gewünscht wird. Der o.g. Artikel stellt die Kernkriterien zusammen, denen das Math-Net in voller Ausbaustufe letztlich genügen soll. Er unterscheidet dabei die folgenden Kriteriengruppen:

- \* Kriterien des Lesers: Für ihn sind bei seiner Suche nach Informationsangeboten vor allem folgende Themen besonders relevant:
  - wissenschaftliche Inhalte, vor allem aus Forschung und Lehre
  - zugehörige Software- und Datensammlungen
  - wissenschaftliche Serviceangebote, einschl. solcher der Bibliothek
  - organisatorische Kontaktangebote bzgl. Lehre und Administration
  - wissenschaftliche Kontaktangebote aus Forschung und Lehre
- Daneben hat er aber auch Fragen
  - nach inhaltlichen Zusammenhängen der Informationen
  - nach neuen Informationen (Aktualisierungen)
  - bzgl. der Benutzerfreundlichkeit und der Dokumentation
- \* Kriterien des Autors von Informationsangeboten bzw. einer Gruppe von Autoren: Relevant sind für sie (zusätzlich) auch Fragen nach der
  - Handhabbarkeit und Bedienerfreundlichkeit der Dokumentenerstellung
  - Möglichkeit, bereits bestehende Informationen einzubeziehen; z.B. Postscript-Dateien (die sich mit “normalen” Wordprozessoren erstellen lassen) und in der Mathematik insbesondere Konvertierung von  $\text{T}_{\text{E}}\text{X}$  - und  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  -Dateien.
  - Koordination mit anderen Autoren
- \* Kriterien des Administrators: Für ihn stehen im Vordergrund die Hilfsmittel zur
  - technischen Installation und Maintenance
  - Unterstützung der Interaktionen von Autoren und Lesern, etwa durch spezielle Diskussionslisten
- \* Strukturelle Kriterien:
  - aus der Sicht des Lesers
  - aus der Sicht der Informationsanbieter (einschl. Administrator)
- \* Kriterien bzgl. des “Verteilten Informationssystems für die Mathematik”
  - Globale und strukturelle Suche nach Information
  - Aktive Vermittlung von Information (e.g. auf der Basis von E-mail)
  - Zusammenspiel mit anderen Typen von Informationssystemen (“Interoperabilität”)

In dem Kriterien-Papier ist ein recht breites Spektrum von Eigenschaften angesprochen, die man von einem mathematischen Informationsangebot erwarten kann. Um einem möglichen Mißverständnis vorzubeugen: Das bedeutet nicht, daß die Partner des Projekts

und insbesondere die mathematischen Fachbereiche nun alle das ganze Spektrum der im Kriterien-Papier angesprochenen Information anbieten müßten. Welche Information ein Partner in das Netz einbringt, liegt in seiner Hand. Nur, wenn er etwas einbringt, sollte dieses der Struktur des Verteilten Informationssystems "harmonisch" entsprechen. Nur so wird das Verteilte Informationssystem auch alle Informationsangebote erschließen können. Das Verteilte Informationssystem sollte allerdings (möglichst) mit allen so strukturierten Informationsangeboten "zu Rande kommen". Für das Math-Net kommen deshalb im Prinzip alle im Kriterien-Papier genannten Kriterien zur Anwendung.

Während die Situation bei den einzelnen Fachbereichen vielleicht noch überschaubar erscheinen mag, gelten für das Math-Net, den geplanten "zentralen" Teil des Verteilten Informationssystems, besondere Bedingungen, die aus dem Umfang des Informationsangebots resultieren, mit dem wir rechnen müssen (s.o.). Ein "Leser" von Hypertexten hat hier besondere Bedürfnisse und Anforderungen. Hierzu nur ein Beispiel:

Wenn er sich über Aktivitäten in der "Optimierung" (oder Teilgebiete davon) informieren will, möchte er vielleicht den MSC-Index zur Hand haben (etwa, indem er in einem Hypertext die Klasse "msc-90" adressiert). Wenn er sich dann, per Hypertext und Mausclicks, tiefer in die Struktur der Information "einwählt", etwa in "Departments", "Institutes" und "Projects" ([Lügger 94]), möchte er dort vielleicht alle Projekte in Deutschland (oder auch, im Math-Net) finden, die etwas mit Optimierung zu tun haben.

Wie kann man nun solch eine Anforderung etwa im World Wide Web erfüllen? Genau das ist ein zentrales Thema des hier vorliegenden Papiers. Wer sich Gedanken darüber macht, wie solch eine Information auf dem neuesten Stand gehalten werden soll, sieht sehr bald, daß ein deutschlandweites oder auch ein globales Informationsangebot nicht mehr mit traditionellen Mitteln (etwa "per Hand") zu pflegen ist und, daß die örtlichen Informationsstrukturen gewisse, formale Voraussetzungen erfüllen müssen, wenn man in dieser Hinsicht eine "Automatisierung" erreichen will. Wir werden sehen, daß eine solche Aufgabe auch nicht allein mit Mitteln eines Hypertextsystems, das auf dem einfachen Link-Node-Model beruht, lösbar ist, sondern die Integration von Datenbankkonzepten voraussetzt.

Die vorliegende Arbeit diskutiert in den Kapiteln 4 und 5 insbesondere einen ersten Vorschlag zur Strukturierung des dezentralen und des zentralisierten Informationsangebots. Ein detaillierter Vorschlag zur Strukturierung des Informationsangebots bei den mathematischen Fachbereichen befindet sich in Arbeit. Eine Vorstellung ist allerdings dabei, die aus zentralen Bedürfnissen resultierenden, strukturellen Anforderungen möglichst minimal zu halten. Die Partner des Projekts sollen genügend Freiheitsgrade erhalten, ihr örtliches Informationsangebot "individuell" zu gestalten, allerdings auch so, daß es vom "zentralen Anteil" des Math-Net "verstanden" und verarbeitet werden kann.

### **2.3 Math-Net als Datenbank, zentralisiertes Informationsangebot**

Reine Hypertexte unterstützen das (eher) "unstrukturierte" bzw. assoziative Suchen nach Information. Jeder, der gerne Buchläden besucht, weiß, was damit gemeint ist – und wie man dabei auf neue und interessante Bücher stoßen kann, die man vorher nicht einmal im Sinn gehabt hat. Diese Art des Herumstöberns hat ihren eigenen Wert und jeder Buchladen hat auch ein eigenes System, seine Bücher interessant anzubieten. Wer jedoch auf die Dienste größerer Bibliotheken angewiesen ist, weiß auch, daß er dabei ohne eine gute Systematik "verloren" ist. Für das elektronische Angebot von Katalogen, z.B. von Uni-Bibliotheken, sind gute Datenbanken unumgänglich.

Vor allem wegen des schier unermesslichen Umfangs an Information, der sich ein Leser im Verteilten Informationssystem gegenüber sehen wird, soll das Math-Net eine "zentrale" Datenbankkomponente enthalten. Es gibt aber noch einen weiteren, organisationstechnischen Grund, warum eine Datenbankkomponente für das Math-Net unentbehrlich ist. Er liegt aber – technisch – etwas tiefer und hat etwas mit dem Aufbau des Verteilten Informationssystem als "mehrdimensional strukturiertes" Hypertextsystem (wie es hier einmal kurz bezeichnet werden soll) zu tun.

Zum einen kann das Verteilte Informationssystem als Menge der verschiedenen Hypertexte bei den Partnern des Projekts gesehen werden. Diese Verteilung bildet die "erste Dimension". Aus Gründen, die in den folgenden Kapiteln noch näher erläutert werden, bietet es sich an, diese Hypertexte hierarchisch – nach Kollektionen – zu strukturieren (gleichwohl können die einzelnen Informationseinheiten mittels "Links" miteinander vernetzt sein). Diese "Dimension" soll deshalb auch als "vertikale Dimension" bezeichnet werden. Die vertikale Dimension wird im wesentlichen von den Informationsbeauftragten der Partner organisiert und gepflegt.

Zum anderen tritt jede Informationseinheit der vertikalen Dimension auch (mindestens je einmal) in der zentralisierten Komponente des Math-Net auf, die nach einem anderen – ebenfalls hierarchischen – Schema gegliedert ist, dem MSC-Index (für Publikationen) [FIZKA 94] vereinigt mit dem GAMS-Index (für Software- und Datensammlungen [NIST]).

Lassen wir an dieser Stelle einmal außer acht, daß es Informationseinheiten geben wird, die in keines dieser Klassifikationsschemata eingeordnet werden können (sie werden einfach jeweils als "unclassified" klassifiziert). Und nehmen wir an, daß der MSC-Index ebenfalls als "hierarchischer" Hypertext organisiert ist. Unter der Klassifikation msc-90 sollte ein Leser dann alle Einträge in das Math-Net finden, die nach dem MSC-Index 90 klassifiziert sind, also alle diesbezüglichen elektronischen Reports, Vorlesungsskripte, Beschreibungen von Instituten und Projekten, Beiträge zum Lebenden Mathematischen Museum etc. – quer durch ganz Deutschland. Diese "Dimension" soll deshalb auch als "horizontale Dimension" bezeichnet werden. Die horizontale Dimension wird aber (das wollen wir in diesem Zusammenhang zunächst einmal annehmen) von anderen Personen organisiert, als den Informationsbeauftragten der Partner.

Abgesehen von der großen Anzahl der MSC-Klassen (z.Z. über 4000) wird jetzt folgendes Problem offensichtlich: Wie koordiniert man die vertikale mit der horizontalen Dimension? Stellen wir uns vor, daß – an mehreren Fachbereichen in Deutschland – mehrere neue Einträge, die nach msc-90 klassifiziert sind, gemacht werden (von mehreren Informationsbeauftragten). Es ist schlechterdings unmöglich, den Schreibzugriff auf die entsprechende horizontale msc-90-Liste allen diesen Personen zu ermöglichen. Die zentralisierte, horizontale "Liste" könnte also (bestenfalls) nur von einer einzigen, verantwortlichen Person gewartet werden.

Damit ist ein grundlegendes, organisatorisches und technisches Problem angesprochen, das – so die These dieser Arbeit – im Rahmen des einfachen Link-Node-Modells von Hypertexten der WWW/HTML/Mosaic-Art nicht lösbar ist. Abgesehen davon, daß man es mit ca. 4000 MSC-Klassen zu tun hat, für die man entsprechende Betreuer finden müßte, die die zentralen Einträge per Hand vornehmen müssen, stellen inhaltliche Änderungen und dann auch Löschungen von Informationseinheiten ein weiteres diffiziles Problem dar. Ob gewisse (bzw. welche) Anteile dieser Aktualisierungen automatisierbar sind, ist ein anderes Thema, das an anderer Stelle noch zu diskutieren ist.

Die These dieser Arbeit ist, daß die oben geschilderte Problematik dadurch lösbar ist, daß



in das Math-Net eine Datenbank integriert ist, die mit der Hypertext-Komponente des Math-Net “eng gekoppelt” ist und die von den Informationsbeauftragten mit Information über neue Einträge, Änderungen und Löschungen beschickt wird, etwa über den oben skizzierten Nachrichtenteil.

Was ist nun unter einer “engen Kopplung” zu verstehen? An dieser Stelle kann man sich – in erster Näherung – vorstellen, daß die horizontale Hypertextstruktur nicht “physisch” existiert, sondern (nur) virtuell: Wenn ein Leser die horizontale Struktur etwa auf “msc-90” abfragt, generiert die Datenbank, in der alle aktuellen Informationseinheiten “registriert” sind, einen entsprechenden Hypertext, der nun vom Leser angeklickt oder auch weiter durchsucht werden kann. Andere Organisationsformen sind ebenfalls denkbar. Die später folgenden Kapitel diskutieren diese Thematik im Einzelnen.

Die Anforderung, daß ein aus der Datenbank entwickelter Hypertext “weiter durchsucht” werden kann, stellt ein weiteres, diffiziles technisches Problem dar, zu dem es bei den heutigen, globalen Hypertextsystemen noch keine vorgefertigte Lösung gibt; diese ist aber aus Nutzersicht absolut notwendig. Das Ergebnis einer Suchanfrage kann ja sehr umfangreich sein, so daß eine weitere Detaillierung der Anfrage erforderlich ist. Im Rahmen des vorgeschlagenen Modells soll z.B. ganz global nach “msc” gesucht werden können, also nach allen Informationseinheiten, die überhaupt nach dem MSC-Index klassifiziert sind (eine solche Anfrage ist z.B. dann sinnvoll, wenn man Software aus der Anfrage ausschließen will, die ja i.d.R. nur nach dem GAMS-Index klassifiziert ist).

Außerdem könnte sich ein Leser des horizontalen Teils ja wünschen, daß er nur über Projekte, die nach msc-90 klassifiziert sind, informiert wird, und nicht über Preprints, Institute etc. Diese Forderung führt zu dem (weiteren) Problem, daß der Ergebnis-Hypertext ähnlich wie die vertikalen Informationsstrukturen bei den Partnern aufgebaut sein sollte: Die horizontale Informationsstruktur sollte der vertikalen “ähnlich” sein. Auch dieses Problem, das in den folgenden Kapiteln näher diskutiert wird, ist mit den gegenwärtigen Mitteln globaler Hypertexte nicht so ohne Weiteres lösbar.

Damit ist aber die “dritte Dimension” des zentralisierten Anteils der Information im Math-Net angesprochen. Diese Informationsstruktur, die ebenfalls “horizontal” verläuft, ist – so der Vorschlag dieser Arbeit – mit der vertikalen Struktur bei den Partnern homogen: Ein Leser des (zentralen) Math-Net sollte grundsätzlich auch die Möglichkeit haben, einen horizontalen Hypertext zu “durchstöbern”, der nach Publikationen, Software, Info-Systemen, organisatorischen Informationen, Kontaktadressen etc. gegliedert ist – quer durch alle Partnerorganisationen, die am Verteilten Informationssystem beteiligt sind. Die Kapitel 4 und 5 entwickeln und diskutieren genau die Struktur dieser dritten Dimension, die als tragendes Skelett auch in der Informationsstruktur der Hypertexte der vertikalen Dimension auftritt.

Innerhalb des Verteilten Informationssystems sind noch weitere Dimensionen für Informationsstrukturen denkbar und auch notwendig: Alle am Math-Net offiziell Beteiligten (e.g. Moderatoren des Nachrichtenteils), insbesondere auch alle, die inhaltliche Beiträge leisten oder die auch nur als (offizielle) Kontaktpartner bei einem der Partner in Frage kommen, müssen z.B. im Rahmen des Math-Net namentlich bekannt sein. Ihre beruflichen Kontaktadressen (E-mail, Post, Telefon, Fax) sollen über das Math-Net abfragbar sein.

Darüber hinaus müssen auch Subscribenten des Math-Net, auch wenn sie in gewisser Weise anonym bleiben wollen, zumindest in der Datenbank des Math-Net bekannt sein, damit ihnen das Math-Net die für sie interessanten Beiträge zustellen kann (Nachrichtenkomponente). Auch diese Zuordnung ist im Rahmen der Datenbankkomponente des

Math-Net zu lösen. Ferner auch alle mit Zugriffsrechten und login-Namen (“Nicknames”) zusammenhängenden Fragen. Damit ist die Aufgabenstellung der Datenbankkomponente des Math-Net global umrissen.

## 2.4 Zusammenwirken – dezentrale und zentralisierte Information

Das Verteilte Informationssystem und das Math-Net sollen eine erste, technische Grundlage für den effizienten, elektronischen Austausch von Information bilden. Eine – sehr knappe – technische Charakteristik ist, daß die eigentliche Information (Volltexte, Hypertexte, Software etc.) ausschließlich dezentral vorgehalten wird. Zentralisiert sollen nur Referenzen auf die jeweils entsprechenden Informationseinheiten gehalten (gespeichert und verteilt) werden.

Die dezentralen Informationsangebote sind “vertikal” strukturiert, d.h. hierarchisch nach Kollektionen geordnet, die selbst wieder nach weiteren “Subkollektionen” untergliedert sein können. Abb. 1 symbolisiert diese Struktur in Form von Dreiecken, die auch jeweils einen “Suchraum” bestimmen, auf den sich eine Datenbankanfrage beziehen kann. Die äußeren Dreiecke repräsentieren dabei die gesamte Menge der angebotenen Information. Eine Suche, die an der Spitze dieser Menge beginnt, schließt also das gesamte örtliche Informationsangebot ein.

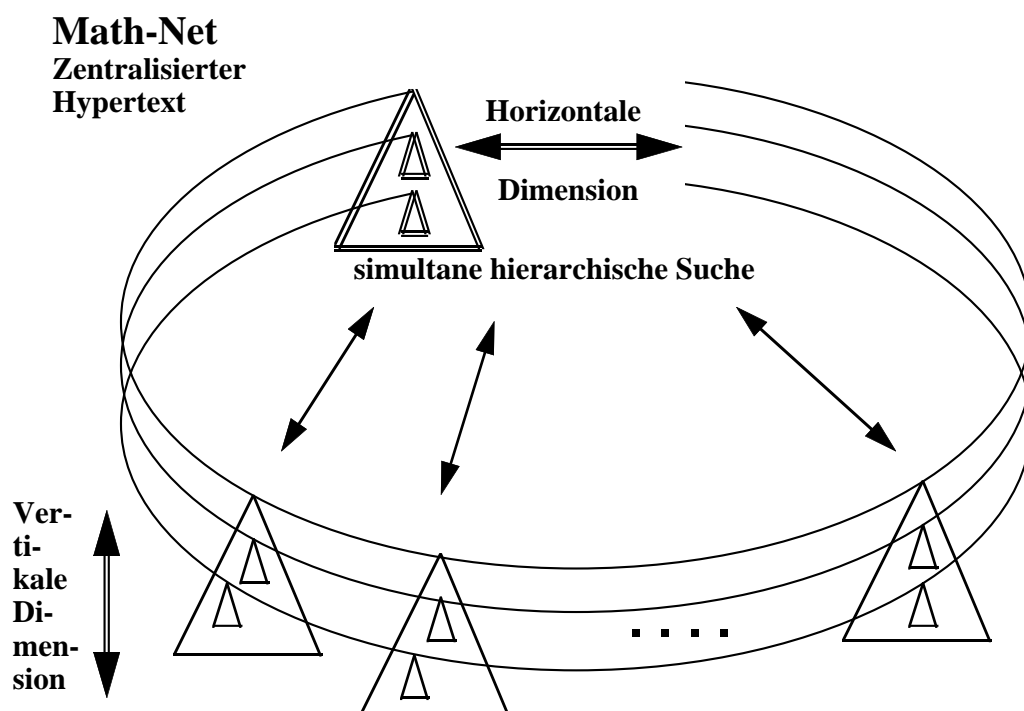


Abb. 1: Vertikale und horizontale Informationsstrukturen

Die einander entsprechenden Informationskollektionen, etwa alle Reports oder alle organisatorischen Beschreibungen, sind “horizontal” zusammengefaßt. Es sind Kollektionen von Kollektionen, die insgesamt wieder “vertikal” (also harmonisch bzw. entsprechend

hierarchisch) gegliedert sind. Der im Math-Net zentralisierte Hypertext repräsentiert also beide strukturellen Komponenten (das soll der doppelte Rand in Abb. 1 symbolisieren). Sein “vertikaler” Anteil erlaubt dem Leser die hierarchische Selektion (mittels Maus-Clicks, also mit Hypertext-Mitteln) nach Kollektionen (also nach fachlichen Gesichtspunkten), wobei an jeder Stelle dieses Hypertextes gleichzeitig die (entsprechende) “horizontale” Struktur zur Verfügung steht – quer durch ganz Deutschland. Der eigentliche Sinn dieser Anordnung besteht darin, daß Leser an jeder Stelle der Hierarchie (im Math-Net) simultan in allen entsprechenden Teilen der Informationsangebote der Partner des Projekts auch mit Datenbankmitteln suchen kann. Näheres dazu findet sich im Teil II dieses Artikels.

Wie repräsentiert nun das Math-Net die dezentral vorgehaltenen Dokumente? Im Falle von Reports kann man sich das so vorstellen, daß im Math-Net etwa nur die erste Seite eines Reports gehalten wird. Sie enthält alle wichtigen Angaben über den Report: den Titel, die Autoren, den Abstract, die MSC-Klassifikationen und entsprechende Schlüsselworte etc., also alle Angaben, die für die Entscheidung eines Lesers notwendig sind, ob er das zugehörige Originaldokument über das Netz transferieren soll oder nicht.

Das Referenz- und das Originaldokument sind über einen sogenannten Hyperlink miteinander verbunden, so daß ein Leser, der ein interessantes Referenzdokument im Math-Net gefunden hat, das zugehörige Originaldokument – mittels Mausclick – sofort auf seine Workstation übertragen kann. In der WorldWideWeb-Welt heißt solch ein Hyperlink “Uniform Resource Locator” (URL) [BernersCGP 92]; er “zeigt” gewissermaßen vom Referenzdokument direkt in den Speicherbereich hinein, in dem das assoziierte Originaldokument abgelegt ist. Als mögliche dezentrale Speichersysteme (also bei den Partnern des Projekts) kommen dabei ftp-, Gopher-, WWW- oder Hyper-G-Server in Betracht.

