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

Martin Grötschel
Joachim Lügger

Aufbau elektronischer Informations- und Kommunikationsstrukturen

Technical Report TR 95–10 (September 1995)

Aufbau elektronischer Informations- und Kommunikationsstrukturen

M. Grötschel, J. Lügger
Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)
Heilbronner Str. 10, 10711 Berlin-Wilmersdorf

Zusammenfassung

Die vorliegende Arbeit ist die (z.T. wesentlich) erweiterte Fassung eines Vortrages auf dem Deutschen Dokumentartag im September 1995 an der Fachhochschule Potsdam. Unser Ziel ist eine Bestandsaufnahme der gegenwärtigen Entwicklungen auf dem Gebiet der elektronischen Information und Kommunikation. Wir diskutieren Chancen und Risiken, beleuchten gesellschaftliche Rahmenbedingungen - vor allem im Vergleich mit den Vereinigten Staaten von Amerika - und skizzieren Aufgaben im Bereich von Forschung, Staat und Wirtschaft, die unserer Meinung nach noch zu bewältigen sind.

Wir führen insbesondere in Gebiete ein, die vermutlich in naher Zukunft die Landschaft des Dokumentierens und Archivierens wesentlich verändern werden. Wir benutzen dabei häufig Beispiele und Modellvorstellungen aus der Mathematik, unserem eigenen Fachgebiet, die u.a. zeigen, wie Elemente des Informierens, Dokumentierens und Archivierens durch die neuen Technologien der elektronischen Datenverarbeitung und Kommunikation ineinander fließen. Mehr Leistung kann mit geringeren Kosten erzielt werden. Wir schließen mit einer Auflistung von neuen Perspektiven und Möglichkeiten zum Handeln in diesem Bereich.

0. Ein Blick zurück

Werfen wir einen Blick zurück - in die Vergangenheit - auf die Tätigkeit eines Dokumentars in einem wissenschaftlichen Dokumentationszentrum, z.B. in Deutschland, das vielleicht mit einem amerikanischen Abstractservice kooperiert, um gemeinsam eine Datenbank zu erstellen. Der Dokumentar wertet wissenschaftliche Literatur aus und schreibt seine Ergebnisse in ein Formular, datenbankgerecht, aber auf Papier. Dieses Formblatt überträgt eine Schreibkraft in eine lesbare Form, mit der Schreibmaschine. Die so in aufwendiger "Handarbeit" erstellten Formblätter sieht ein Redakteur noch einmal durch, sammelt sie und schickt sie - eventuell nach weiteren Korrekturdurchgängen - in die Vereinigten Staaten, per Post. Dort erst erfaßt sie der Produzent der Datenbank in elektronischer Form. So war es doch damals. Aber, liegt diese Zeit wirklich weit hinter uns zurück? War es nicht erst gestern?

Wir haben einen Blick in die vergangene Ära der zentralen Großrechner getan, so kommt es uns vor. Nur unter dieser Perspektive scheint uns die geschilderte Verfahrensweise noch verständlich zu sein. Leistungsfähige Rechner waren damals außerordentlich teuer, ebenso die Kommunikationsverbindungen, die sehr langsam waren und oft nur über Standleitungen realisiert. Personalkosten fielen dagegen weit weniger ins Gewicht. Die mit den teuren Datenbanken operierenden Fachkräfte waren darauf geschult, elektronische Betriebsmittel äußerst sparsam einzusetzen. Spezielle Informationsvermittler mußten auf der Nutzerseite dafür Sorge tragen, mit möglichst "genau gezielten" Anfragen (was immer darunter zu verstehen ist) ein Maximum an relevanten Antworten aus einer solchen Datenbank "herauszukitzeln". Darauf wurden sie besonders "gedrillt". Wissenschaftler haben kaum die Zeit, diesen hohen Aufwand zu treiben, und sie nehmen die - für sie - recht umständliche Informationsvermittlung nur dann in Anspruch, wenn es sich nicht vermeiden läßt, zum Beispiel, wenn ihre Forschungsarbeit besondere finanzielle Risiken birgt, die sich auf diese Weise vielleicht verringern lassen.