Der Fall eines Reports, dessen Original vielleicht nur aus einem einzigen  $\text{\LaTeX}$  - oder Postscript-File besteht, ist nur der einfachste Fall. Schon Reports, die Grafiken enthalten, sind aus mehreren Files zusammengesetzt. In diesem Falle zeigt der Hyperlink (bzw. der URL) nur auf das “oberste Element” im dezentralen Speicher, etwa auf ein Directory im ftp-Bereich, in dem das Original abgespeichert ist. Falls das Originaldokument in einem Gopher-Server bereitsteht, zeigt der Hyperlink vielleicht auf das entsprechende Gopher-Submenü. Falls das Original in einem WWW-Server bereitsteht, zeigt der Hyperlink auf das entsprechende HTML-Dokument mitten in der Hypertext-Struktur, von dem ausgehend sich der Leser dann – mittels Mausclicks – hindurchwählt, um an die für ihn interessante Information zu kommen.

In jedem Falle ist das Originaldokument durch ein einziges “Dokument” vertreten, das die Spitze der Informationseinheit darstellt (und auch ein Directory oder ein Menü sein kann): Diese “Dokument” soll im folgenden TOP-Dokument heißen. Es ist mit dem Originaldokument identisch, sofern dieses nur aus einem einzigen File besteht. Im anderen Fall repräsentiert es eine File-Hierarchie (ftp-Archiv) oder eine Submenüstruktur oder eine Hypertextstruktur (WWW oder Hyper-G).

Wie erfahren nun die Partner des Projekts davon, daß neue Dokumente in das verteilte Informationssystem eingebracht werden? Rein technisch gesehen wäre es möglich, daß eine Institution Preprints oder auch gewisse Multi-Media-Dokumente an alle Partner verteilt (Multi-Media-Mail; siehe z.B. [Borenstein 92]). Dieses Vorgehen erscheint jedoch nicht sinnvoll (und ist bei komplexeren Strukturen auch nicht immer möglich). Es genügt, wenn die Institution ein zu dem neuen Beitrag korrespondierendes Referenz-Dokument an alle Partner verteilt, das, neben einer Kurzinformation über den Inhalt (in Textform), auch einen Hyperlink auf das Top-Dokument des Originals enthält.

Dazu bedient sich die beitragende Institution des Math-Net's, das die weitere Distribution übernimmt. Das Math-Net wird das Referenz-Dokument also nicht nur in seine Datenbank integrieren, sondern auch direkt an alle diejenigen Subscribenten verschicken, die sich für die entsprechende Informationsklasse interessieren. Dabei ist stillschweigend vorausgesetzt, daß das Referenz-Dokument auch entsprechend klassifiziert ist. Die über das Math-Net verteilbaren Referenz-Dokumente müssen also einer gewissen (minimalen) Struktur genügen. Weiteres dazu im Teil II dieses Artikels.

Für die Startphase des Math-Net kann man sich vorstellen, daß die Partner ihre neuen Beiträge "per Hand" ankündigen. Im Verlauf des Projekts sollen Verfahren entwickelt werden, diesen Prozeß nach Möglichkeit zu automatisieren. Unter der Voraussetzung, daß bei der Bereitstellung von Information gewisse Konventionen eingehalten werden, sollte sich der Prozeß der Integration neuer Information in das Math-Net – etwa mit im Softwarebereich heute schon üblichen Mirror-Verfahren – zum großen Teil auch automatisieren lassen. Von irgendeinem Partner neu (auf seinem Server) bereitgestellte Dokumente sollten dann – so das Fernziel – automatisch allen anderen Lesern des Math-Net (etwa per E-mail) bekannt gemacht werden können.

Jede Institution (weltweit), die mathematische Information bereitstellt, soll das Math-Net beschicken können. Das Annoncieren von mathematischer Information über das Math-net soll allerdings nur autorisierten Personen möglich sein, deren Adressenangaben zuvor verifiziert sind und die den allgemeinen Regelungen des Math-Net zugestimmt haben. Ihnen sind insbesondere die besonderen Formen der Strukturierung bekannt, mit denen sie Referenz-Dokumente in bestimmte Kollektionen des Math-Net (wie z.B. Preprints, Referate, Software, Datensammlungen, Beiträge zum Mathematischen Museum etc) integrieren. Außerdem kennen sie die im Math-Net üblichen Verfahren der Klassifizierung (etwa per MSC- oder GAMS-Index).

Das Math-Net wird sich, so wie es technisch angelegt ist, insbesondere auch dafür eignen, zu speziellen Themen elektronische Mailverteiler bzw. Diskussionsforen einzurichten, deren Beiträge ebenfalls im Math-Net langfristig gespeichert werden. Solche Diskussionsforen sollen i.d.R. durch einen Moderator betreut werden, der die Information sichtet (und eventuell auch mögliche Rückfragen klärt), bevor diese weiter verteilt wird. Alle Interessenten (selbst anonyme Personen) können Beiträge zu einem Diskussionsforum leisten, sofern es durch einen Moderator betreut wird.

Die von projektexternen mathematischen Institutionen autorisierten Personen spielen die Rolle der Informationsbeauftragten der Partner des Projekts. Sie sollten, wie diese und wie die Moderatoren des verteilten Informationssystems, namentlich (mit Adressen) bekannt sein. Leser der Nachrichten des Math-Net sollen allerdings nicht gezwungen sein, ihre Adressen der Öffentlichkeit bekannt zu geben. Sie können ihre Adressangaben in das Math-Net einbringen (Whitepage im Sinne des Opt-Net), aber sie brauchen sie nicht zu publizieren. Für die Mehrzahl der Leser des Math-Net sollte es sogar genügen, wenn sie nur durch ihre E-mail Adresse bekannt sind.

Ein kombiniertes, elektronisches Nachrichten- und Hypertextsystem dieser Art ist sowohl für den "passiven" Nutzer geeignet, der sich (per E-mail) über Neuankündigungen zu bestimmten Sachgebieten informieren lassen möchte (hier wird das System "aktiv"), als auch für den "aktiv" zugreifenden Nutzer, der sich auf dem Retrieval-Wege (via Datenbank plus Hypertext) einen einfachen Zugang zu den aktuell verfügbaren Informationsquellen verschaffen will (hier verhält sich das System "passiv").

Die Beiträge zum Math-Net sind dabei nicht auf (nur) Ankündigungen von neuen Informa-

tionsressourcen beschränkt. Es können auf diesem Wege auch Kommentare und Referate zu den Ressourcen bereitgestellt und verbreitet werden, die jeweils mit der entsprechenden Zugriffsinformation auf die originale Quelle versehen sind. Wenn auch Revisionen und Neuauflagen von Informationsressourcen (und wieder Kommentare bzw. Referate dazu) über das Math-Net verbreitet werden, findet sich im Verteilten Informationssystem der ganze Life-Cycle einer elektronischen Publikation wieder (living document), sei es nun ein Report, ein Buch oder ein Stück Software oder eine Sammlung von Bildern (im Mathematischen Museum) oder irgend eine andere Darstellung mathematischer Art.

### **3 Übergang zur elektronischen Fachinformation** – Math-Net und Zentralblatt

#### **3.1 Von traditionellen zu elektronischen Publikationsformen**

Der Bereich des wissenschaftlichen Publizierens und der wissenschaftlichen Kommunikation befindet sich in einer Phase des Umbruchs [GrötschelLS 93], [GrötschelL 95b]. In den letzten Monaten wurden bereits etliche voll-elektronische, mathematische Journale gegründet, auf die ein Leser – i.d.R. kostenfrei – über das Internet zugreifen kann. In der American Mathematical Society diskutieren führende Mathematiker bereits die Ablösung traditioneller, mathematischer Journale (auf Papier) durch elektronische Formen [Franks 93], [Odlyzko 95], [Quinn] (in Europa auch [Collart 94]). Auch die DMV bereitet ein universelles, elektronisches Journal vor, die “Documenta Mathematica” [Rehmann 95]. Odlyzko sagt das baldige Ende der traditionellen, wissenschaftlichen Publikationen (in einem Zeitraum von 10 bis 20 Jahren) voraus.

Die Gründe für diese Annahmen liegen vor allem in dem weltweiten Wachstum der wissenschaftlichen Literatur sowie in der Schwerfälligkeit des traditionellen Publikationswesens und den erheblichen Kosten, die von wissenschaftlichen Institutionen und Bibliotheken dafür aufzubringen sind. Der Markt für traditionelle Papierjournale ist aufgrund der stark gestiegenen Preise und der aktuellen, finanziellen Situation der Bibliotheken beschränkt, wenn nicht sogar rückläufig. Das Bedürfnis, mathematische Ergebnisse zu publizieren, wächst aber mit der steigenden Anzahl der Wissenschaftler weiter an. Hier stoßen die neuen elektronischen Fachzeitschriften in eine große Lücke, die im Rahmen des traditionellen Publikationswesens kaum noch geschlossen werden kann.

Die Informationsanbieter der Mathematik (und ihrer Anwender), die Verlage, die Bibliotheken, die Fachinformationszentren und – nicht zuletzt – die mathematischen Fachbereiche und Forschungsinstitute sehen sich damit vor eine große Herausforderung gestellt: dem Übergang von traditionellen Formen, miteinander zu kommunizieren (Briefe, Mitteilungen, Journale auf Papier) zu elektronischen Formen der Weitergabe und Verbreitung von fachlicher Information (E-mail, Diskussionsforen, E-Journals). Die Ergebnisse wissenschaftlicher Arbeiten konkretisieren sich heute außerdem auch in mathematischen Software- und Datensammlungen. Die Bedeutung dieser wichtigen Ressourcen, die vom traditionellen Publikations- und Referatewesen bisher kaum erfaßt wurden, steigt vor allem in den Anwendungen der Mathematik und in interdisziplinären Bereichen zunehmend an. Längerfristig werden Videos und komplexe Multimediadokumente hinzukommen, die es auch heute schon gibt – allerdings in der Mathematik noch in relativ geringer Zahl.

Traditionelle Publikationsformen werden, wenigstens für einige Zeit, vermutlich aber auch langfristige, weiterhin Bestand haben. Die These dieser Arbeit ist, daß die traditionellen,

wissenschaftlichen Journale zunächst durch entsprechende, elektronische Publikationsformen ergänzt werden. Sie treten damit in einen "Wettbewerb" mit dem neuen Medium ein, dessen Ausgang noch ungewiß ist und der vermutlich auch nicht für alle Journale in gleicher Weise entschieden werden wird. Die besten traditionellen Journale werden möglicherweise, trotz der mit ihnen verbundenen Engpässe (e.g. lange Warteschlangen wegen der nur begrenzt zur Verfügung stehenden Druckseiten), überleben, während etliche traditionelle Journale der elektronischen Konkurrenz unterliegen werden, nicht zuletzt aus Kostengründen.

Bücher, insbesondere Lehrbücher und qualitativ hochwertige Monographien, in der üblichen Form wird es weiterhin geben. Sie werden jedoch durch elektronische Versionen ergänzt werden, die Hypertext-Links erlauben, den Zugriff auf umfangreiches Datenmaterial und damit z.B. den Ablauf von komplizierten Algorithmen an interessanten Beispielen ermöglichen. Geeignete Graphiken, Visualisierungstechniken, statistisches Material oder sogar Videosequenzen können zu einer erheblichen Verbesserung von Lehrmaterial beitragen. Bei Spezialliteratur (z.B. Monographien mit engem Focus und entsprechend geringem Leserkreis, Proceedings von kleinen Tagungen) mag sich ein Übergang zu elektronische Publikationsformen wie bei wissenschaftlichen Zeitschriften entwickeln. Es ist durchaus nicht unwahrscheinlich, daß (langfristig gesehen) Bücher zunächst elektronisch angeboten werden und erst dann, wenn sich genügend Nachfrage zeigt, gedruckte und leserfreundlich aufbereitete Bände erscheinen.

Bei den Preprints stellt sich die Situation jedoch völlig anders dar. Preprints haben schon heute – und sehr weitgehend – wichtige Funktionen wissenschaftlicher Journale übernommen. Mathematiker aus aller Welt tauschen Preprints schon heute elektronisch aus, oft per E-mail und zunehmend auch, indem sie Preprints auf mail- oder ftp-Servern füreinander bereitstellen (zunehmend auch im WorldWideWeb). Ein exemplarischer Kostenvergleich (siehe Anlage) macht deutlich, warum diese Art des Austausches wissenschaftlicher Ergebnisse nicht nur sehr effizient, sondern auch kostensparend, und damit für mathematische Institute fast unumgänglich ist:

Die Vervielfältigung und der Versand eines typischen Preprints auf Papier kostet ein mathematisches Institut heute

DM	8,35	für	1	Exemplar, also
<hr style="width: 100%; border: 0.5px solid black;"/>	<hr style="width: 100%; border: 0.5px solid black;"/>	<hr style="width: 100%; border: 0.5px solid black;"/>	<hr style="width: 100%; border: 0.5px solid black;"/>	<hr style="width: 100%; border: 0.5px solid black;"/>
DM	2.500,-	für	300	Exemplare.

Setzt man die Jahresproduktion eines Fachbereichs oder eines größeren Instituts im Mittel mit etwa 50 Reports pro Jahr an, so entstehen dem Institut heute Kosten in Höhe von

DM	125.000,-	für	15.000	Exemplare (50 Reports a 300 Exemplare).
----	-----------	-----	--------	---

Für die Berechnung wurde ein Report von 24 Seiten (mit handelsüblichem Einband) einschließlich eines Betrags für den Briefumschlag und den Postversand zugrunde gelegt. In elektronischer Form (als T<sub>E</sub>X -Code plus komprimiertem Postscript-Code) belegt solch ein Report etwa 200K Byte auf einem Plattenspeicher. Auf eine handelsübliche Platte von 4 GByte Speicherkapazität passen also ca. 20.000 Reports in dieser Form. Bei einer Übertragung des Reports über eine in Deutschland im WIN (dem Wissenschaftsnetz) heute gängige 2MBs-Leitung, entstehen (bei Anschluß an das Internet) damit anteilige Kosten in Höhe von

DM	0,40	für	jedes erste	und
DM	7,60	für	je 300	Exemplare, also rund
DM	500,-	für	die gesamte	Jahresproduktion

Die anteiligen Kosten für die notwendige Hard- und Software sind mit einer Abschreibungszeit von 3 Jahren in diesen Betrag mit eingerechnet.

Jede weitere Übertragung jeweils beider Formen (TEX und Postscript) des Reports kostet dann (Hard- und Software sind ja bereits anteilig abgeschrieben) nur noch

DM	0,025	pro	Einzelexemplar,	also rund
DM	25,-	für	1.000 weitere	Exemplare.

Ein mathematisches Institut in Deutschland könnte also zusätzlich praktisch alle mathematischen Institute der westlichen Welt (wenn man deren Anzahl mit etwa 1000 ansetzt) für ca. DM 25,- (an echten Selbstkosten) mit (s)einem Report versorgen.

Plattenspeicher und Workstation gehen in die Selbstkosten aber nur in Höhe von etwa DM 0,40 je Report ein. Die hauptsächlichen Kosten entstehen in Deutschland durch die anteilige Belastung des Netzes. Es ist aber zu erwarten, daß Netzkosten in wenigen Jahren etwa auf denselben Betrag fallen, der in den Vereinigten Staaten schon heute üblich ist, nämlich auf etwa ein zehntel dessen, was man in Deutschland dafür zu zahlen hat.

Bei einem Server, der in den USA steht, führt eine analoge Rechnung dazu, daß die elektronische Weitergabe eines Reports (schon) heute das mathematische Institut mit etwa nur

DM	0,40	für	das erste	und
DM	2,50	für	1.000	weitere Exemplare

belastet. Wird das Institut etwa 1.000 Rechnungen schreiben, um diesen Betrag bei seinen Partnern wieder einzutreiben? Oder ist es nicht sinnvoller, wenn sich alle mathematischen Institute auf dieser Basis gegenseitig – und “kostenlos” (jedes Institut trägt nur seine eigenen Kosten für das Informationsangebot) – mit ihren Arbeitsergebnissen versorgen?

Diese relativ ausführliche Diskussion der Kosten sollte plausibel machen, welche dramatischen Einsparungen für Anbieter wissenschaftlicher Information möglich sind, wenn sie elektronischen Netze – wie das Internet – benutzen. Die eigentlichen Kosten für den Druck können sie so auf die empfangenden Institute bzw. die Leser verlagern. Natürlich müssen diese dann, was ja auch fair ist, die Kosten für die Vervielfältigung tragen, wenn sie es nicht vorziehen, die erhaltenen Paper nur in einer Datenbank einzulagern, damit man diese bei Bedarf recherchieren und abrufen kann.

Es sollte jetzt auch verständlich sein, vor welchen neuartigen Problemen kommerzielle Anbieter wissenschaftlicher Information stehen. Angesichts ihrer echten Kosten werden sie sich um völlig neue Konzepte und Verfahren der Abrechnung bemühen müssen. Die Thematik des Copyrights ist in diesem Zusammenhang noch nicht einmal angesprochen. Das eigentliche Problem der kommerziell arbeitenden Informationsanbieter wird aber wohl sein, den anstehenden Übergang in die elektronische Welt auch organisatorisch zu verkraften. Dieser Übergang könnte bzgl. der wissenschaftlichen Journale ziemlich abrupt vonstatten gehen [Odlyzko 95]. Die Informationsanbieter werden sich aber nur schrittweise auf die neuen Verhältnisse einstellen können.

Die DMV sieht die Mathematik in diesem Zusammenhang nicht so sehr als Nutznießer dieser Entwicklung an, sondern vielmehr als Mitglied einer Gemeinschaft der Informationsabnehmer und -anbieter. Sie bezieht deshalb die Verlage, die Bibliotheken, die Zentren der Fachinformation und die Informationsbibliotheken ebenso in ihre Überlegungen (und das geplante Projekt) mit ein, wie z.B. "isolierte" Mathematiker in der Industrie [DaltzGLS 94]. Es ist ihre erklärte Absicht, alle Beteiligten zum gemeinsamen Dialog und Erfahrungsaustausch zusammenzuführen und den anstehenden Übergang in das Informationszeitalter gemeinsam zu bewältigen.

## 3.2 Zentralblatt und Math-Net

Dieser Abschnitt beschreibt Möglichkeiten der Einbettung des Referateorgans "Zentralblatt für Mathematik" und der assoziierten Datenbanken MATH und CompactMath in das Verteilte Informationssystem.

### 3.2.1 Allgemeine Zielsetzungen in bezug auf Fachinformation

Mit dem Verteilten Informationssystem soll ein elektronisches Nachrichtensystem (mit Datenbank und Hypertextanschluß) für die Verbreitung (einschl. Speicherung und Retrieval) von Nachrichten und Mitteilungen mathematischer Art entstehen. Im technischen Rahmen des Math-Net sollen sich weitere, elektronische Nachrichtensysteme (oder Hierarchien davon) auf einfache Weise einrichten und betreiben lassen. Insbesondere soll damit eine elektronische Form des Referatewesens aufgebaut werden, die von der Redaktion des Zentralblatt und ihren gegenwärtigen Referenten genutzt werden kann.

Mittelfristig geht es um die Steigerung der Attraktivität und der Konkurrenzfähigkeit des Zentralblatt durch:

- \* Integration des vollständigen Angebots an mathematischer Fachinformation, insbesondere auch von elektronisch vorliegender Information (*Vollständigkeit*). Neben mathematischer Literatur, die elektronisch bzw. im Rahmen von E-Journals angeboten wird, gewinnen hier mathematische Algorithmen (in Form von Software und Datensammlungen), entsprechende, internationale Informationssysteme sowie mathematische Darstellungen (Visualisierungen bis hin zu digitalen Videos) zunehmend an Bedeutung. Diese können nicht auf traditionelle Art und Weise publiziert werden, sondern nur noch elektronisch (neuerdings im Rahmen von globalen Hypertext- und Hypermediasystemen).
- \* Prinzipielle Beschleunigung des Referatewesens durch frühzeitigen Nachweis mathematischer Ergebnisse schon zum Zeitpunkt ihrer Verfügbarkeit (*Aktualität*). Mathematische Literatur liegt heute in Form von Preprints und elektronischen Skripten oft schon zwei bis drei Jahre vor ihrer Veröffentlichung in Papierjournalen vor. Aktualität ist eines der wichtigsten Qualitätsmerkmale eines Referateorgans.
- \* Wahrung und Steigerung der bestehenden Aussagekraft des Zentralblatt durch Einführung von Referaten (nicht nur von Abstracts) auch für elektronisch vorliegende Fachinformation, die bisher nicht referiert wird (Software, Datensammlungen etc.). Ausgewählte elektronische Dokumente (Hypertexte auf CD-ROM oder im Netz) und mathematische Algorithmen sollen nicht nur nachgewiesen, sondern ebenfalls referiert werden (*Qualität*).



- \* Verbesserung der Wirtschaftlichkeit des Zentralblatt durch weitergehende Beteiligung der mathematischen Community am Referatewesen. Fachverlage und größere Bibliotheken mit mathematischen Archiven sollen möglichst in den Prozeß der Bereitstellung bibliographischer Nachweise mit einbezogen werden (*Kostendämpfung*).

Die DMV strebt mit dem Verteilten Informationssystem an, einen Beitrag zur Anpassung des mathematischen Referate- und des Publikationswesens zu leisten. Sie erwartet, daß hier in den nächsten fünf bis zehn Jahren ein grundsätzlicher Wandel – in Richtung auf elektronische Verfügbarkeit von Fachinformation – eintreten wird.

### 3.2.2 Erweiterung des Math-Net – auch für traditionelle Publikationen

Der Einsatzbereich des Math-Net ist im Prinzip nicht auf elektronische Informationsangebote (im Internet) beschränkt. Es können auch Ankündigungen, Abstracts und Referate zu Veröffentlichungen im Sinne des traditionellen Publikationswesens über das Math-Net verteilt werden. Statt eines Hinweises (Hyperlink bzw. URL) auf den elektronischen Standort enthalten solche Beiträge dann die entsprechenden bibliographischen Angaben. Wenn mathematische Institute z.B. die Ankündigung von (traditionellen) Veröffentlichungen (auf Papier) ihrer wissenschaftlichen Mitarbeiter – mit Abstracts – zum Zeitpunkt der Annahme eines Papers über das Math-Net distribuieren, werden nicht nur die Leser der Math-Net-Nachrichten sofort über die kommende Veröffentlichung informiert, sondern die Ergebnisse der Arbeit werden auch frühzeitig in der Datenbank des Math-Net recherchierbar.

Das Math-Net soll von vornherein so ausgelegt sein, daß eine große Anzahl von Mathematikern bzw. Anwendern der Mathematik (ohne Beschränkung auf die Bundesrepublik Deutschland) elektronisch angesprochen und auch in den Prozeß des Informationsaustausches aktiv einbezogen werden kann. Dabei verschickt das Math-Net aber nicht jede Ankündigung an alle Math-Net-Leser, sondern nur an Leser mit bestimmten Interessengebieten (Vermeidung des Information Overload), denn die “Interessengebiete” sind dem Math-Net über den Mechanismus der Subscription zu gewissen Kollektionen und Indizes (MSC oder GAMS) ja im Prinzip bekannt.

Auch Verlage und Softwarehäuser können das Math-Net für Ankündigungen ihrer Produkte nutzen. Sie können auf diese Weise bibliographische oder Inhaltsangaben zu Neuerscheinungen von Büchern (auch von CD-ROM's), Inhaltsverzeichnisse von Journalen (bzw. einzelner Hefte) oder auch Beschreibungen neuer Softwareprodukte zielgenau an bestimmte Interessengruppen richten. Sie brauchen dafür nicht notwendig selbst über einen Anschluß an das Internet zu verfügen. Elektronische Post ist die einzige, notwendige technische Voraussetzung und – natürlich – die Verfügbarkeit ihrer Annoncen in einer vom Math-Net interpretierbaren, elektronischen Form (etwa in einer konkreten Form gemäß SGML-Standard). Solch ein Service kann vor allem für kleinere Verlage, die über keine eigene DV-Abteilung verfügen, äußerst nützlich sein.

Die Redaktion des Zentralblatt wird deshalb das Math-Net auf mehrfache Art und Weise für eine Neuorganisation des Referatewesens – in elektronischer Form – einsetzen können:

- \* Auswertung aller Nachrichten (Ankündigungen, Abstracts, Referate, Annoncen etc.), die über das Math-Net verteilt werden und Überführung dieser “Rohdaten” in eine dem Zentralblatt und seinen assoziierten Datenbanken (MATH, Compact-MATH, MATHDI) formal und qualitativ entsprechende Form (Steigerung der Aktualität und der Vollständigkeit).

- \* Einrichtung von Diskussionforen für Referenten des Zentralblatt zur elektronischen Organisation des Referatewesens (Standardisierungs- und Abstimmungsfragen etc.). Die Diskussionen in solchen Liste sind möglicherweise nicht für alle Math-Net-Leser "offen", weil sie vielleicht Zentralblatt-interne Themen betreffen. Nachrichtensysteme des Math-Net sollten also auch "geschlossen" eingerichtet werden können, d.h., daß letztlich nur der Moderator (er wird der Redaktion des Zentralblatt angehören) einer solchen Liste bestimmt, wer an den Diskussionen teilnehmen kann oder nicht.
- \* Solche spezifischen Nachrichtensysteme kann die Redaktion des Zentralblatt auch für den rein internen Geschäftsverkehr einsetzen, etwa, wenn es dazu kommen sollte, daß sich mehrere europäische Länder an der redaktionellen Arbeit beteiligen wollen. In diesem Falle wäre die Redaktion des Zentralblatt vielleicht selbst als (über Europa) verteilt zu denken. Denkbar wäre auch, daß solche elektronischen Nachrichten in Fragen der Abstimmung mit der Redaktion der Math. Reviews (z.B. bei der Fortschreibung des MSC-Index) hilfreich sein könnten.
- \* Aufbau einer Datenbank von Referenten des Zentralblatt, etwa dadurch, daß sich Referenten mit einer Whitepage-Angabe (siehe Opt-Net) mit entsprechender Kennzeichnung in das Math-Net eintragen. Ihre Whitepage braucht nicht allgemein, sondern nur der Redaktion des Zentralblatt zugänglich zu sein. Referenten tragen vielleicht auch Informationen darüber in ihre Whitepage ein, auf welche Journale sie in der Bibliothek ihres Instituts zugreifen können, ferner auch Angaben dazu, mit welchen Internet-Tools sie auf elektronische Informationen zugreifen. Die Redaktion des Zentralblatt könnte solche Informationen auswerten und u. U. damit eine Reduktion ihrer Kosten erreichen.

Bei hinreichender Beteiligung am Math-Net durch Mathematiker aus aller Welt werden sich auch neue Referenten für eine aktive Mitarbeit am Math-Net bzw. am Zentralblatt gewinnen lassen, insbesondere zu neuen Kategorien, wie z.B. zu Software- und Datensammlungen oder zu Multi-Media-Publikationen (Hypertexten). Durch das Math-Net besteht die Chance, daß auch solche Mathematiker Referate und Kommentare beisteuern, die nicht explizit von der Redaktion des Zentralblatt dazu aufgefordert wurden.

Der oben ausgeführte Vorschlag zielt darauf ab, die Mathematische Community in Bezug auf das Math-Net und das Zentralblatt zu involvieren. Die Nutzung der von der Community erbrachten Leistungen und Beiträge sollte deshalb kostenfrei durch die Community selbst möglich sein. Die vom Zentralblatt auf der Grundlage von "Rohdaten" erstellten "Qualitätsdaten" (die dann in die Datenbanken CompactMATH, MATH und in die gedruckte Version des Zentralblatt integriert werden) sollten dagegen nur bei einer angemessenen finanziellen Beteiligung genutzt werden können.

### 4 Das Skelett der Informationsstruktur, Hierarchie der Kollektionen

Dieses Kapitel entwickelt den Kern der vertikalen Informationsstruktur – inhaltlich – aus der Aufgabenstellung des Verteilten Informationssystems. Er beinhaltet einen ersten Vorschlag zur homogenen Strukturierung der verteilt organisierten, mathematischen Information (1. und 3. Dimension der im zweiten Kapitel bereits diskutierten Informationsstruktur).

Diese technische Struktur des Math-Net folgt aus der Art der bei den Partnern vorliegenden Originaldokumente und Informationsangebote, die wieder direkt aus den Zielsetzungen bzw. den konkreten Aufgabenstellungen der Teilaufgaben und Teilprojekte resultiert. Wir wollen sie uns kurz vergegenwärtigen.

#### 4.1 Teilaufgaben und Teilprojekte

Als Aktivitäten – mit Focus auf mathematische Fachinformation – sind geplant (gemäß [Grötschel 94]):

- \* Teilaufgabe “Elektronische Informationsstationen”: Aufbau und Betrieb von Informationsstationen (eine je Partner) und Clients (auf allen Workstations) bei den Partnern auf der Basis der im Internet verbreiteten Informationsdienste. Errichtung einer offenen, technischen Infrastruktur für Informationsdienste der Partner zur effizienten Informationsvermittlung auch für die DMV und Fachgruppen der DMV; transparente Darstellung von Leistungs- und Serviceangeboten.
- \* Teilaufgabe “Verteilte Preprints”: Bereitstellung von Volltexten (Reports, Bücher und Skripte) zum Zeitpunkt der Fertigstellung in den Institutionen. Das Zentralblatt integriert die Abstracts in die Datenbank MATH und erstellt ggf. Referate.
- \* Teilaufgabe “Verteilte Software- und Datensammlungen”: Erschließung örtlicher Mathematischer Software für die Mathematik und für andere Fachgebiete (Naturwissenschaften, Ingenieurwissenschaften, etc.). Bereitstellung des nationalen und internationalen Angebots (eLib, NetLib, Reduce Network Library etc.).
- \* Teilaufgabe “Informationssysteme in der Mathematik”: Zusammenstellung mathematischer Informationssysteme – weltweit – in Hypertext-Dokumenten, die online per Maus-Click anwählbar sind (auch suchbarer Index zu externen Systemen). Öffnung des Verteilten Informationssystems in Richtung auf europäische und nordamerikanische Regionen.
- \* Pilotvorhaben “Elektronische Referate”: Aufbau eines Mailverteilers und einer elektronischen Referate-Organisation für das “Zentralblatt für Mathematik” zur Beschleunigung des Referatewesens.
- \* Pilotvorhaben “Elektronische mathematische Journale”: Entwicklung eines qualitativ hochwertigen, exportierbaren Modells für das elektronische Angebot mathematischer Journale. Modellhafte Erprobung neuer Publikations- und Vertriebskonzepte.
- \* Pilotvorhaben “Historische Dokumente und Bücher”: Einscannen und Bereitstellung wertvoller mathematischer Literatur aus der Schatzkammer der Bibliotheken. Kostengünstige Erstellung von Reprints.

- \* Pilotvorhaben: “Lebendes Mathematisches Museum”: Aufbau einer Kernorganisation eines elektronischen (“virtuellen”) Mathematischen Museums und Ausweitung des Konzepts auf andere Länder.
- \* Pilotvorhaben: “Zugriff auf elektronische Kataloge von Bibliotheken”: Anschluß der Kataloge von ausgewählten Fachbibliotheken und zentralen Bibliotheken an das Internet. Installation von Clients bei allen Partnern für den Zugriff auf diese Kataloge.
- \* Pilotvorhaben: “Elektronische Dokumentenlieferung”: Elektronische Auslieferung von (auch eingescannten) Artikeln (durch ausgewählte Partner), die bisher nur im Rahmen des traditionellen Publikationswesens angeboten wurden. Erprobung von elektronischen Bezugsmöglichkeit auch älterer, mathematischer Literatur.
- \* Teilaufgabe “Elektronische Projektorganisation”: Elektronischer Verbund der Partner bzw. der Projektmitglieder zum Zwecke des effizienten, informellen Nachrichtenaustausches.

## 4.2 Verteilte Kollektionen von Informationen

Die folgende Tabelle stellt eine erste Übersicht und Gliederung möglicher elektronischer Informationsangebote dar, die etwa ein Fachbereich in das Verteilte Informationssystem einbringt. Zunächst sollen nicht die zentralen (oder weltweiten) Informationsangebote betrachtet werden, sondern nur örtliche Beiträge. Wenn in dieser Aufstellung vom Mathematischen Museum die Rede ist, geht es also nicht um die zentralisierte, vollständige Sammlung, sondern zunächst nur um entsprechende Beiträge des Fachbereichs.

Den Originaldokumenten (z.B. Hypertexte oder Files im ftp-Archiv), die dezentral vorgehalten werden, sind Referenzdokumente zugeordnet, die im zentralisierten Teil des Math-Net verfügbar und dort recherchierbar sein sollen. Diese sind in der dritten Spalte angeführt. Sie enthalten, damit sie im Math-Net auch recherchierbar sind, neben einer Titelzeile auch Klassifikationen, Keywords, einen knappen, beschreibenden Text und sie sind per Hypertext-Link mit den ihnen entsprechenden Originalen verbunden.

Referenzdokumente sind i.d.R. reine ASCII-Texte, die in der Form eines Hypertextes vorliegen (HTML- oder HTF-Format). Sie können – in das Math-Net – aber auch als einfacher Text eingebracht werden, am einfachsten per E-mail in einer “schwach” strukturierten Form (weiteres dazu weiter unten). Ein Fragezeichen in der dritten Spalte deutet an, daß die örtliche und die zentrale Bereitstellung der Angaben optional ist und in den Händen des Fachbereichs (bzw. der Person, der das Dokument zuzuordnen ist) liegt.

Die letzte Spalte zeigt, nach welchem Schema diese Gruppe (vielleicht) zu klassifizieren ist. Wenn ein Original nicht klassifizierbar ist, ist es der Klasse “unclassified” zugeordnet. Ein Fragezeichen in der vierten Spalte zeigt an, daß nicht alle Referenzdokumente der betreffenden Kollektion entsprechend klassifizierbar sind.

Kollektion	Original	Referenzdokument	Schema
Organisation	Fachbereich	Prospekt	MSC?
	Dekanat/Sekretariat	Personalangaben/Whitepage?	
	Institut	Kurzbeschreibung	
	Leitung/Sekretariat	Personalangaben/Whitepage?	
	Mitarbeiter	Personalangaben/Whitepage?	
	Projekt	Aufgabenbeschreibung	
	Bibliothek	Sammelgebiete	
	Service-Einheit	Aufgabenstellung	
	Lageplan (Bild)	Ortsbeschreibung	
	Raumplan (Bild)	Tourbeschreibung	
Publikationen	Preprint/Report	Abstract	MSC
	Buch (auch gescannt)	Titelseite/Inhaltsverzeichnis	
	Skript (auch gescannt)	Titelseite/Abstract	
	Hausmitteilungen	Titelseite/Inhaltsverzeichnis	
	E-Journal	Editorial	
	Referat/Review	das Referat	
Software	Programm	Readme	GAMS
	Programmbibliothek	Index (netlib-Form)	
	Softwaresystem	Datenblatt	
	Testdaten	Kurzbeschreibung	
	Datensammlung	Index (netlib-Form)	
Info-System	WWW Top-Page	diese Page	GAMS/MSC?
	Gopher Top-Menü	dieses Menü	
	ftp-Archiv	Kurzbeschreibung	
	telnet-Archiv	Kurzbeschreibung	
	News-Angebot	Kurzbeschreibung	
Math-Museum	histor. Text (gescannt)	Abstract/Referat	GAMS/MSC?
	histor. Buch (gescannt)	Titelseite/Inhaltsverzeichnis	
	Photo	Kurzbeschreibung	
	Mathematiker-Photo	Kurzbiographie	
	Video	Kurzbeschreibung	
	Computergraphik	Referat oder Top-Page	
	Living Book	die Top-Page	
	Report/Skript	Abstract/Titelseite	
	Math. Experiment	die Top-Page	
Personen	Mitglied des Fachb.	Personalangaben/Whitepage?	
	Gast	Personalangaben/Whitepage?	
	Student. Mitgl.	Personalangaben/Whitepage?	

Diese Zusammenstellung ist nicht vollständig und sie soll auch nicht implizieren, daß ein Fachbereich nicht darüber hinausgeht. Natürlich können Fachbereiche ihr Informationsangebot durch lokale "Spezialitäten" ergänzen, aber sie können auch Kollektionen auslassen, zu denen sie keine originären Beiträge erbringen. Wenn Fachbereiche allerdings

eigene Kollektionen und Informationen einbringen, die für alle Partner und “Leser” des verteilten Informationssystems von Interesse sind, sollte dieses in einer Form geschehen, die mit der Struktur des Math-Net verträglich (harmonisierbar) ist.

Es sollte jetzt außerdem deutlich sein, daß das Informationsangebot des Verteilten Informationssystem sehr heterogen ausfallen wird. Obendrein können, technisch gesehen, die Originaldokumente aus Files der verschiedensten Formate bestehen bzw. zusammengesetzt sein. Hier kommen in erster Linie in Betracht:

$\text{T}_{\text{E}}\text{X}$ , $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$	mathematische Texte
Postscript, DVI	mathematische Texte und Grafiken
ASCII, Text	einfache Texte (etwa Readme oder Index-Files)
MATH, MR, $\text{BIB}_{\text{E}}\text{X}$	bibliographische Formate (ASCII Darstellung) des Zentralblatt oder der Math. Reviews bzw. das weit verbreitete $\text{T}_{\text{E}}\text{X}$ -orientierte Format.
HTML, HTF	Hypertext-Formate (ASCII Darstellung)
SGML	strukturierte Dokument gemäß SGML-Standard
netlib-index	Text-Formate für Software-Library-Indizes
whitepage	Adresseninformation (nicht standardisiert)
MPEG	Video-Formate
JPEG, GIF, TIFF	Bild-Formate, Pixelgrafik-Formate

Auch diese Zusammenstellung ist nicht vollständig. Im Prinzip sollen alle Formate zugelassen sein, die sich im Internet bzw. im WorldWideWeb präsentieren lassen. Dazu gehören heute insbesondere auch die Formate gemäß MIME-Standard (Multi-Media-Mail) [Borenstein 92]. Das System sollte auch für künftige Formate offen sein, d.h. an diese angepaßt werden können.

Der Begriff “WorldWideWeb” ist dabei nicht im engeren, technischen Sinne zu verstehen, also nicht als WWW-Server oder WWW-Browser, sondern als das gesamte, weltweite Netz der über WWW-Tools erreichbaren Informationen. Dazu gehören z.B. auch Angebote, die mittels spezieller Datenbanken oder Dokumentenarchive in das Web eingebracht werden (auch wenn das spezifische Datenbanksystem selbst nicht zu den Internet- bzw. WWW-Tools zählt).

Die Partner des Projekts entscheiden selbst, auf welche Art und Weise sie ihre Informationen bereitstellen, insbesondere entscheiden sie, welche Systeme (Server) sie dafür verwenden. Hier kommen, neben Hypertext-Systemen, wie WWW oder Hyper-G, auch ftp- und Gopher-Server in Betracht. Auch die Anbindung von speziellen Datenbanksystemen sollte möglich sein, die vor allem im Bibliotheksbereich auftreten können.

### 4.3 Ein Vorschlag zur Strukturierung des Informationsangebots

Dieser Abschnitt konzentriert sich auf die Bereitstellung von Information mittels Hypertext-Systemen vor Ort. Diese Systeme erlauben heute eine sehr einfache Art des Zugriffs auf Information. Sie lassen sich i.d.R. ohne größeren Lernaufwand benutzen.

Bei Hypertexten mittleren (ab ca. 100 Dokumenten) oder größeren Umfangs (ab 1000 Dokumenten) besteht jedoch ein Problem, das in der Literatur als “Lost in Hyperspace Syndrom” bekannt ist. Leser verirren sich leicht in größeren Hypertexten, insbesondere,

wenn diese vielfältig vernetzt sind, weil sie sich dann kein genaues Bild von der Struktur des gesamten Informationsangebots mehr machen können.

Die Arbeit von D. M. Edwards und L. Hardmann [EdwardsH 89] z.B. diskutiert dieses Problem, indem es das kognitive Verhalten von (ungeübten) Lesern an inhaltlich vergleichbaren Hypertexten untersucht, die aber unterschiedlich strukturiert sind. Die Autoren kommen zu dem Schluß, daß hierarchisch strukturierte Informationsangebote dem Leser die Navigation wesentlich erleichtern und solchen Hypertexten deutlich überlegen sind, die allein nach einem Index strukturiert sind. Beide Verfahren sind wieder einem "beliebig" vernetzten Hypertext vorzuziehen, bei dem die Versuchspersonen die größten Schwierigkeiten hatten, sich darin zu orientieren.

Auch die obige Zusammenstellung möglicher Kollektionen legt eine hierarchische Anordnung als hauptsächliches Strukturierungsmittel nahe. Abb. 2 stellt einen ersten Vorschlag hierzu als n-ären Baum dar. Sie enthält nur die Knoten des Baumes und keine Blätter, also noch keine Dokumente, und sie ist auch nicht als vollständig anzusehen. Ein Fachbereich kann auf jeder Ebene eigene Ergänzungen machen, und er sollte auch die Möglichkeit haben, zwischen zwei Knoten Ergänzungen anzubringen, solange der gesamte Baum nur inhaltlich konsistent bleibt.

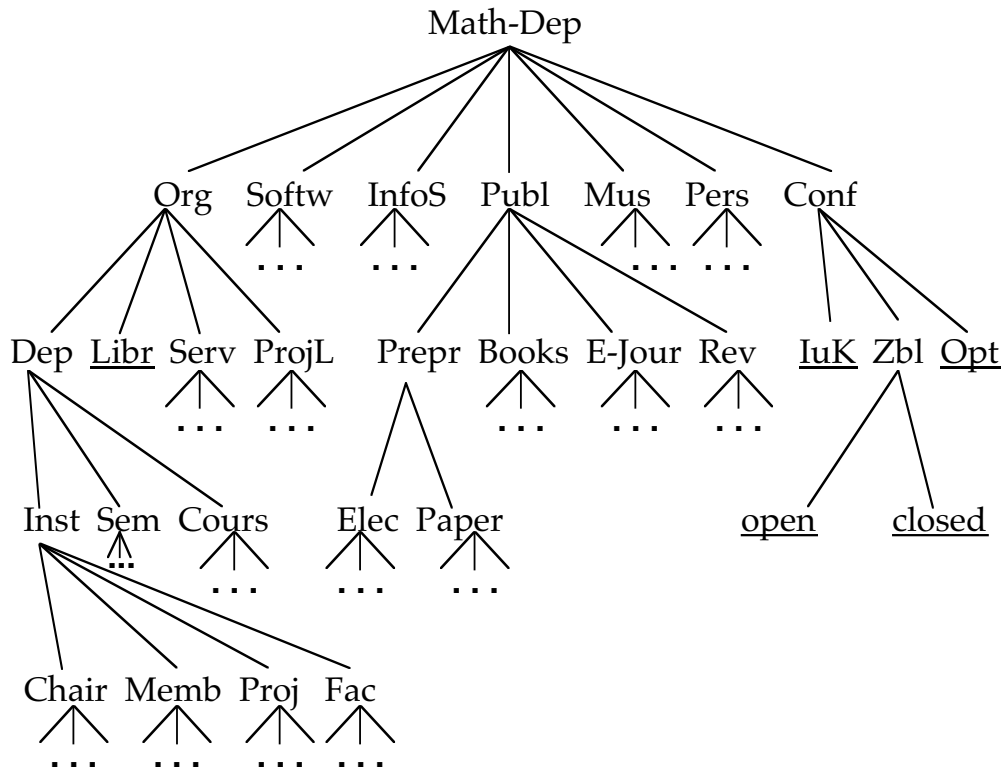


Abb. 2: Hierarchische Struktur von Kollektionen

Wenn sich z.B. in einem Fachbereich mehrere Bibliotheken befinden, sollte der Fachbereich den Verweis auf die Datenbank seiner Bibliothek (mit dem unterstrichen Text "Libr" ist eine telnet-Verbindung gemeint) durch eine entsprechende Kollektion von Bibliotheken ersetzen. Wenn ein Fachbereich (Kürzel: "Dep"; gemeint ist der Knoten unterhalb des Knotens "Org" für das organisatorische Informationsangebot) organisatorisch in mehrere

Teilbereiche gegliedert ist, z.B. in “Mathematik A”, “Mathematik B”, dann kann er zwischen “Org” und “Dep” etwa einen Knoten (also eine neue Kollektion) “Areas” anbringen, der selbst wieder etwa auf Subknoten “Math-A”, “Math-B” etc. verweist. Und, wie bereits angeführt: Wenn ein Fachbereich keinen Beitrag zum Lebenden Mathematischen Museum (“Mus”) leisten kann, tritt der entsprechende Ast bei ihm einfach nicht auf.

In diesem Sinne ist auch zu verstehen, wie Institutionen, die anders als Fachbereiche aufgebaut sind, sich harmonisch an solch eine Struktur anpassen können. Das Konrad-Zuse-Zentrum ist z.B. eine Forschungseinrichtung, das die auf Fachbereiche bezogenen Begriffe in der folgenden Weise interpretieren wird:

Math-Dep:	Fachbereich	⇒	Forschungseinrichtung
Inst:	Institut	⇒	Abteilung
Proj:	Projekt	⇒	Projekt- oder Arbeitsgruppe

Nun ist das ZIB in mehrere Bereiche aufgeteilt: Es wird also zwischen “Org”, der organisatorischen Gliederung, und “Inst”, das der Gliederung in Abteilungen entspricht, noch die Kollektion

Area:	⇒	Bereich
-------	---	---------

einfügen. Beim ZIB gibt es ferner Projekte, an denen mehrere Abteilungen und sogar mehrere Bereiche beteiligt sind. Solche Projekte sollten dann auf der entsprechenden Ebene der Hierarchie auftreten, wobei noch zu diskutieren ist, ob sie auch auf der Ebene der Abteilung auftreten sollten. Den Fall, daß es auch Projekte gibt, die fachbereichsübergreifend gelagert sind, lassen wir an dieser Stelle einmal außer Betracht. Er könnte so gelöst werden, daß in jedem der beteiligten Fachbereiche ein Verweise auf die Projektinformation existiert, die von einem – dem federführenden – Fachbereich verwaltet wird.

In jedem Fall wird es Dokumente geben, die nur ein einziges Mal vorhanden sind und die an mehreren Stellen der Hierarchie auftreten. Dokumente, die Personen beschreiben, sind nur ein Beispiel. Dem obigen Vorschlag zufolge sollte der Zugriff auf Adresseninformationen möglichst direkt erfolgen können. Auf der obersten Ebene der Hierarchie befindet sich dem entsprechend eine Kollektion, in der alle Mitglieder des Fachbereichs auftreten (“Pers”). Diese könnte einerseits alphabetisch geordnet

All:	Alle Mitglieder des Fachbereichs
------	----------------------------------

und andererseits nach Funktionen, also etwa gemäß

Dean:	Dekanat
Dirs:	Institutsdirektionen
Secr:	Sekretariate
Prof:	Professoren
Sci:	Wissenschaftliche Mitarbeiter
Guest:	Gäste
Doc:	Doktoranden
Stud:	Studentische Mitarbeiter

zusammengestellt sein. Angaben zu Personen finden sich aber, z.B. im organisatorischen Teil (“Org”), praktisch auf allen Ebenen. Weiterhin ist auch denkbar, daß ein elektronisches Paper, das unter den Preprints (“Prep”) zu finden ist, auch im Museum auftritt. Die oben angeführte Struktur ist also kein Baum, sondern ein gerichteter, azyklischer Graph.



Zum gegenwärtigen Zeitpunkt ist noch nicht geklärt, ob sich die – schon in Deutschland – sehr unterschiedlichen Strukturen auf diese Art und Weise alle in solch eine Struktur harmonisch einbetten lassen. Die Ausarbeitung eines detaillierten Vorschlags steht noch aus. Und, sie muß natürlich mit allen beteiligten (potentiellen) Partnern im einzelnen abgestimmt werden.

#### **4.4 Anmerkungen zur Implementierung mit WWW- oder Hyper-G-Servern**

Der Strukturvorschlag soll nicht implizieren, daß das Informationsangebot nicht noch weiter vernetzt werden kann. Der Vorteil von Hypertextsystemen besteht ja gerade darin, daß die Information in gewisser Weise “beliebig” vernetzt und damit von den Lesern “assoziativ” erschlossen werden kann.

Bei Verwendung eines WWW-Servers kann die hierarchische Struktur jedoch nur mit Hyperlinks (lokale URLs) implementiert werden und es ist unklar, wie strukturellen Hyperlinks von den anderen zu unterscheiden sind. Die entsprechende Hypertext-Sprache HTML kennt nur einen einzigen Typ von Links, eben die unidirektionale URL. Abgesehen davon, daß die Autoren von HTML-orientierten Hypertexten alle hierarchischen Links explizit in die betreffenden Dokumente eintragen müßten, ist mit HTML-orientierten Hypertexten noch ein weiterer Nachteil verbunden.

Das Verschieben oder Löschen eines einzigen Dokumentes innerhalb der einmal aufgebauten HTML-Struktur hat zur Folge, daß man nun in allen HTML-Dokumenten, die darauf verweisen, die entsprechenden Änderungen an den URLs explizit vornehmen muß, damit diese nicht ins Leere verweisen. Eine konsistente strukturelle Änderung von Directories, in denen solche HTML-Texte liegen, ist damit nur sehr schwer möglich, wenn nicht sogar – bei umfangreichen Mengen von Dokumenten – praktisch ausgeschlossen.

Insbesondere deshalb empfiehlt sich die Verwendung von Hyper-G, eines Hypertextsystems der zweiten Generation, das bidirektionale Links verwendet. Die oben ausgeführte hierarchische Struktur läßt sich direkt mittels Hyper-G-Kollektionen realisieren, ohne daß dafür in den Dokumenten ein einziger Link gesetzt werden müßte. Solche Kollektionen lassen sich auch umstrukturieren, sei es dadurch, daß man einzelne Dokumente oder daß man ganze Kollektionen verschieben will.

Hyper-G-Kollektionen erscheinen bei Präsentation (via WWW-Gateway) im WorldWideWeb unter allen Browsern, wie z.B. in Mosaic, allerdings wieder als URL-Hyperlinks, die nicht von zusätzlichen, “beliebigen” Links zu unterscheiden sind. In diesem Falle ist innerhalb der WWW-Welt (ohne nähere Kenntnis der Hyper-G-Welt) der Unterschied zwischen “strukturellen” und “beliebigen” Hyperlinks nicht sofort auszumachen. Die Autoren von Hyper-G-orientierten Texten sollten also ebenfalls mit zusätzlichen Links sorgfältig umgehen, wenn sie den Lesern aus dem WorldWideWeb das “Lost in Hypertext Syndrom” ersparen wollen.

Die vollen, strukturellen Vorteile von Hyper-G können Leser nur bei Einsatz eines entsprechenden Hyper-G-Browsers (Harmony oder Amadeus) nutzen. Diese trennen “beliebige” Hyperlinks in modularer Weise von den hierarchischen Strukturen (Hierarchie von Kollektionen), die in einem anderen Fenster dargestellt werden (dem Session Manager). Die Verwendung von Hyper-G bietet, neben einigen anderen Dingen, außerdem noch die Möglichkeit, daß Leser in den hierarchisch geordneten Kollektionen auch hierarchisch suchen können. Diese Anforderung, die auch für das Math-Net wünschenswert und erforderlich ist, diskutiert der nächste Abschnitt.

## 4.5 Suchstrategien in Hypertextsystemen

Leser von Hypertexten werden normalerweise nicht immer das gesamte Informationsangebot durchgehen. Das wollen sie – vielleicht – nur das erste mal, wenn sie ein neues Institut “besuchen” oder im WorldWideWeb ein weiteres Informationssystem finden (aber auch nur dann, wenn das Angebot nicht zu umfangreich ist).

### 4.5.1 “Lost-in-Hypertext-Syndrom” – erneut

Leser haben spezifische Interessen und Themen, zu denen sie die für sie wertvolle Informationen suchen; [Lügger 94] diskutiert entsprechende Anforderungen für die Mathematik. Sie stehen damit vor folgenden Problemen:

- \* Wie finde ich etwas zu meinem Thema?
- \* Wie finde ich es ohne großen Zeitverlust (effizient)?

Jeder Leser, der das WorldWideWeb etwas intensiver einsetzt, hat sicher auch schon die Erfahrung gemacht, daß er etwas Interessantes gesehen hat, und nun vielleicht nur noch die entsprechende “Home Page” kennt, aber nicht mehr herausfinden kann, wo die eigentlich interessante Stelle war (man kann ja nicht alles in seiner Hotlist speichern), also:

- \* Wie finde ich etwas wieder?

Selbst geübte Leser, die sich das vollständige Angebot einer Institution anschauen wollen, können sich nicht sicher sein:

- \* Habe ich alle wichtige Information gesehen?

Denn wer kann schon eine “beliebig” vernetzte Struktur im Kopf behalten? Noch schwieriger wird es, wenn ein Leser etwas sucht, was im Angebot aber vielleicht nicht enthalten ist:

- \* Wie entscheide ich, ob die Suche ergebnislos ist?

Bei Problemen dieser Art, die in größeren Hypertextangeboten typischerweise auftreten, hilft i.d.R. auch eine hierarchische Struktur nicht mehr. Zur Lösung solcher Fragen bieten WWW-Server eine Datenbankschnittstelle an.

Im Internet werden hier meist einfache Indizes eingesetzt, die eine Suche nach Schlüsselwörtern in den Titelzeilen von Dokumenten verwenden, die linear in einer Liste angeordnet sind. Etwas leistungsfähiger ist schon die Anbindung eines WAIS-Servers, der es zusätzlich erlaubt, eine Volltextindizierung der HTML-Dokumente anzubieten. Mit WAIS kann aber immer nur die gesamte Information eines Servers, also i.d.R. eines Fachbereichs, durchsucht werden, was beim Leser leicht zum “Information Overload” führt, weil zu viele Ergebnisse geliefert werden, die dann immer einzeln inspiziert werden müssen.

Wer sich z.B. nur für Reports interessiert, will natürlich nicht mit Information über Software, Datensammlungen oder gar Angaben zu Personen überhäuft werden. Oder, Leser, die sich mittels Mausclicks bis zur Seite eines Instituts vorgearbeitet haben, wollen vermutlich nichts mehr über “Informationssysteme Weltweit” wissen, die sich mit Sicherheit auch im Informationsangebot des Instituts befinden, aber aus diesem “herausführen”.

### 4.5.2 Hierarchische Suche in hierarchischen Kollektionen

Der Grundgedanke der hierarchischen Suche in hierarchischen Kollektionen von Hypertext-Dokumenten geht auf Hyper-G zurück [KappeM 93b]. Dieser Artikel diskutiert ihn jedoch unabhängig von Hyper-G, weil er die Perspektive des Benutzers in den Vordergrund stellt. Insbesondere die angestrebte Integration der zentralen Math-Net-Komponenten (Nachrichten-, Hypertext- und Datenbank-Teil) stellt Anforderungen, die sich mit der gegenwärtigen Implementierung von Hyper-G nicht so ohne weiteres erfüllen lassen.

Um ein Beispiel für das Retrieval von Information zu nennen: Hyper-G bietet zwar sehr allgemeine (für alle möglichen Informationsmodule geltende), aber doch nur wenige (vom System vorgegebene) Kategorien, (Title, Keywords, Autor eines Hypertext-Dokuments etc.) auf die sich eine Suche beziehen kann. Klassifikationen fehlen ganz – sie müssen als Schlüsselworte codiert werden, wenn man die Suche nach ihnen erlauben will. Daraus ergeben sich aber nur eingeschränkte Möglichkeiten der Suche nach Subkategorien, z.B. des MSC-Index und Kombinationen davon.

Ferner: Die Suche nach Literatur verlangt nach anderen, spezifischen Kategorien als die Suche nach Software. Es ist in Hyper-G nicht möglich, auf bestimmten Dokumentemengen weitere, anwendungsspezifische Suchkategorien zu definieren. Diese Möglichkeit würde in Hyper-G eine Erweiterung der Konzepte erfordern, auch wenn sie prinzipiell realisierbar erscheint.

Die in Abb. 2 dargestellte Struktur könnte als Hypertext der folgenden Form gedacht bzw. realisiert werden (siehe Abb. 3). Dieses Bild entspricht der vertikalen (und der im 2. Kapitel angeführten dritten) Dimension. Die einfachen Pfeile stellen die Kollektionshierarchie dar und die doppelten Hypertext-Links zu der eigentlichen Information, also zu Dokumenten oder Dialogverbindungen zu Datenbanken, deren Inhalt außerhalb des hierarchischen Suchbereichs liegen kann.

Die grau unterlegten Fenster stellen Eingabemöglichkeiten für Suchbegriffe dar, die sich auf die Information des jeweiligen Hypertextkollektion und ihrer Subkollektionen einschließlich aller auf diesem Wege erreichbaren, indizierten Dokumente beziehen – und nur auf diese. Leser, die sich mittels Mausclicks über das Top-Dokument bis zu der Kollektion hindurchgewählt haben, auf die der Department-Verweis “Dep” zeigt, können dort eine Suchanfrage formulieren, die sich auf alle Informationen zum Department, und nur auf diese, bezieht. Alle anderen Informationseinheiten bleiben bei dieser Suche ausgeblendet.

Die gezielte, hierarchische Suche in hierarchischen Kollektionsstrukturen ist für Leser von örtlichen Informationsangeboten eine wesentliche Hilfe zur Vermeidung des “Information Overload”. Sie ist insbesondere für Leser des (zentralisierten) Math-Net-Teils, der umfangreiche Mengen von Informationseinheiten (quer durch ganz Deutschland) anbietet, von besonderem Interesse. Hierbei tritt jedoch ein weiteres Problem auf, das sich bei Einsatz von Hyper-G derzeit nicht lösen läßt.

Das Suchergebnis in großen Informationsmengen kann selbst wieder recht umfangreich sein. Aus Sicht der Leser ist die Forderung sinnvoll, daß die Ergebnismenge wieder genauso, wie die durchsuchte Kollektion, hierarchisch strukturiert wird und damit die Auswahl von Subkollektionen erlaubt und mit detaillierteren Anfragen weiter durchsucht werden kann. Hyper-G liefert das Suchergebnis jedoch als (begrenzte) lineare “Liste” von Dokumenten und Kollektionen im Suchfenster aus. Ein Weitersuchen in Suchergebnissen, wie man es aus der Verwendung von Datenbanken kennt, ist damit nicht mehr möglich.

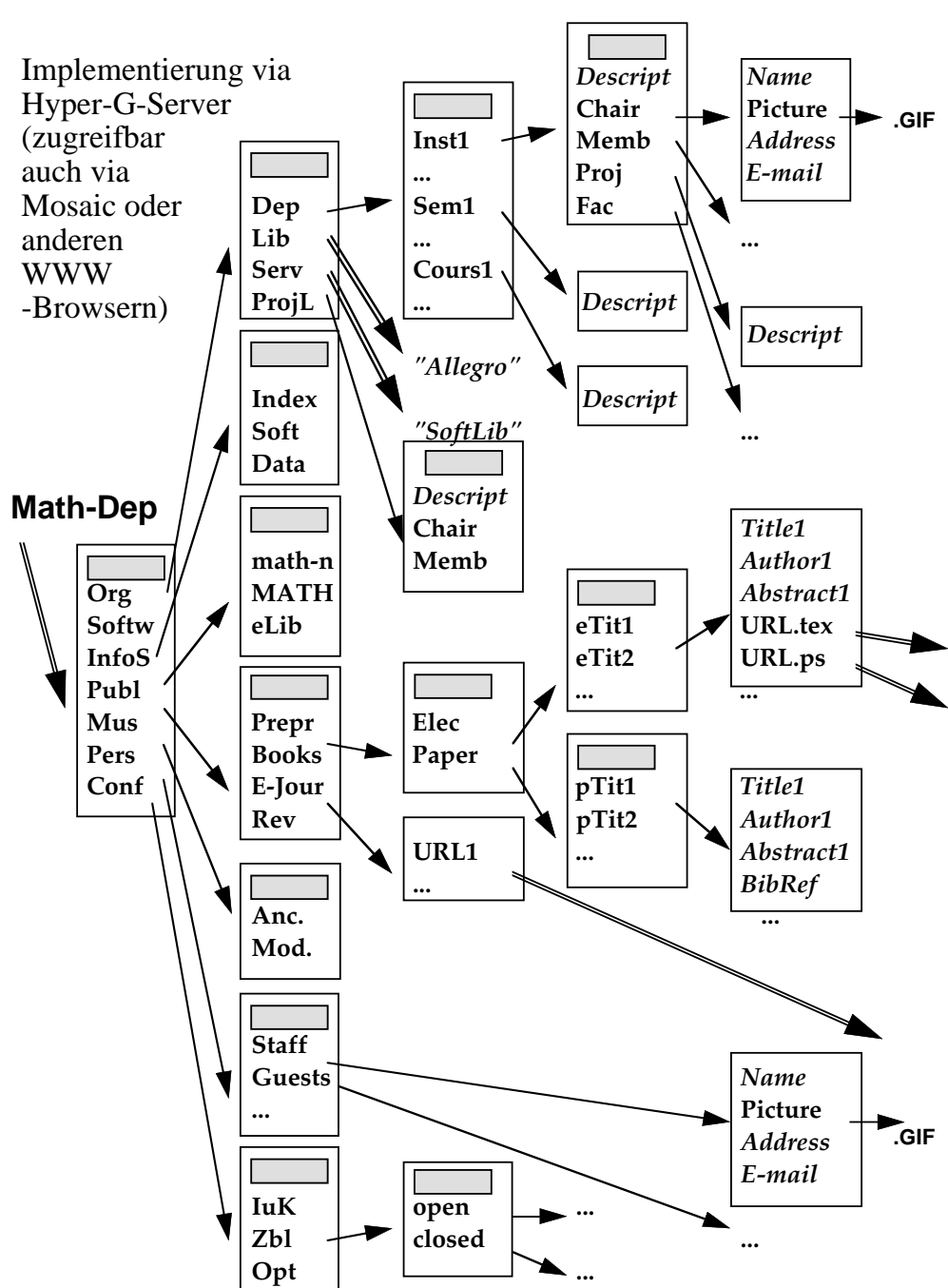


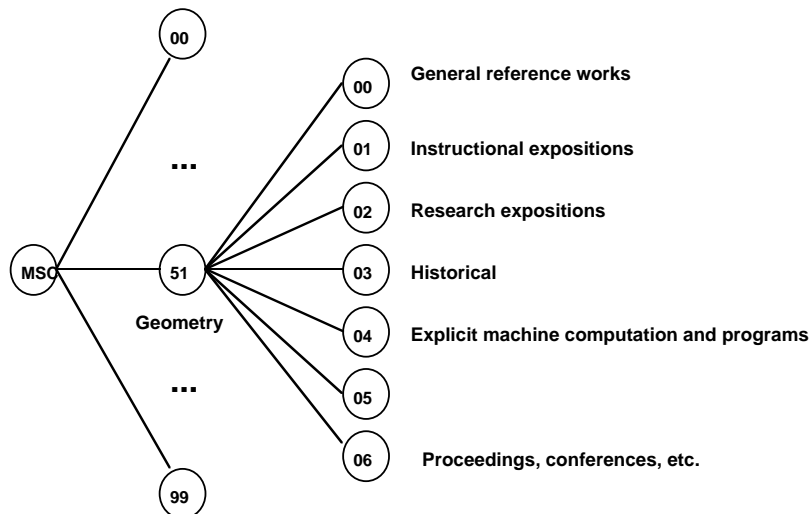
Abb. 3: Kollektions-orientierte hierarchische Suche

Bei der Suche in sehr umfangreichen Informationsmengen, die obendrein heterogen strukturiert sind, treten die Begrenzungen gegenwärtiger, globaler Hypertextsysteme besonders deutlich zutage. Der oben erwähnte Datenbankmodul des Math-Net soll Fragen dieser Art lösen. Der Einsatz von Hyper-G bei den Fachbereichen bzw. den Partnern des Projekts ist nur ein erster Schritt in Richtung auf ein verteiltes Informationssystem.

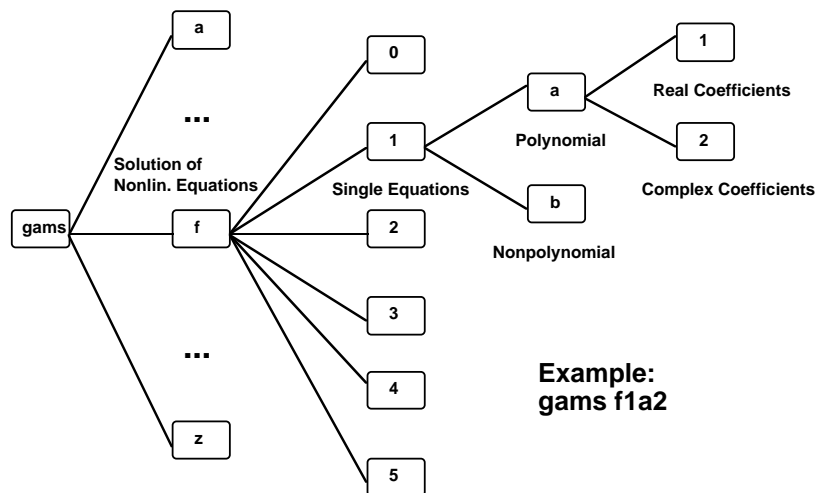
## 5 Hierarchien und Klassifikationsschemata, Automatisierungsfragen

Die Verwendung des MSC-Index [FIZ-KA 94] ist eine wichtige Hilfe insbesondere bei der Suche nach mathematischer Literatur. In Bezug auf mathematische Software- und Datensammlungen spielt der GAMS-Index eine vergleichbare Rolle [NIST]. Dabei handelt es sich um mehrstufige Hierarchien von mathematischen Begriffen, deren Verwendung im Bereich der Mathematik weltweit verbreitet ist; siehe Abb. 4 für einen sehr kleinen Ausschnitt. Der GAMS-Index ist allerdings weniger verbreitet und anerkannt; R. Boisvert, der Autor einer ersten Fassung, nimmt obendrein eine Revision in Angriff.

### I. MSC: "Mathematics Subject Classification"



### II. GAMS: "Guide to Available Mathematical Software"



### III. "unclassified"

Abb. 4: Verwendung von Klassifikationsschemata

Beide Schemata bieten sich auch für die Anordnung von Hypertexten an, die von mathematischen Lesern und Anwendern mathematischer Software ohne besondere Instruktionen

verstanden wird – jedenfalls in den ihnen bekannten fachlichen Ausschnitten.

Beide Klassifikationsschemata sind jedoch sehr breit angelegt und ein einzelner Fachbereich wird z.B. die meisten Klassen leer lassen müssen, wenn er seine Information mit ihrer Hilfe strukturiert. Die einzelnen Partner des Projekts werden deshalb ihre Informationsangebote zwar entsprechend klassifizieren (indem sie die MSC-Kategorie in den Stichworten angeben), damit die Leser diese Indizes bei der Suche verwenden können, sie werden aber vermutlich selbst keine Kollektionshierarchien aufbauen, in denen die Informationen nach dem MSC- oder GAMS-Index sortiert sind.

## 5.1 Hierarchische Kollektionen und MSC-Index im Hypertext-Teil

Dieser Abschnitt skizziert zunächst nur den Einsatz des MSC-Index in Verbindung mit dem Hypertext-Anteil; für den GAMS-Index gilt entsprechendes. Die Zielvorstellung – aus Sicht der Leser – scheint offensichtlich und recht einfach (darstellbar) zu sein.

Auf der obersten Ebene findet ein Leser nur eine Übersicht mit den Hauptklassifikationen, die der “Table of Contents” entspricht, wobei er auf dieser Stufe vielleicht schon eine Angabe darüber erwartet, wieviele Informationseinheiten in den jeweiligen Abschnitten zu finden sind. Wählt er dort eine Kategorie, etwa msc-90 “Economics, operations research, programming, games”, so findet er zur weiteren Detaillierung seiner Suche dort zunächst die Überschriften der Abschnitte 90-XX, 90Axx,...,90Dxx. Erst bei Anwahl von z.B. des Hypertextdokumentes 90Cxx “Mathematical Programming” wird er dann eine vollständige Detaillierung finden, ebenfalls mit Angabe der zu erwartenden, potentiellen “Treffer”. Auf dieser Ebene sollten auch die im MSC-Index definierten Querverweise zu verwandten Abschnitten [hier: 49Mxx und 65Kxx] in Form von Hypertext-Links zu finden sein.

Bei Auswahl der Kategorie 90C06 “Large Scale Problems” wird der Leser eine “Zusammenstellung aller Items” finden, die das Math-Net zu diesem Thema zu bieten hat, aller diesbezüglichen elektronischen und nicht-elektronischen Reports, aller Forschungsgruppen und Projekte (in Deutschland oder auch außerhalb, sofern dem Math-Net solche Aktivitäten angezeigt wurden), vielleicht auch aller entsprechenden “Produkte” (Softwareentwicklungen, Testdaten etc.) oder möglicherweise auch aller Personen (Kontaktadressen), die an diesem Thema interessiert sind.

Vorausgesetzt, daß es sich bei 90C06 um ein “aktives” Gebiet handelt, bei dem mit der Zeit eine ganze Reihe von Einträgen existieren, (dann) entsteht ein Problem, das allein mit den heutigen Hypertext-Mitteln kaum zu lösen ist: Der Leser wird nämlich eine weitere Stufe der Detaillierung (oder auch Ordnung) erwarten, die ihm (im Math-Net) etwa die Gliederung nach Kollektionen bietet, etwa nach Organisation, Software, Publikationen, etc. Während die “harte” Kodierung der MSC-Klassen etwa in Hyper-G als Kollektionsstruktur noch möglich ist, führt die weitere Untergliederung der untersten MSC-Klassen zu einer “Explosion” der insgesamt zu codierenden Klassen.

Da solch eine Strukturierung aus Sicht eines Lesers dennoch erforderlich ist, kann diese nicht mit den “hart” zu codierenden Hypertextstrukturen erzielt werden, sondern nur als virtuelle Sicht, wie sie durch den Einsatz eines spezifischen (z.B. eines objektorientierten) Datenbanksystems erzielt werden kann. An dieser Stelle sollte deshalb deutlich sein, daß es sich bei den “Links” der untersten MSC-Klassen nicht um “normale” Hypertext-Links handeln kann, sondern um Retrieval-Anfragen (an den Datenbankteil des Math-Net).

Es ist aber zu erwarten (vor allem, weil beim Math-Net zunächst die elektronischen Informationsangebote im Vordergrund stehen), daß viele der sehr fein unterteilenden Klassen

des MSC-Index nur leer oder schwach besetzt sein werden – vor allem in der Anlaufzeit des Math-Net. Auch hier ist aus der Sicht der Leser eine wie oben beschriebene Detaillierung nach Kollektionen hilfreich und sinnvoll – also auch schon auf den oberen Ebenen der MSC-Klassen.

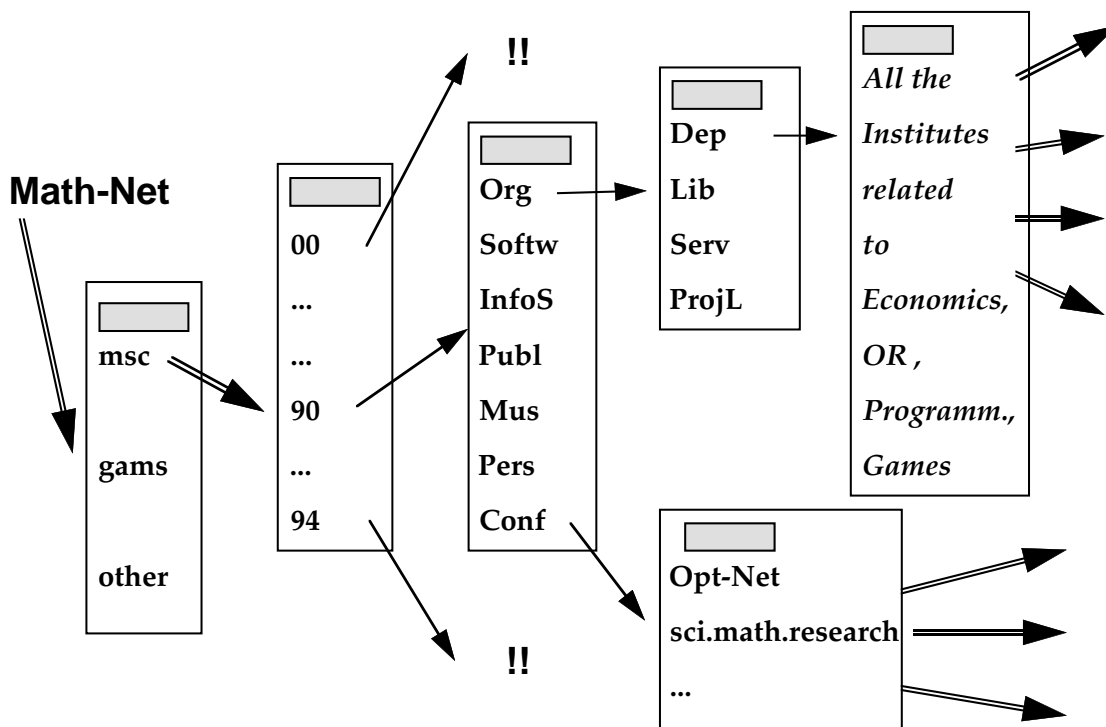


Abb. 5: Hierarchisches "Information Retrieval"

Leser des Math-Net sollen prinzipiell die Möglichkeit erhalten, auf allen Ebenen der Klassifikation auf eine Detaillierung nach Kollektionen umzuschalten. Im Beispiel von Abb. 5 wählt ein Leser etwa den MSC-Index 90 an und erhält eine nach den obersten Begriffen der Kollektionenstruktur des Math-Net geordnete "Liste" diesbezüglicher Elemente – mit einer Angabe der hier möglichen "Treffer". Der entsprechende Pfeil, der diese Auswahl kennzeichnet, ist wieder kein "normaler" Hypertext-Link, sondern er repräsentiert das Ergebnis einer Datenbankanfrage (die bei einem Maus-Click ausgeführt wird).

Wählt der Leser sich auf diese Weise über die organisatorische Struktur ("Org") bis zu den Einträgen über die Fachbereiche durch, so wird er dort eine Zusammenstellung aller Institute, Forschungsgruppen und Projekte finden, die am Verteilten Informationssystem teilnehmen. In dieser Menge von Items wird er wiederum suchen wollen, etwa nach Dekanen, Instituts- oder Projektleitern etc., oder auch nach fachlichen Inhalten (z.B. wenn Projektbeschreibungen vorliegen und diese volltextindiziert sind).

Hierarchisches Suchen und dynamisches Umschalten sind auf jeder Ebene der Detaillierung des Retrievals sinnvoll. Das stellt besondere Anforderungen an den Datenbankteil des Math-Net, die z.B. mit Hyper-G derzeit nicht zu erfüllen sind, weil Hyper-G eine feste Kodierung der Kollektionenstruktur voraussetzt. An eine Hartcodierung aller möglichen

Varianten des Umschaltens in Form einer vorgegebenen Hypertext- oder Kollektionsstruktur ist aber aus den o.g. Gründen nicht zu denken. Die Probleme, solch eine Hypertextstruktur zentral und dezentral zu koordinieren, sind in diesem Zusammenhang noch nicht einmal angesprochen.

Dieser Artikel beinhaltet den Vorschlag, die oben geschilderte Art des Retrievals zunächst modellhaft durch Kombination einer objektorientierten Datenbank mit den heute verfügbaren Mitteln des WorldWideWeb (also auf der Basis von HTML) zu entwickeln und zu erproben. Darüber hinaus soll mit den Entwicklern von Hyper-G auf der Grundlage der Erfahrungen, die mit einer entsprechenden Modellimplementierung gewonnen werden, diskutiert werden, ob Hyper-G nicht um derartige Funktionalitäten erweitert werden kann.

Für das Verteilte Informationssystem wäre aber schon der Einsatz von Hyper-G ein erster, wichtiger Schritt in die Richtung der Suche nach klassifizierten Items innerhalb Deutschlands. Hyper-G erlaubt ja bereits heute eine simultane Suche in mehreren Hyper-G-Servern, wobei aber z.Z. die betroffenen Server immer explizit, als Option bei der Suche, zu spezifizieren sind. Wenn es gelingt, für jede Stufe der Hierarchie der Kollektionsstruktur des Math-Net eine "Sammlung" von relevanten Servern bereitzustellen und innerhalb Deutschlands bzw. des "mathematischen Teils der Hyper-G-Welt" auch zu koordinieren, sollte eine Hyper-G-weite Suche nach MSC-klassifizierten Items auf jeder Ebene der Hierarchie der Math-Net Kollektionen möglich sein.

Eine Fokussierung allein auf Information, die nur innerhalb der Hyper-G-Welt zur Verfügung steht, würde aber den restlichen Teil der WorldWideWeb-Welt "abschneiden". Eine Suche nach Information in WWW- oder Gopher-Servern, oder gar in ftp-Archiven kann auf diese Weise nicht realisiert werden und es ginge die für das Verteilte Informationssystem angestrebte Offenheit bzgl. des Internet und insbesondere der WWW-Welt verloren. Damit steht die Entwicklung von generelleren, aber trotzdem einfachen Mechanismen und Verfahrensweisen an, die den Eintrag "externer" Information in das Math-Net in effizienter Weise erlauben. Dieses Thema behandelt das sechste Kapitel.

## 5.2 Hierarchische Kollektionen und MSC-Index im Nachrichtenteil

Hypertextsysteme verhalten sich Lesern gegenüber passiv. Ein Leser muß sie (aktiv) "besuchen", um herauszufinden, welche Informationen sie bieten. Schwierig wird es – insbesondere bei großen Systemen – für ihn immer dann, wenn er wissen will, welche Neuigkeiten in einer Informationsstruktur, die er bereits kennt, angeboten werden.

WWW-basierte Systeme bieten dem Leser hierfür keinerlei Unterstützung. Er müßte so geartete Informationsangebote vollständig durchsuchen, um auf dem Laufenden zu bleiben. Das kann aber nicht die Lösung sein. Manche WWW-Administratoren unterhalten deshalb eine Rubrik, die über Neuigkeiten informiert. Abgesehen davon, daß solche Rubriken bei großen Systemen selten vollständig sind, ändern sie sich nach dem vom Administrator vorgegebenen "Timing". Verpaßt ein Leser eine Änderungsankündigung, bleibt er nur unvollständig informiert.

Hyper-G-basierte Informationssysteme bieten die Möglichkeit, eine Suchanfrage durch Datumsangaben einzugrenzen. Im Prinzip kann ein Leser sich also von Zeit zu Zeit über alle neuen Einträge informieren. Das Ergebnis solcher Suchanfragen ist selbst aber wieder unstrukturiert. Hyper-G liefert es in Form einer linearen Liste aus. Bei umfangreichen Informationsangeboten (z.B., wenn es mehrere Fachbereiche umfaßt) führt dieses Verfahren leicht zum "Information Overload". Dem Leser bleibt natürlich – als Abhilfe – im Math-



Net die Möglichkeit, seine Suche hierarchisch zu spezialisieren, indem er sie auf gewisse Kollektionen oder durch Angabe spezifischer MSC- oder GAMS-Klassen eingrenzt. Dann entsteht aber (möglicherweise) das entgegengesetzte Problem.

Zu bestimmten mathematischen Teilgebieten (bzgl. des MSC-Index) werden nämlich relativ selten (und in unregelmäßigen Abständen) Neuigkeiten erscheinen, manchmal nur alle paar Wochen oder Monate. Leser, die sich für solche Teilgebiete interessieren, werden vermutlich deshalb zumeist ins Leere stoßen, weil sich in ihrem Teil des Informationsangebots nichts geändert hat. Möglicherweise werden sie bald aufgeben, wenn sie immer wieder nur auf dieselbe Information stoßen, die ihnen nichts Neues mehr bietet (“Information Underload”).

Um dieses Problem zu vermeiden, muß und soll das Math-Net eine “aktive” Nachrichtenkomponente erhalten, die den Leser (der sich nun “passiv” verhalten kann) – mittel elektronischer Post – informiert, sobald etwas für ihn Neues in das Math-Net eingebracht wird.

Für Hyper-G ist solch eine aktive Komponente bereits angedacht [KappeM 93a]. Das Einfügen eines neuen Dokuments in einen Hyper-G Server soll eine E-mail auslösen können, die über diesen Vorgang informiert. Nutzer von Hyper-G-Servern sollen ihren Wunsch, auf diese Weise informiert zu werden, mit bestimmten Kollektionen assoziieren können, also z.B. mit der Kollektion aller Informationsangebote eines Fachbereichs. Die o.g. Komponente kann aber aus solchen Botschaften noch nicht die nur jeweils interessanten herausfiltern, etwa diejenigen, die sich auf bestimmte MSC-Indizes beziehen. Leistungen dieser Art wären mit den Entwicklern in Graz abzusprechen.

Das Math-Net selbst kann dann – mit seiner Nachrichtenkomponente – als Empfänger solcher Nachrichten über neue Einträge eingerichtet werden, die innerhalb der Hyper-G-Welt bzw. innerhalb der mathematischen Institute (in Deutschland) in die örtlichen Informationsangebote eingebracht werden. Sobald es Nachrichten über neue Einträge erhält, wird es diese an alle verteilen können, die sich zu bestimmten MSC- oder GAMS-Klassen oder zu speziellen Kollektionen (oder entsprechenden Kombinationen davon) eingetragen haben.

Abb. 6 zeigt ein Referenzdokument, welches das Math-Net – via E-mail – über die Kollektionen “Publ” und “Mus” bzw. die MSC-Indizes “msc-00-02” und “msc-51-05” an die betreffenden Subskribenten verteilt. Dieses Bild ist so zu verstehen, daß das Math-Net nicht nur Leser, die sich für die ganz spezifischen MSC-Klassen interessieren, informiert, sondern auch alle, die sich zu “darüber” liegenden MSC-Klassen eingetragen haben, also auch diejenigen, die etwa “msc-51” oder “msc-00” subskribiert haben.

Leser, die sich etwa zu “msc” eingetragen haben, werden also alle Nachrichten zu Neuigkeiten erhalten, die überhaupt nach dem MSC-Index klassifiziert sind. Entsprechendes gilt für die Hierarchie der Kollektionen. Wer sich etwa an der “Spitze” der Math-Net-Kollektionen (also bei der Kollektion “math-net”) einträgt, den informiert das Math-Net über alle Beiträge, die es überhaupt erhält.

Die auf diese Weise verteilten Nachrichten müssen natürlich gewissen formalen Anforderungen genügen, die das nächste Kapitel diskutiert. Vorerst soll für diese Darstellung genügen, daß das Math-Net einer solchen Nachricht neben den entsprechenden Klassifikationen und Angaben zur Einordnung in die Hierarchie der Kollektionen auch den Hinweis auf den Standort des Originaldokuments (URL) sowie eine kurze Beschreibung des Inhalts des Originals entnehmen können muss. Falls das Original z.B. ein Report ist, könnte das Referenzdokument auch die “erste Seite” (Titel, Autor, Abstract etc. – auch in  $\LaTeX$ -Form) enthalten, die um solche Angaben ergänzt ist.

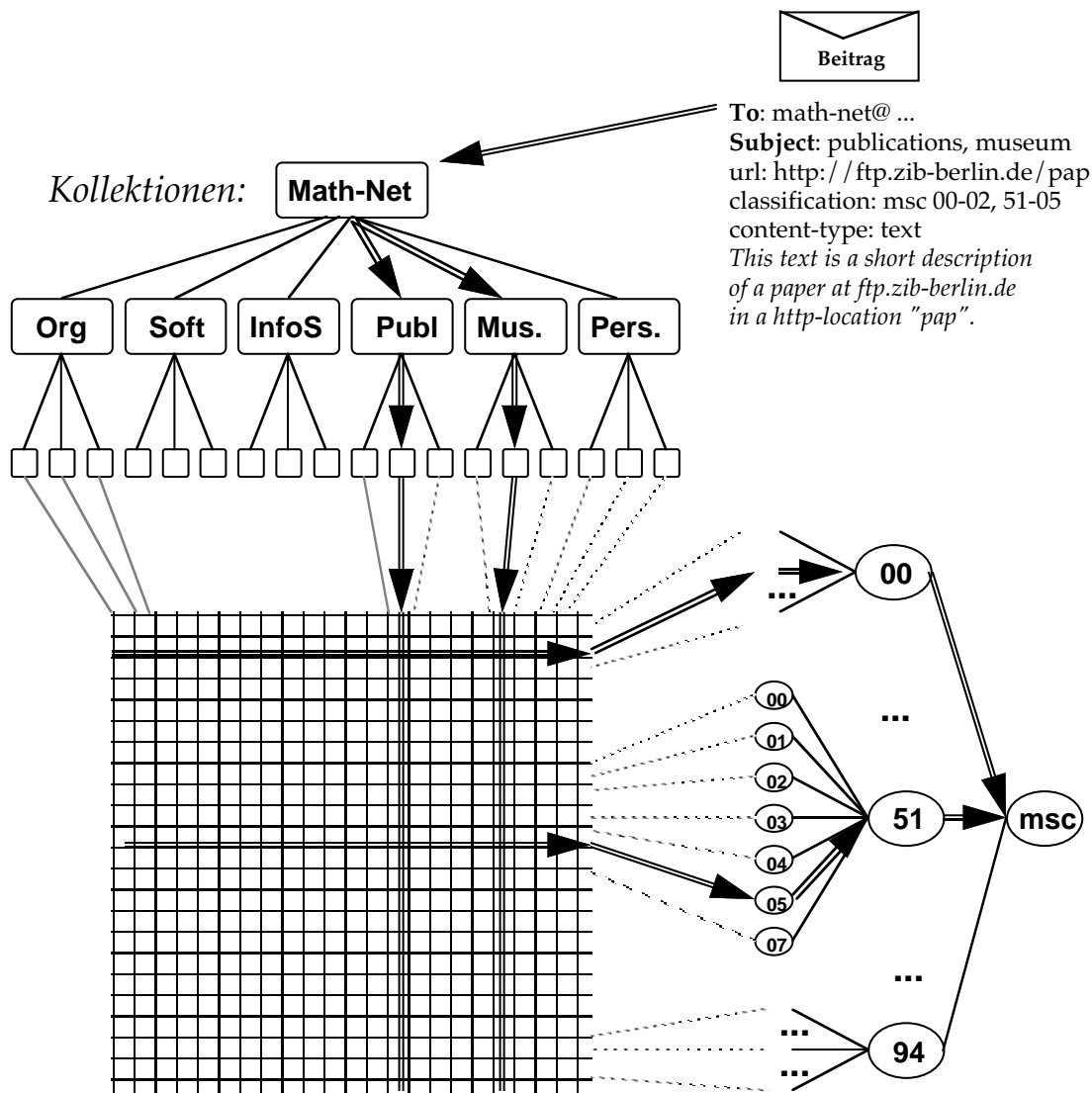


Abb. 6: E-mail-basiertes Nachrichtensystem

Auch die “Sprache”, mit der sich Leser subscribieren, ist noch zu entwickeln. Sie soll es zulassen, alle möglichen Kombinationen von Klassifikationsschemata und Hierarchien von Kollektionen zu spezifizieren. Die Datenbank-Komponente des Math-Net wird nicht nur die Adressenangaben und Subskriptionswünsche der Leser speichern, sondern auch dafür sorgen, daß diesen die angeforderten Ankündigungen korrekt (und genau einmal) zugestellt werden.

Das hier bezüglich Hyper-G beschriebene Verfahren läßt sich – mit gewissen Einschränkungen – auch global in die WWW-Welt übertragen, da es technisch nur auf (strukturierter) elektronischer Post als Mittel der Übertragung beruht. Dazu müssen mathematische Institutionen, die ihre Informationen außerhalb der Hyper-G- Welt anbieten wollen (also etwa mittels ftp, WWW- oder Gopher-Servern), ihre wichtigsten Dokumente per strukturierter E-mail über das Math-Net verteilen bzw. im Math- Net registrieren.

Ein Institut, das seine Preprints nur in einem ftp-Bereich anbietet, erhält damit – via Math-Net – die Möglichkeit, entsprechende Abstracts (mit Hinweisen auf den Standort) zielgenau an Interessenten zu versenden. Das Math-Net erfüllt in diesem Falle also die Funktion eines Preprint-Servers vom Ginsparg-Typ [Ginsparg 94], wobei das Math-Net zusätzlich die Information für Leser nach fachlichen Gesichtspunkten filtert. Mail-Server vom Ginsparg-Typ archivieren Originaldokumente zentral. Dieses ist im Rahmen des Math-Net-Konzepts nicht erforderlich (und aus verschiedenen Gründen der Sache auch nicht angemessen), weil die Partner des Projekts ihre Dokumente bereits elektronisch vorhalten. Die Frage, wie ein Leser sich über alle via Math-Net erreichbaren Dokumente informiert, ist ein anderes Thema, das im nächsten Kapitel vertieft behandelt wird.

Mittels Mirror-Verfahren lassen sich neue Einträge in einem ftp-Bereich heute automatisch herausfiltern. Wenn ein Institut nun auch seine Referenzdokumente im ftp-Bereich ablegt, die den strukturellen Anforderungen bzgl. des Math-Net genügen, kann der Vorgang der expliziten Ankündigung via E-mail an das Math-Net (mittelfristig) automatisiert werden. Mathematische Institutionen brauchen – bei Einsatz solcher Verfahren – dann nur noch (ein einziges Mal) ihren ftp-Bereich beim Math-Net zu registrieren, damit Leser des Math-Net über alle Änderungen in diesem Bereich informiert werden.

Sofern ein solcher ftp-Bereich “harmonisch” auf die Struktur des Math-Net abgestimmt ist, d.h. wenn die ftp-Pfade den “Pfadern” der Hierarchie der Kollektionen entsprechen, sollte sich auch die Zuordnung zu den Kollektionen des Math-Net automatisieren lassen. Klassifikationen, etwa nach dem MSC-Index, lassen sich hingegen kaum automatisch generieren. Sie müssen dem Referenzdokument i.d.R. “per Hand” mitgegeben werden.

Gopher-Server sollten sich in ähnlicher Weise in das Math-Net integrieren lassen, denn sie sind ebenfalls hierarchisch strukturiert. Zu prüfen bleibt allerdings die Frage, ob es in der Gopher-Welt Mirror-Verfahren gibt, die für einen Einsatz im Zusammenhang mit dem Math-Net geeignet sind.

Mit den heute verwendeten WWW-Servern verhält es sich jedoch anders, weil sie i.d.R. “beliebige” Link-Strukturen verwenden, die sich nicht auf die hierarchischen Strukturen des Math-Net abbilden lassen. Hier bleibt – aus heutiger Sicht – wohl nur die Möglichkeit, daß die Administratoren und Autoren solcher Informationsstrukturen ihre wichtigsten, für das Math-Net relevanten Seiten (die Top-Pages gemäß dem zweiten Kapitel) mittels von Hand erzeugter E-mail über das Math-Net ankündigen.

Sofern solche Server ausschließlich mathematische Informationen enthalten und sich in ihrer Struktur auf das Math-Net abbilden lassen, wäre (in Zukunft) zu überlegen, ob sich in diesem Zusammenhang nicht auch gewisse WWW-Robots zur Automatisierung von Ankündigungen einsetzen lassen.

## **6 Architektur des Verteilten Informationssystems**

Das Verteilte Informationssystem besteht – technisch – aus den bei den Partnern des Projekts örtliche vorliegenden Informationssystemen und einem zentralisierten Anteil, der die bei den Partnern “verteilten” Informationsangebote erschließt. Die folgenden Kapitel fassen die Ergebnisse der vorangegangenen Diskussionen zu einem technischen Gesamtbild zusammen.

## 6.1 Zentralisierter Teil der Math-Net-Architektur

Aus Sicht der Leser mathematischer Informationen sind – wie oben diskutiert – beide Betriebsmodi für das Verteilte Informationssystem sinnvoll und notwendig. Das Math-Net wird sie in folgender Form enthalten; einer

1. aktiven Informationskomponente, einem Nachrichtensystem, das insbesondere
  - \* Neuankündigungen via E-mail distribuiert und dabei
  - \* nach Kollektionen und Klassifikationen filtert;
2. passiven Informationskomponente, einem System, das
  - 2.1 die assoziative Suche (“Browsen”) in Hypertexten und
  - 2.2 die hierarchische Suche (“Query”) in hierarchischen Kollektionen gestattet, wobei die Suche nach
    - \* Schlüsselworten (in Dokumententiteln und -texten)
    - \* Klassifikationen bzw. Indizes
    - \* Datumsangaben etc. möglich sein soll.

Durch die Klassifikation der einzelnen Informationseinheiten und die hierarchische Strukturierung der gesamten (örtlichen) Information sollen Leser des Math-Net sowohl dem Problem des “Information Overload” als auch dem des “Information Underload” begegnen können.

Abb. 7 faßt die bereits diskutierten technischen Komponenten in einer knappen Graphik zusammen. Strukturierte Nachrichten (Referenzdokumente), die – “per Hand” oder automatisch – an das Math-Net geschickt werden, verteilt dieses einerseits an die betreffenden Interessenten. Es integriert sie andererseits aber auch derart in (s)einer Datenbank, daß Leser die Menge der Referenzdokumente (oder Teilhierarchien davon) sowohl mit Hypertextmitteln (vertikal und horizontal) “browsen” können, als auch mit Datenbankmitteln (hierarchisch) befragen können.

Das Ergebnis einer Datenbankanfrage ist wieder ein Hypertextdokument, hinter dem sich eine (virtuelle) hierarchische Hypertextstruktur verbirgt, die nun die entsprechend durch die Datenbankanfrage gefilterte eingeschränkte Hierarchie von Kollektionen enthält, und die horizontal – also quer durch Deutschland – auf alle so spezifizierten Informationseinheiten verweist.

Ein technisches Fernziel ist, daß Leser in allen im Math-Net zentralisierten Hypertextdokumenten zwischen den verschiedenen Hierarchien, der Hierarchie der Kollektionen und den durch die Klassifikationsstrukturen definierten Hierarchien, nach Bedarf (beliebig) hin- und herschalten können. Die Strategie für die Realisierung der mit dieser Vorstellung verbundenen “virtuellen Hierarchien” von Kollektionen muß allerdings einer anderen Arbeit vorbehalten bleiben.

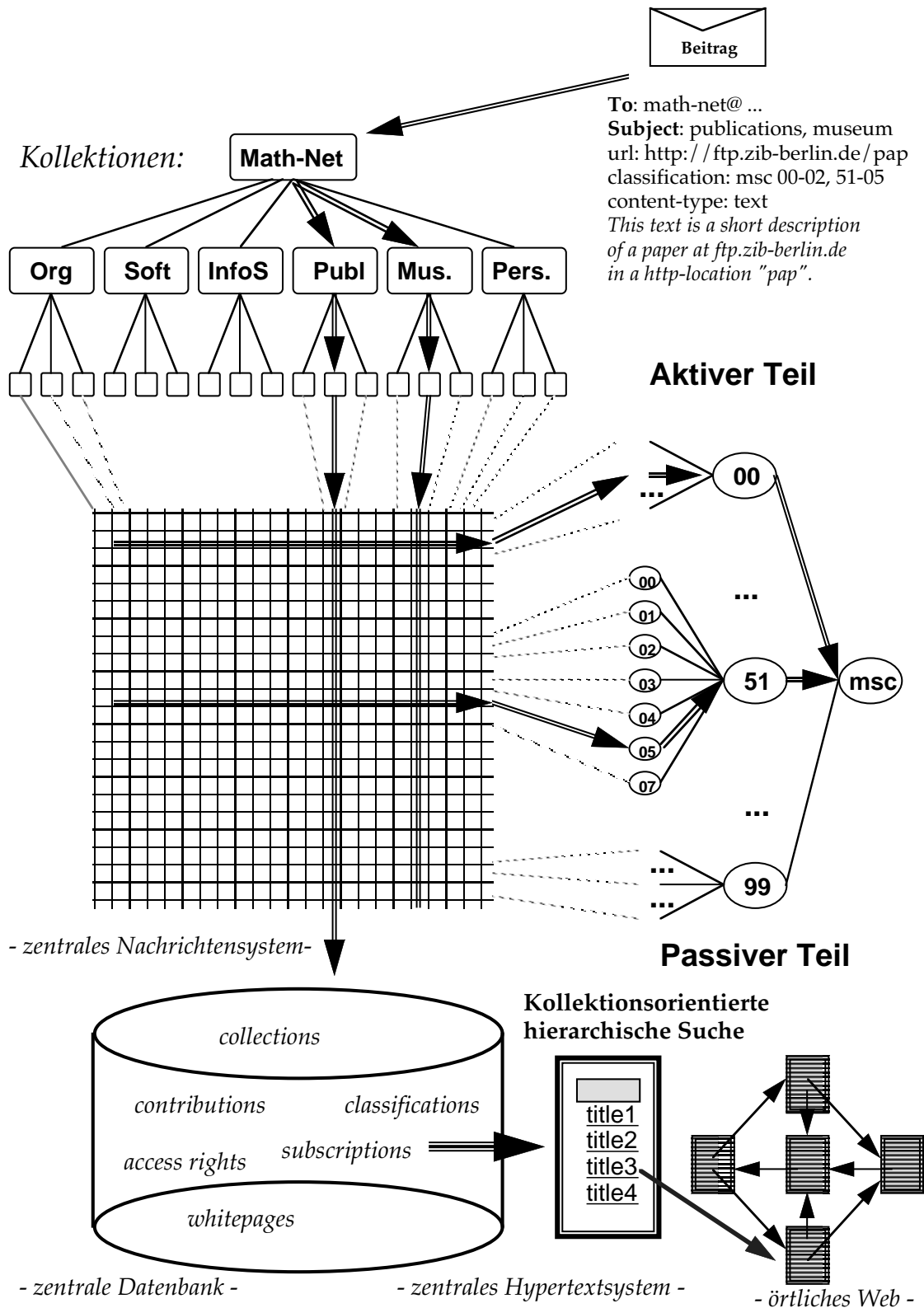


Abb. 7: Zentrale Math-Net-Architektur

## 6.2 Integration von Information in das Math-Net via E-mail

Innerhalb der “mathematischen Hyper-G-Welt” kann neue Information automatisch in das Math-Net eingebracht werden, denn, sofern alle mathematischen Hyper-G-Server gemäß derselben Struktur aufgebaut sind, kann diese Struktur im Math-Net derart reflektiert werden, daß neue dezentrale Einträge (und Löschungen) gleichzeitig im zentralisierten Teil (des Math-Net) erscheinen. Die eigentliche Arbeit der Koordinierung der Information übernimmt in diesem Falle das im Internet verteilte System von Hyper-G-Servern. Die dafür notwendige, “fachliche” Struktur der mathematischen Information ist Gegenstand eines anderen Papiers, das gegenwärtig mit den mathematischen Fachbereichen abgestimmt wird.

Dieses Kapitel skizziert einen Weg (und einen ersten Vorschlag), wie neue Information von “außerhalb” (der homogenen Hyper-G-Welt) in das Math-Net eingebracht werden kann, sofern dieses auf der Basis von Hyper-G implementiert ist. Der Grundgedanke ist, daß solche Informationen mittels strukturierter, elektronischer Post, die alle für eine Einordnung notwendigen Angaben enthält, an den bereits erläuterten Nachrichtenteil des Math-Net zu schicken sind (das sie auf diese Weise auch “registriert” und speichert). Fragen der Automatisierung sollen in diesem Zusammenhang zunächst ausgeklammert sein. Der Vorschlag ist, daß “externe” Administratoren (von WWW-Servern) ein Referenzdokument (zunächst “per Hand”) an das Math-Net schicken, das sie um die entsprechenden Angaben ergänzt haben.

Die für eine Einordnung im Math-Net absolut notwendigen (minimalen) Angaben bzw. Kategorien sind:

- \* collection: Name bzw. “Pfad” der Informationskollektion  
(Angabe der “Ablage” im hierarchischen Hypertext)
- \* title: Titelzeile des Beitrags, so wie er im Math-Net erscheint  
(nicht notwendig identisch mit der Titelzeile vor Ort)
- \* class: MSC-Index, GAMS-Index, Keyword  
(auch mehrere Angaben zulässig)
- \* url: Uniform Resource Locator  
(elektronischer Standort im Sinne des WorldWideWeb)
- \* owner: Kurzbezeichnung des Einsenders (identisch mit seinem  
“nickname”, sofern er im Math-Net registriert ist)  
oder E-mail-Adresse des Einsenders
- \* date: Datum, zu dem der Beitrag vorgelegt  
oder auch aktualisiert wurde
- \* content-type: “Datentyp” des Referenzdokuments

Der eigentliche Inhalt, das Referenzdokument, wird i.d.R. ein ASCII-Text sein. Dieser kann aber selbst wieder strukturiert sein, je nach der Art der Kollektion, in die er einzuordnen ist. Verweist das Referenzdokument auf bibliographische Information, so kann er z.B. im BIBTEX -Format oder im MATH-Format (das in der Datenbank des FIZ Karlsruhe verwendet wird) oder im MR-Format (der Datenbank MathSci der “Mathematical Reviews”) geschrieben sein. Handelt es sich um einen Abstract, so kann er in TEX oder LATEX vorliegen. Verweist das Referenzdokument auf die HomePage eines Fachbereichs oder auf die Top-Page einer mathematischen Information “in” einem Hypertext-Web, so liegt er vielleicht als HTML- oder HTF-Text vor. Verweist es auf Software, so ist das

Referenzdokument vielleicht identisch mit dem README- oder Index-File (einem reinen ASCII-Text).

Welche spezifische Form das Referenzdokument hat, ist für die Einordnung in das Math-Net nicht wesentlich, solange es nur für eine (Volltext-) Indizierung herangezogen (und in einem “Viewer” gezeigt) werden kann. Es wäre also insbesondere auch möglich, Postscript-Files (im Attachment der E-Mail) zuzulassen. Die Angabe zum “content-type” (durch den Einsender) hilft dem Math-Net bei der richtigen Interpretation.

Ein externer Einsender von Referenzdokumenten an das Math-Net soll es möglichst leicht haben, neue Informationen “per Hand” einzutragen. Stellen wir uns vor, daß er via Mosaic auf ein Web-Dokument stößt, das für das Verteilte Informationssystem von Interesse ist. Er sollte dann einfach mit Hilfe des “Mail-To”-Features aus Mosaic heraus seinen Beitrag an das Math-Net schicken können. Angenommen, daß er die Home-Page der “Anima Computer Gallery” gefunden hat, die für die (jetzt noch fiktive) Kollektion “Art” des “Mathematischen Museums” von Interesse ist. Dann soll es genügen, wenn der Einsender seinen Beitrag (im Mail-Fenster von Mosaic) in folgender Form spezifiziert:

```
To:      math-net@...
Subject: math.mus.art // Anima Computer Gallery // msc00A99
```

Diese E-mail wird im “Body” dann die Home-Page in Text- oder in HTML-Form enthalten, je nachdem was er in seinem Mosaic-Client eingestellt hat (auch Postscript wäre möglich). Die Angaben zu “url”, “date” und “content-type” ergänzt Mosaic automatisch im “Header” der Mail. Die Angabe zu “owner” kann das Math-Net aus der Reply-Adresse evaluieren, sofern der Einsender, der vielleicht ein Informationsbeauftragter ist, sich mit seiner E-mail-Adresse (einmal beim Math-Net) bekannt gemacht hat. Die Frage der Identifizierung solcher anonymer Einsender soll an dieser Stelle einmal außer Acht gelassen werden. Alle übrigen notwendigen Angaben stehen in der Subject-Zeile der Mail.

### 6.3 Distribution des zentralisierten Teils des Math-Net

Das Verteilte Informationssystem ist von vorneherein als verteiltes System konzipiert. Das bedeutet insbesondere, daß sein zentralisierter Anteil, das Math-Net, verteilt zu denken ist. D.h., die im Math-Net zentralisierte Information soll nicht nur auf einem zentralen Server zur Verfügung stehen, sondern, bei jedem der Partner des Projekts, direkt zugreifbar sein (bzw. “aufliegen”). Der hauptsächliche Vorteil einer solchen Organisationsform liegt darin, daß auf diese Weise Engpässe vermieden werden, die sonst bei zentral organisierten Datenbanken über kurz oder lang zu erwarten sind.

Innerhalb der “mathematischen Hyper-G-Welt” läßt sich die Verteilung des Math-Net mit Standardwerkzeugen organisieren. Bei voller Öffnung der Math-Net-Konzepts bzgl. der WWW-Welt sind aber andere Konzepte notwendig. Ein erster Vorschlag zu diesem Thema ist, die Verteilung der (neuen) Beiträge aus der (Hyper-G-externen) WWW-Welt ebenfalls auf der Basis elektronischer Post mittels der Nachrichtenkomponente des Math-Net zu implementieren.

Abb. 8 zeigt, wie ein Math-Net Server “A”, der innerhalb der Hyper-G-Welt liegt, neue Ankündigungen an einen Server “B” verschickt, der außerhalb der Hyper-G-Welt eingerichtet ist. Der Server “B” ist dabei als Subscribent der Kollektion “math-net” eingerichtet. Er empfängt also von “A” alle Nachrichten, die überhaupt über das Math-Net verteilt werden, und integriert diese auch in seine Datenbankkomponente (wobei er gleichzeitig

alle Interessenten informiert, die sich vor Ort subscribiert haben). Die Frage, wie diese Datenbank zu gestalten ist, muß an dieser Stelle offen bleiben (sie ist der Gegenstand einer anderen Arbeit).

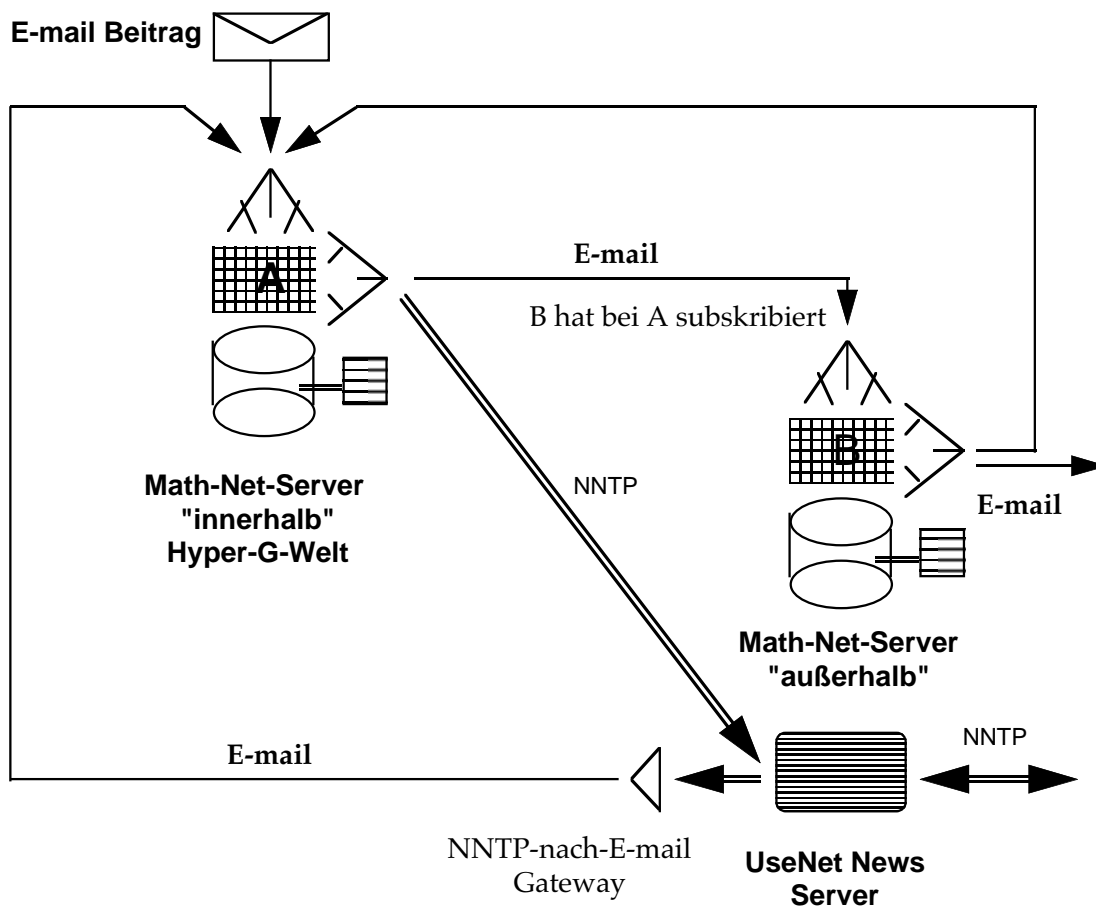


Abb. 8: Distribution des zentralisierten Teils

Mittels der Mail-Schnittstelle des Math-Net könnten auch die UseNet News (bzw. spezielle Diskussionsgruppen der News) mit mathematischer Information versorgt werden, wobei noch zu klären ist, ob hierfür nicht sogar das übliche News-Protokoll (NNTP) zum Einsatz kommen kann. Auch der umgekehrte Weg ist technisch – prinzipiell – denkbar, daß Nachrichten aus den News über einen E-mail-Gateway in das Math-Net eingebracht werden können.

Ob solch ein Verfahren auch inhaltlich richtig ist, wäre noch zu überlegen. Die News arbeiten i.d.R. unmoderiert. Anonyme Beiträge sind nicht nur möglich, sondern sogar die Regel. Das Verteilte Informationssystem soll hingegen von Personen beschriftet, gesteuert und moderiert werden, deren organisatorische Einbettung bekannt ist und die sich inhaltlich verantwortlich fühlen.

Damit ist ein weiteres Thema angeschnitten, das sich innerhalb der Hyper-G-Welt besser regeln läßt als in einer fast vollständig offenen Internet-Welt: die Bewertung und Kontrolle der Information, die in das Verteilte Informationssystem integriert wird. Hyper-G erlaubt



die dedizierte Vergabe von Schreibrechten und beinhaltet ein abgestuftes Konzept zur Definition verschiedener Ebenen von Zugriffsrechten (mit entsprechenden Passwort- bzw. login-Mechanismen).

Informationsbeauftragte, die ihren Server mit Hyper-G steuern, können von Fall zu Fall entscheiden (und kontrollieren), welche Personen sie in den Prozeß der Bereitstellung von Information einbeziehen (und wie). Mathematische Institutionen können so nicht nur entscheiden, welche Information sie in das Verteilte Informationssystem einbringen, sondern sie erhalten auch die technischen Mittel zur abgestuften Steuerung dieses Vorgangs.

Für die Kontrolle von Information, die “von außen” in das Math-Net eingebracht wird, fehlen derzeit noch entsprechende Konzepte. Ein erster Gedanke hierzu ist, beliebige, externe Beiträge nicht automatisch zu integrieren, sondern diesen Prozeß durch Moderatoren zu steuern, die z.B. die Inhalte gewisser Kollektionen verantwortlich betreuen.

Ungeklärt ist zum heutigen Zeitpunkt auch noch die Frage, wie sich das – mit Hyper-G-Mitteln heute schon weitgehend realisierbare – Math-Net-Konzept bzgl. der WWW-Welt vollständig generalisieren läßt. Dazu sind, mit dem bereits oben erwähnten Strukturkonzept, erste Vorstellungen erarbeitet worden, deren Realisierung allerdings noch zu klären ist. Die eigentlichen Schwierigkeiten bei der Errichtung des Verteilten Informationssystems werden aber weniger technischer Art, sondern vielmehr organisatorischer Natur sein.

Werden die mathematischen Institutionen – gleich, ob innerhalb oder ob außerhalb der Hyper-G-Welt – bereit sein, die bei ihnen elektronisch vorhandenen Informationen in ihre örtlichen Hypertextsysteme einzubringen? Werden sie auch Willens sein, die für die Leser ihrer Hypertexte notwendige Strukturinformation in das Verteilte Informationssystem einzubringen? Werden Sie dabei auch dem Vorschlag folgen, ihre einzelnen Dokumente entsprechend zu klassifizieren? Und, werden sich auch Mathematiker finden, die in der Lage sind, den Prozeß des Einbringens und Verteilens von zentralisierter Information entsprechend zu bewerten und zu moderieren?

Das ist das eigentliche Zentrum des Verteilten Informationssystems: Es ist der Kreis der Personen, die – auf freiwilliger Basis – bereit sind, es aufzubauen und langfristig zu tragen.

## 7 Zusammenfassung, Leitlinien und Designprinzipien

Das organisatorische und technische Design des Verteilten Informationssystems orientiert sich an folgenden Leitlinien:

**Aktualität und Vollständigkeit:** Mathematische Ergebnisse sollen zum Zeitpunkt ihrer Verfügbarkeit sofort elektronisch angeboten werden; dazu gehören nicht nur Preprints und Vorveröffentlichungen von Büchern (Volltexte), sondern auch mathematische Software und Datensammlungen sowie aktuelle Nachrichten in Form von elektronischen Journalen. Auch historische Dokumente und Bücher, sowie Darstellungen der mathematischen Moderne (“lebendige” Experimente und Visualisierung, Computerkunst) sollen in das Angebot eingehen. Kontaktadressen, Institutsprofile, Informationen zur Lehre und Forschung, Statusberichte über laufende Projekte etc. sollen das Informationsangebot abrunden.

**Organisation als Querschnittsaktivität:** Das Projekt soll nicht nur mathematische Fachbereiche und Forschungsinstitute mit einbeziehen, sondern auch mathematische Forschungslabors aus der Industrie und die Informationsversorger der Mathematik, insbesondere die wissenschaftlichen Verlage, die Technische Informationsbibliothek Hannover, die Bibliotheken der Universitäten und das Fachinformationszentrum Karlsruhe mit dem Zentralblatt für Mathematik.

**Dezentrale Organisation:** Das bisherige zentrale Angebot soll durch das reichhaltige dezentrale Informationsangebot ergänzt werden (mittels an das Internet angeschlossener Server). Die Partner des Vorhabens entscheiden selbst über ihr eigenes Informationsangebot und ordnen es nach gemeinsamen Regeln und Standards.

**Offenheit auf der technischen Grundlage des Internet:** Die Partner des Projekts sollen nicht nur untereinander Information austauschen können, sondern auch mit ihren Fachkollegen aus anderen Ländern und ihren Kooperationspartnern und Anwendern aus anderen Fachrichtungen. Der Zugriff auf Informationssysteme in anderen Ländern oder in anderen Fachdisziplinen soll gleichermaßen möglich sein.

**Einfache, universelle Benutzbarkeit, Flexibilität:** Für den Zugriff auf die angebotene Information kommen nur die einfach zu benutzenden Werkzeuge zum Einsatz (Hypertext-Systeme wie Mosaic, World Wide Web oder Hyper-G) und Systeme, die im Internet weit verbreitet und anerkannt sind (E-mail, ftp, Gopher etc.). Langwieriges Erlernen von Kommandosprachen kommt nicht in Betracht. Auf diese Werkzeuge ist das Projekt aber nicht festgelegt. Der Übergang auf neuere und bessere Hilfsmittel soll jederzeit möglich sein.

**Freizügigkeit:** Der Zugriff auf dezentrale Information soll – einschließlich aller Volltexte – auf der Basis des gegenseitigen Austausches (kostenfrei) möglich sein. Für den Zugriff auf zentrale (kommerzielle) Angebote (insbes. Datenbanken) werden faire Preise angestrebt.

**Effektivität:** Die Kommunikation der Partner untereinander und mit allen an der Mitnutzung interessierten Personen wird mit elektronischen Mitteln organisiert. Dabei kommen insbesondere auch E-mail, elektronische Nachrichtensysteme und fachgebietsspezifische Informationsnetze zum Einsatz.

## 8 Ursprung und Dank

Die Ausgangsidee für das technische Konzept geht auf einen Leitgedanken von A. Odlyzko zurück, den er in seiner für den Bereich des elektronischen Publizierens in der Mathematik grundlegenden Arbeit “Tragic loss or good riddance? The impending demise of traditional scholarly journals” [Odlyzko 94] ausgesprochen hatte. Hier war zum ersten Mal von einem universellen Speicher die Rede, der die gesamte mathematische Literatur umfaßt. Dieser Speicher sollte mehrstufig, hierarchisch gegliedert sein und in gewisser Weise eine bewertende Struktur (refereeing structure) elektronisch reflektieren. Allerdings war dieser Speicher in den ersten Versionen seiner elektronisch (weithin) diskutierten Arbeit noch nicht als in einem Netz (Internet) global verteilt gedacht und sollte auch nur klassische mathematische Publikationen enthalten.

Die Verbreitung des Internet (und seine Rolle) in der wissenschaftlichen Welt und die weltweite Akzeptanz der einfach zu benutzenden Internet-Tools, wie Gopher (Ende 1993) und – vor allem – Mosaic (in 1994) im World Wide Web (sowie die Fülle der schon seit langem im Internet vorhandenen ftp-Archive für Software und auch Preprints in diversen Wissenschaftsbereichen) führten auf natürliche Weise zu dem Gedanken eines über die mathematische Welt verteilten, universellen mathematischen Archivs, das auch Software, Datensammlungen, organisatorisch bedeutsame Information etc. enthält.

Das Konrad-Zuse-Zentrum für Informationstechnik hatte schon seit längerer Zeit – auf eine Initiative von P. Deuffhard (ZIB) hin und in Kooperation mit E. Grosse von AT&T und A.C. Hearn (RAND Corp.) – Erfahrungen mit einer elektronischen Bibliothek eLib

für mathematische Software gewonnen, die mit der netlib von AT&T und mit der Reduce Network Library der RAND Corp. "vernetzt" war. Die eLib wurde überwiegend von Herrn Dalitz (ZIB) aufgebaut und bot (und bietet auch heute) eine Fülle von Algorithmen (Software, Daten und Dokumente) aus den Bereichen des "Scientific Computing" und der Computeralgebra. Spezifische Erfahrungen in Deutschland führten uns jedoch zu der Einsicht, daß der in den genannten elektronischen Bibliotheken realisierte zentrale Speicher an prinzipielle Grenzen stößt und besser "verteilt" zu organisieren ist.

Auf Initiative von M. Grötschel (ZIB) und in Zusammenarbeit mit U. Zimmermann und seinen Mitarbeitern (Univ. Braunschweig) haben wir (Herr Dalitz und der Autor) das elektronische Diskussionsforum Opt-Net für die Fachgruppe Optimierung der DMV implementiert. Ein wesentlicher Grundgedanke, den wir U. Zimmermann verdanken, war die Kombination dieses "Mailverteilers" (der dem NA-Net von J.J. Dongarra und B. Rosener von der Univ. of Tennessee in vielem nachempfunden war) mit einem hierarchisch gegliederten Speicher. Gedacht war dabei ursprünglich an die Anbindung einer Datenbank, ein technisches Ziel, das wir aus diversen Gründen aber leider noch nicht für das Opt-Net erreichen konnten.

Letztlich zeigte aber die Akzeptanz des Opt-Net, daß elektronische Diskussionforen ein wesentlicher Bestandteil eines verteilten, mathematischen Informationssystems sind, das über aktuelle Ereignisse informiert. Sie sollten aber (aus qualitativen Gründen) fachlich "gefiltert" arbeiten, d.h. wie das Opt-Net durch einen Moderator betreut werden. Auch der Betrieb eines technisch sehr viel einfacheren Mailverteilers im DMV-Projekt "Fachinformation", den W. Sperber (ZIB) moderiert, bestätigte diese Einsicht.

Anfang 1994 begannen H.-H. Frese, P. Müller und M. Dörrschuck im Bereich Rechenzentrum des ZIB mit dem Aufbau eines WWW-Servers, dem ZIB- Informationssystem ZIBIS. In einer Arbeitsgruppe zusammen mit W. Dalitz und W. Sperber und dem Autor unternahmen wir gemeinsam den Versuch, eine homogene Informationsstruktur für das ZIBIS zu finden, die mit den Zielsetzungen des Verteilten Informationssystems verträglich war. Diese Anstrengungen waren jedoch damals leider noch zum Scheitern verurteilt. Unsere Konzepte waren nur ansatzweise angedacht und es wurde auch klar, daß die technischen Mittel zur Realisierung einer homogenen, absuchbaren und über die Mathematik in Deutschland "verteilten" Informationsstruktur uns noch fehlten.

Mit den uns damals bekannten Internet-Tools des World Wide Web (Mosaic und WWW-Server, etwa in Kombination mit WAIS) waren unsere Ziele technisch nicht in zuverlässiger Weise realisierbar. Dennoch waren gerade die hier geführten Gespräche recht wertvoll, weil sie uns die Problematik der gegebenen technischen Mittel deutlich vor Augen führten, insbes. Fragen der Zuverlässigkeit von Links in Hypertexten, der Änderbarkeit bestehender Informationsstrukturen, der Behandlung umfangreicher Dokumentenmengen, der hierarchischen Gliederbarkeit und Suchbarkeit von heterogenen Daten sowie nicht zuletzt dem Retrieval in über mehrere Standorte verteilten Datenbeständen.

Nach mehreren Gesprächen mit H. W. Kröger, O. Ninnemann, R. Welk und C.-D. Ziegler von der Redaktion des Zentralblatt sowie W. Dalitz, W. Sperber und dem Autor vom ZIB (in denen über elektronische Mittel zur "Modernisierung" des Zentralblatt nachgedacht wurde) kam der Gedanke auf, auch hier einen hierarchisch strukturierten Mailverteiler einzusetzen, der etwa nach dem MSC-Index gegliedert ist. Wir dachten zunächst daran, eine Modifikation des bekannten UseNet einzusetzen, um den Prozeß der Referierens zu unterstützen. Eine homogene Einordnung dieses Konzepts war zum damaligen Zeitpunkt jedoch nicht absehbar. Dennoch waren auch diese Gespräche für uns von Bedeutung, und

den Beteiligten sei an dieser Stelle ausdrücklich gedankt.

Auf Grund einer Kooperation mit dem Arbeitskreis Fachinformation der GI erfuhren wir schließlich von Hyper-G, einem System, das ursprünglich F. Kappe und H. Maurer von der TU-Graz konzipiert hatten. Das System befand sich damals (Mai 1994) noch in der Entwicklung, jedoch konnte F. Kappe an einem ersten Prototyp (im ZIB) bereits wesentliche strukturelle Eigenschaften demonstrieren. Viele der Probleme, denen wir zuvor begegnet waren, fanden in Hyper-G zumindest theoretisch eine Lösung, Grund genug, unsere ersten Schritte in Richtung auf ein Verteiltes Informationssystem auf der Basis von Hyper-G zu wagen.

Die ersten Design-Überlegungen wurden dann den Fachinformationsbeauftragten der Mathematik bei ihrem Workshop in Halle (Juni 94) vorgestellt. In einer elektronischen Diskussion auf dem Mailverteiler des DMV-Projekts wurde anschließend der Versuch unternommen, die Grundlagen für die Gestaltung von elektronischen Informationsangeboten zu klären. Inzwischen haben W. Dalitz und G. Heyer (ZIB) mehrere Tutorials zu Hyper-G durchgeführt, um die Fachinformationsbeauftragten mit den technischen Gegebenheiten von Hyper-G vertraut zu machen und um ihnen eine Bewertung von Hyper-G zu ermöglichen. Sie haben in diesem Zusammenhang auch die Strukturüberlegungen für das Verteilte Informationssystem (weiter-) entwickelt und mit den Teilnehmern der Tutorials diskutiert. W. Dalitz bereitet derzeit (federführend) einen entsprechenden Vorschlag aus, zu dem G. Heyer auf dem Hyper-G-Server des ZIB auch eine Probeimplementierung erstellt hat. Die Übertragung der Strukturüberlegungen auf die mathematischen Fachbereiche und Forschungsinstitute bedarf jedoch noch der eingehenden Diskussion mit den Fachinformationsbeauftragten.

In dem hier vorliegenden Artikel habe ich den Versuch gemacht, die Ideen, Anregungen und Kommentare aus vielen Diskussionen (und Quellen) zu integrieren. Leider ist es mir nicht möglich, alle Personen mit ihrem Beitrag namentlich in Verbindung zu bringen. An vorrangiger Stelle möchte ich W. Dalitz und W. Sperber für die vielen Ideen, Anregungen und auch kritischen Diskussionen danken, die wir im Rahmen der Vorbereitung des Verteilten Informationssystem führen konnten. Ohne die ständige Ermutigung durch M. Grötschel und seine vielfältigen Anregungen wäre diese Arbeit sicher nicht möglich gewesen. Das Manuskript habe ich (erst) in der zweiten Jahreshälfte 1994 niederschreiben können. In diesem Zusammenhang möchte ich G. Heyer (ZIB), der erst im August 1994 zu uns gestoßen ist, für seine außerordentlich gründliche redaktionelle Begleitung und Korrektur dieser Arbeit danken. Auch Frau S. Mattrisch (ZIB) gebührt ein besonderer Dank für den sorgfältigen Satz des Artikels in TeX.

## 9 Literatur und elektronische Referenzen

[AndreessenB 94] M. ANDREESSEN, E. BINA; NCSA Mosaic: A Global Hypermedia System; Internet Research, Vol. 4 No. 1, Mecklermedia 1994, pp. 7-17

[BernersCGP 92] T. BERNERS, R. CAILLIAU, J.-F. GROFF, B. POLLERMANN; World-Wide Web: the Information Universe; Electronic Networking, Vol. 2, No. 1, 1992, pp. 52-58

[Borenstein 92] BORENSTEIN, N., FREED, N.; RFC 1341; MIME (Multipurpose Internet Mail Extensions): Mechanisms for specifying and describing the format of Internet message;

- [Collart 94] S. COLLART; The Future of Publishing in Mathematics (Editorial); Euro-math Bulletin, Vol. 1, No. 2, 1994, pp 5-6
- [DalitzGLS 95] W. DALITZ, M. GRÖTSCHHEL, J. LÜGGER, W. SPERBER; Verteiltes Informationssystem für die Mathematik; DMV-Mitteilungen, Heft 1, 1995, S. 52-54
- [DalitzGLS 94] W. DALITZ, M. GRÖTSCHHEL, J. LÜGGER, W. SPERBER; Neue Perspektiven eines verteilten Informationssystems für die Mathematik; GAMM Rundbrief; Sept. 1993; auch Report TR-94-7, ZIB-Berlin, August 1994; <ftp://elib.zib-berlin.de/pub/zib-publications/reports/TR-94-7>
- [DMV 95] DEUTSCHE MATHEMATIKER-VEREINIGUNG; Entwurf eines Projektantrages an den BMBF zur Realisierung eines Verteilten Informationssystems für die Mathematik in Deutschland, Redaktion: W. Dalitz, M. Grötschel, G. Heyer, J. Lügger, W. Sperber; Stand: 20. April 1995
- [DongarraR 91] J. J. DONGARRA, B. ROSENER; NA-NET - Numerical Analysis NET; Computer Science Dept. Univ. of Tennessee, 1991
- [DPG 94] DEUTSCHE PHYSIKALISCHE GESELLSCHAFT E.V.; Integrierendes Informationsmanagement im Bereich der Physik; Entwurf eines Antrags an dem BMFT (fachlicher Teil) vom 25. März 1994
- [EdwardsH 89] D. M. EDWARDS, L. HARDMANN; Lost in Hyperspace: Cognitive Mapping and Navigation in a Hypertext Environment; in: Hypertext: Theory and Practice; R. McAleese (Ed.); Blackwell Scientific Publications Ltd., 1989, pp 105-125
- [GI 94] GESELLSCHAFT FÜR INFORMATIK E.V.; Fördervorhaben Fachinformation; Entwurf eines Antrags an den BMFT (fachlicher Teil) vom 24. Mai 1994
- [FIZ-KA 94] FACHINFORMATIONSZENTRUM KARLSRUHE UND HEIDELBERGER AKADEMIE DER WISSENSCHAFTEN (HERAUSGEBER); 1994 Mathematics Subject Classification; Springer Verlag, 1994
- [Franks 93] J. FRANKS; The impact of electronic publication on scholarly journals; Notices Amer. Math. Soc. 40 (1993), pp 1200-1202
- [Ginsparg 94] P. GINSPARG; First Steps Towards Electronic Research Communication; Computers in Physics, Vol. 8, No. 4, Jul/Aug 1994
- [GrötschelL 95a] M. GRÖTSCHHEL, J. LÜGGER; Die Zukunft wissenschaftlicher Kommunikation aus Sicht der Mathematik; Spektrum der Wissenschaft, März 1995, S. 39-43; elektronisch via <http://elib.zib-berlin.de:8000/Cmath.rd.vismath.wisskomm>
- [GrötschelL 95b] M. GRÖTSCHHEL, J. LÜGGER; Wissenschaftliche Kommunikation am Wendepunkt – Bibliotheken im Zeitalter globaler elektronischer Netze; erscheint in "Zeitschrift für Bibliothekswesen und Bibliographie", Vol. 42, Heft 3, 1995
- [Grötschel 94] M. GRÖTSCHHEL et al.; Distributed Electronic Information System for Mathematics - Short description of a project planned by the DMV; DMV-Projekt-Leitung; ZIB-Berlin, September 1994; electronically via <http://www.zib-berlin.de/SoftInf/NewPerspShort.html>

- [GrötschelLS 93] M. GRÖTSCHHEL, J. LÜGGER, W. SPERBER; Wissenschaftliches Publizieren und Elektronische Fachinformation im Umbruch: ein Situationsbericht aus der Sicht der Mathematik; Jahrbuch Überblicke Mathematik 1994, Hrsg.: Chatterji, Fuchssteiner, Kulisch, Liedl; Vieweg 1993, S. 154-174
- [KappeAF 94] F. Kappe, K. Andrews, J. Faschingbauer; Hyper-G: A New Tool for Distributed Hypermedia IICM Report 388, TU-Graz 1994 report388.ps in <ftp://iicm.tu-graz.ac.at/pub/Hyper-G/papers>
- [Kappe 93] F. KAPPE; Hyper-G: A Distributed Hypermedia System; Proc. INET'93, San Francisco, Calif., pp. DCC-1 - DCC-9, Aug. 1993; auch inet93.ps in <ftp://iicm.tu-graz.ac.at/pub/Hyper-G/papers>
- [KappeM 93a] F. KAPPE, H. MAURER; From Hypertext to Active Communication/Information Systems IICM Report 363, TU Graz, 1994; auch report363.ps in <ftp://iicm.tu-graz.ac.at/pub/Hyper-G/papers>
- [KappeM 93b] F. KAPPE, H. MAURER; Hyper-G: A Large Universal Hypermedia System and Some Spin-Offs; ACM SIGGRAPH's experimental special online issue of "Computer Graphics" (May 1993); auch IICM Report 364, TU-Graz, 1993; report364.ps in <ftp://iicm.tu-graz.at/pub/Hyper-G/papers>
- [Krol 94] E. KROL; The Whole Internet - Users Guide & Catalog, Second Edition; O'Reilly & Associates, Inc., 1994; 544 S.
- [LANL] LOS ALAMOS NATIONAL LABORATORY; Internet Preprint Services für die Hochenergiephysik via electronic Mail mit Archiv im WorldWideWeb: <http://xxx.lanl.gov/>
- [Lügger 94] J. LÜGGER; Kriterien zur Analyse und Bewertung von Informationssystemen und zur Gestaltung von elektronischen Informationsangeboten auf der Basis globaler Hypertextsysteme; Vorläufige Version vom 8. September 1994; ZIB-Berlin 1994; Diskussionsbeitrag zur Liste [dmv-ug-pvefm@zib-berlin.de](mailto:dmv-ug-pvefm@zib-berlin.de) vom 13.9.94
- [LüggerD 93] J. LÜGGER, W. DALITZ; Opt-Net: A New Communication Facility for Scientists Interested in Optimization; GMÖOR Newsletter 2, 1993, pp. 7-9
- [Nielsen 90a] J. NIELSEN; The Art of Navigating Through a Hypertext; Comm. ACM 33, 1990, pp 267-310
- [Nielsen 90b] J. NIELSEN; Hypertext & Hypermedia; Academic Press, 1990
- [NIST] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, USA GAMS: Guide to Available Mathematical Software; <http://gams.nist.gov/>
- [Odlyzko 95] A. M. ODLYZKO; Tragic loss or good riddance? The impending demise of traditional scholarly journals; To be published in Intern. J. Human-Computer Studies (formerly Intern. J. Man-Machine Studies). Condensed Version to appear in Notices Amer. Math. Soc., Jan. 1995; auch elektronisch via <ftp://elib.zib-berlin.de/netlib/att/math/odlyzko/index>

- [Rehmann 95] U. REHMANN; Das elektronische Journal "Documenta Mathematica" der DMV; Vortrag im Rahmen des Workshops "Neue Wege der wissenschaftlichen Information und Kommunikation"; TU Berlin, 8. - 10. März 1995; Vortragsfolien via <ftp://elib.zib-berlin.de/pub/Workshop/TU-Berlin/Rehmann>
- [Stix 95] G. STIX Publizieren mit Lichtgeschwindigkeit; Spektrum der Wissenschaft, März 1995, S. 34-39
- [Quinn] F. QUINN; Roadkill on the electronic highway? The threat to the mathematical literature; To appear in Notices Amer. Math. Soc.

## A Vergleichende Kostenbetrachtung für die Weitergabe von Reports durch mathematische Institutionen

Dieser Anhang enthält eine vergleichende Abschätzung der Kosten, die einem mathematischen Fachbereich oder Forschungsinstitut entstehen, wenn

- a) die Vervielfältigung und der Versand von Reports (Preprints etc.) auf traditionelle Art und Weise (via Papier) organisiert wird, gegenüber
- b) den Selbstkosten für die elektronische Bereitstellung auf einem Server, der an das WIN (des DFN) bzw. das Internet angeschlossen ist.

### A.1 Annahmen für die Kostenabschätzung

Für die Abschätzung sei ein

Report mit dem Umfang von 24 Seiten (mit farbigem Einband)

zugrunde gelegt, der mit weißer Titelseite (natürlich ohne Einband) in

elektronischer Form	40K Byte als komprimierter $\text{T}_{\text{E}}\text{X}$ / $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ -Code
	160K Byte als komprimierter Postscript-File

---

insgesamt also                      200K Byte Speicher auf der Platte belegt.

Zur Vereinfachung der Abschätzung sei vorausgesetzt, daß ein mittelgroßes Forschungsinstitut etwa 50 Reports dieser Art pro Jahr erstellen wird, die es an bestimmte Kooperationspartner verschickt bzw. vermittelt. Diese Annahmen entsprechen im wesentlichen denen von A. Odlyzko, der eine ähnliche Abschätzung für mathematische Publikationen vorgenommen hat [Odlyzko 94].

Ferner sei angenommen, daß es sinnvoll ist, einen "elektronischen Report" in zwei Formen bereitzustellen: als Postscript-File, der für den Bibliothekar eines Instituts oder Fachbereichs geeignet ist, und als Tex- oder  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ -File, der für den Mathematiker eher geeignet ist. Die Postscript-Form dient vor allem dem originalgetreuen Abdruck mittels Laserwriter. Die textuelle Version des Reports kann nicht nur abgesucht und auch weiterverarbeitet, sondern, wenn sie in einer Datenbank integriert wird, auch lokal recherchierbar bereitgestellt werden.

### A.2 Vervielfältigung und Versand auf Papier

Für die Berechnung der Kosten für die Papierversion sei zugrunde gelegt, daß ein Report in einer Auflage von 300 Exemplaren erstellt wird. Dann kosten die Vervielfältigung (Preisangaben mit Stand vom Dezember 1994, in Berlin) und der Versand etwa:



DM	6,13	pro gebundenes Exemplar (incl. MWST.)
DM	0,22	Personalkosten (1/2 Minute einer Hilfskraft Eintüten, Adressieren, Frankieren) plus Umschlag
DM	2,-	Porto (Mittelwert für Versand in Europa)
<hr/>		
DM	8,35	1 Exemplar
DM	2.500,-	300 Exemplare
DM	125.000,-	15.000 Exemplare (50 Reports, je 300 Exemplare)

In diesem Betrag sind der Einfachheit halber nicht die Kosten für den Eigenbedarf enthalten. Auch die Kosten für die notwendige Infrastruktur (Personalkosten und Raumbedarf für Mitarbeiter der Verwaltung und der Bibliothek) sind in dieser Betrachtung außer acht gelassen, weil diese Infrastruktur ohnehin vorgehalten werden muß.

### A.3 Elektronische Bereitstellung im Netz (Internet)

Die Selbstkosten für die elektronische Bereitstellung setzen sich aus verschiedenen Komponenten zusammen aus jeweiligen Anteilen für:

- \* Plattenspeicher, Workstation und Wartung der Hardware
- \* Betriebssystem und allgemeine Software (incl. Upgrades),
- \* die Belastung des Netzes (Internet).

Um zu einem fairen Vergleich zu kommen, sollen diese Selbstkosten in die ersten dreihundert elektronischen "Exemplare" voll hineingerechnet werden, dabei sollen allerdings nur die jährlichen Kosten angerechnet werden. Für die Abschätzung wird angenommen, daß die Hardware in drei Jahren abgeschrieben ist. Der Einfachheit halber wird die Software ähnlich behandelt.

Zu den Investitionen: Auf eine heute handelsübliche Platte von 4G Byte Speicherkapazität passen also 20.000 Reports (zu je 200K Byte). Solche Platten kosten heute – bei Neuanschaffung – etwa 4.000,- DM. Eine einfache Workstation oder ein entsprechender (leistungsfähiger PC) kostet heute etwa 15.000,- DM. Die Software wird insgesamt mit 1.200,- DM (plus 200,- DM für Upgrades) veranschlagt. An jährlichen Kosten entstehen:

DM	1.334,-	Plattenspeicher
DM	5.000,-	Workstation
DM	1.000,-	Wartung
DM	600,-	Software
<hr/>		
DM	7.934,-	Summe pro Jahr, also
DM	0,40	je Exemplar (Abschreibung 3 Jahre)

Zu den Netzkosten: Diese Abschätzung geht von den in Deutschland z.Z. für den Wissenschaftsbereich geltenden Preisen für einen 2MBs-Anschluß (2 Megabit pro Sekunde) im WIN (Wissenschaftsnetz) mit Mehrwertdiensten (Internet Anschluß mit entspr. Diensten) aus. Bei einer Inanspruchnahme durch eine mittelgroße Institution (wie dem ZIB) fallen hier jährlich DM 400.000,- an. Dafür steht solch ein Anschluß im Prinzip rund um die Uhr (das ganze Jahr) zur Verfügung (insgesamt 525600 Minuten bzw. 31536000 Sekunden). Die Inanspruchnahme des Netzes kostet also anteilig einen Betrag von DM 0,01268391679350583460 pro Sekunde (Kosten des Informationsanbieters).

Dafür kann man aus allen Teilen der Welt auf solch einen Anschluß zugreifen. Natürlich ist diese Leistung nicht “echt”. Das Netz wird nicht “rund um die Uhr” voll in Anspruch genommen und manchmal gibt es auch Ausfälle (des Netzes oder der Hardware vor Ort). Die entsprechenden Leistungsverluste sind leider nur sehr schwer abzuschätzen. Um auf der sicheren Seite zu liegen, sei angenommen, daß ein 2MBs-Anschluß nur 1MBs echte Übertragungsleistung bietet, das sind 125K Byte/sec. Da auch die für die Anwendung der Dienste notwendigen Protokolle ihren Leistungsanteil beanspruchen (und es manchmal auch Engpässe oder fehlerhafte Übertragungen gibt), gehen wir im folgenden von einer echten Übertragungsleistung von nur 100K Byte/sec aus (dieser Ansatz ist insgesamt sicherlich konservativ).

Demnach belastet die Übertragung von Reports (jeweils in T<sub>E</sub>X /L<sup>A</sup>T<sub>E</sub>X- UND in Postscript-Form) das Netz mit einem Anteil von

2	sec	pro	Report;	das entspricht	DM	0,02536
600	sec	für	300	“Kopien”	DM	7,60
30.000	sec	für	15.000	“Kopien”	DM	380,50

Die Gesamtkosten für das elektronische Angebot eines mathematischen Reports im Internet belaufen sich für ein wissenschaftliches Institut damit (im Jahr) auf

DM	0,40	für	1	Exemplar	(Hardware, Software, Wartung)
DM	7,60	für	300	“Kopien”	(Netzkostenanteil)
DM	8,-	Summe pro	Exemplar	mit 300	“Kopien”

bzw. – bei einer Produktion von 50 Reports im Jahr – auf

DM	20,-	für	50	Exemplare	(Hardware, Software, Wartung)
DM	380,50	für	15.000	“Kopien”	(Netzkostenanteil)
DM	500,50	für die gesamte	gesamte	Jahresproduktion	

Die Kosten für die notwendige Infrastruktur sind in diesen Betrag nicht eingerechnet, weil ein mathematisches Institut diese heute ohnehin vorhalten muss. Auch die Personalkosten für das Umkopieren von Reports auf den öffentlich zugänglichen Speicher sind nicht extra aufgeführt, da die hauptsächliche Arbeit für die Erstellung von Reports in T<sub>E</sub>X - oder L<sup>A</sup>T<sub>E</sub>X -Form ohnehin bei den Sekretariaten eines mathematischen Instituts anfällt und eine ordnungsgemäße Archivierung in jedem Falle (sowohl bei der Vervielfältigung in traditioneller Form als auch bei der elektronischen Bereitstellung) vorgenommen werden muß (in der Regel von Mitarbeitern der Bibliothek des Hauses).

Nach der ersten Bereitstellung und der Abnahme (“Kopie” über das Netz) können – im elektronischen Falle – im Prinzip beliebig viele, weitere “Kopien” kostengünstig abgegeben werden, denn hierbei sind ja die bereits abgeschriebenen Hard- und Softwarekosten nicht mehr anzurechnen.

Die Abgabe von weiteren “Kopien” eines elektronisch bereitgestellten Reports kostet das Institut nunmehr nur noch

2.000 sec für 1.000 weitere Kopien, DM 25,40 (anteilige Netzkosten)

Für etwa DM 25,- könnte ein mathematisches Institut in Deutschland also theoretisch praktisch alle mathematischen Institute der westlichen Welt (wenn man deren Anzahl mit etwa 1000 ansetzt) zusätzlich mit (s)einem Report versorgen.

## A.4 Diskussion einiger Auswirkungen

Die hauptsächlichlichen Kosten entstehen in Deutschland durch die anteilige Belastung des Netzes. Es ist zu erwarten, daß Netzkosten in wenigen Jahren etwa auf denselben Betrag fallen, der in den Vereinigten Staaten schon heute üblich ist, nämlich auf etwa ein Zehntel dessen, was in Deutschland dafür zu zahlen ist.

Bei einem Server, der in den USA steht, führt eine analoge Rechnung dazu, daß die elektronische Weitergabe eines Reports das mathematische Institut schon heute mit etwa nur

DM 0,40 für das erste und  
DM 2,50 für 1.000 weitere Exemplare

belastet. Wird solch ein Institut etwa 1.000 Rechnungen schreiben, um diesen Betrag bei seinen Partnern wieder einzutreiben? Oder ist es nicht sinnvoller, wenn sich alle mathematischen Institute auf dieser Basis gegenseitig – und “kostenlos” (jedes Institut trägt nur seine eigenen Kosten für das Informationsangebot) – mit ihren Arbeitsergebnissen versorgen?

Außerdem muß man berücksichtigen, daß die Kosten für das elektronische Informationsangebot in den kommenden Jahren weiterhin drastisch fallen (nicht nur in den Vereinigten Staaten). Erfahrungsgemäß reduzieren sich die Preise für die Hardware alle zwei Jahre um die Hälfte bzw. man bekommt für denselben Betrag etwa das Doppelte an Leistung. Bezogen auf Netze (Steigerung der Übertragungsleistung) ist eine ähnliche Entwicklung zu beobachten, auch in Deutschland, wo vor wenigen Jahren Übertragungsraten von 2MBs noch nicht einmal angeboten wurden.

Die Kosten für traditionelle Informationsangebote auf Papier werden in Zukunft dagegen eher zunehmen. Damit deutet sich ein grundlegender Wandel in der Art und Weise an, in der Information in Zukunft vermittelt wird.

Der dramatische Kostenunterschied ergibt sich für Informationsanbieter im wesentlichen dadurch, daß beim elektronischen Angebot der Anteil für die Vervielfältigung (auf Papier) auf die Abnehmer von Information verschoben wird. Diese haben dann natürlich die eigentlichen Kosten der Vervielfältigung (einschließlich der Bereitstellung der dafür notwendigen Infrastruktur) zu tragen, was letztlich auch fair ist.

Das ist aber genau die Situation, vor die sich die traditionellen Informationsanbieter (Verlage etc.) gestellt sehen, wenn sie ihre Produkte elektronisch anbieten. Sie werden völlig neue Formen und Verfahrensweisen der Distribution entwickeln müssen, um ihren Platz in der elektronischen Welt der Kommunikation zu behaupten.

Diese Kostenbetrachtung sollte plausibel machen, warum – nicht nur im Wissenschaftsbereich – auf aller Welt Server entstehen, die ihre Informationen kostenfrei (für Nutzer) anbieten. Denkt man an die Weitergabe von Reports durch ein (mittelgroßes) mathematisches Institut, so belastet die Speicherung und die Übermittlung dieser elektronischen Dokumente die dafür notwendige Hardware (Workstation und Platte) nur zu einem Bruchteil ihrer Leistungsfähigkeit und das Netz nur zu Bruchteilen von Minuten oder Stunden selbst dann, wenn man einen Report direkt an alle per E-mail erreichbaren Mathematiker der Welt verschicken würde.

Wissenschaftliche Einrichtung können Leistung dieser Art, sofern sie über die notwendige, technische Infrastruktur verfügen, praktisch nebenbei abgeben. Die dafür notwendige

Hardware können sie (hauptsächlich) für andere Zwecke einsetzen, z.B. im Rahmen ihrer wissenschaftlichen Arbeiten.

Wohin die Reise geht, soll eine Zielvorstellung verdeutlichen, die im Zusammenhang mit den Datenautobahnen genannt wird, die z.Z. in den Vereinigten Staaten entstehen. Dort soll im Prinzip nicht nur jede größere Einrichtung, sondern jeder Haushalt an die "Infobahn" angeschlossen werden. Eine der kommerziell interessantesten Anwendung ist es z.B., die Haushalte auf diesem Wege mit Spielfilmen zu versorgen (video on demand). Die Übertragung eines Videos von 90 Minuten Länge sollte dazu aber nicht mehr als ein paar Dollar kosten. In digitalisierter Form belegen Videos heute etwa 10 Gigabyte (bzw. 2,5 Gigabyte in komprimierter Form) an Speicher. Einer Abschätzung von A. Odlyzko [Odlyzko 94] zufolge ist das eine Größenordnung, die etwa der gesamten Weltproduktion an mathematischen Publikationen eines Jahres entspricht.

Es ist zu erwarten, daß auch der Wissenschaftsbereich von diesen Entwicklungen profitieren wird. Die heute diskutierte Kommerzialisierung des Internet wird das Internet sicherlich in Zukunft nicht teurer, sondern erheblich preiswerter machen. Für die elektronische Distribution von Reports werden die Mathematischen Institute in Zukunft nur noch Bruchteile von Pfennigen aufwenden müssen, selbst dann, wenn sie diese an alle interessierten Mathematiker der Welt, die an das Internet angeschlossen sind, persönlich verteilen.