1. Über das Internet zu den Datenautobahnen

Aber hat nicht die elektronische Revolution unsere Wirklichkeit längst verändert? *PCs* und *Workstations* sind sehr leistungsfähig und preiswert geworden, und sie sind sehr weit verbreitet. Wer erstellt denn heute seine Texte nicht elektronisch? In der Mathematik, der Physik und der Informatik setzen die Autoren ihre Arbeiten selbst, in professionellem Stil, mit $\LaTeX$ [Lamport 86]. Moderne Textprozessoren, die in anderen Wissenschaftsbereichen zum Einsatz kommen, erzeugen ebenfalls elektronische Texte in hoher drucktechnischer Qualität *(Postscript)*. Wissenschaftliche Arbeiten liegen also bereits elektronisch vor, lange bevor sie auf Papier publiziert und - in Handarbeit - dokumentiert werden.

Wir wenden uns jetzt dem jüngsten Kapitel der elektronischen Revolution zu, dem Internet und den Datenautobahnen. Die neuen Technologien der elektronischen Kommunikation beginnen ebenfalls unser Leben zu verändern - nicht nur in den Wissenschaften.

1.1 Die transformierende Kraft des Internet

Wissenschaftler verteilen ihre Artikel heute weit im Vorfeld der Publikation, als *Preprint*, und schon vorher - elektronisch - als "Manuskript". Sie diskutieren und kommentieren im Kreis von Fachkollegen und *in-groups* ihre Ergebnisse, per E-mail im Internet, bilateral oder - per öffentlichem *E-mail Verteiler* - gleich in größeren Gruppen. Sie stellen Preprints zunehmend in *ftp-Archiven* bereit, die man oft auch absuchen kann (*Archie, WAIS, Jughead*; für eine Einführung in die Begriffswelt des Internet siehe [ACM 94], [Krol 94] und [Flynn 95]). Das interaktive *World Wide Web*, kurz *Web* genannt, hat den Zugang zum Internet überaus einfach gemacht. Es beginnt bereits die Art und Weise zu transformieren, in der Wissenschaftler miteinander kommunizieren. Die Theoretischen Hochenergiephysiker und andere Physikergruppen verteilen ihre Preprints heute fast nur noch elektronisch und halten sie - recherchierbar und kostenfrei - im Web bereit[1] [Ginsparg 94] [Stix 95], ein Verfahren, das schon in anderen Wissenschaftsbereichen[2] Schule macht; siehe [Grötschel 95a] für eine kritische Betrachtung. In den Naturwissenschaften ist die Arbeit ohne Anschluß an das Internet kaum noch denkbar. Wer hier den "Anschluß" verpaßt, ist bald isoliert.

Wissenschaftliche Fachgesellschaften, Forschungsinstitute und Universitätsfachbereiche

[1]Hier und im weiteren listen wir die elektronischen Quellen auf, die wir im Text zitieren, insbesondere damit der Leser die von uns gemachten Angaben verifizieren kann. Mittels eines Browsers wie Mosaic, Netscape etc. können die jeweiligen elektronischen Archive über die folgenden URLs abgefragt werden:
http://xxx.lanl.gov/ - TH-HEP Archiv (P. Ginsparg's Server)

[2]http://eprints.math.duke.edu/archive.html - Duke Universität und MSRI Preprints

können jetzt mit geringem Aufwand eigene elektronische Zeitschriften herausgeben - und sie tun dies auch. In der Mathematik gibt es heute bereits an die 30 elektronische Journale, alle referiert, die ihre Artikel *online* im Web offerieren, als Volltexte in $\LaTeX$- oder in *Postscript*-Form. Wer über einen *Browser* wie *Netscape* oder *Mosaic* verfügt, kann darauf zugreifen, rund um die Uhr und kostenfrei, direkt von seinem Arbeitsplatz aus. Einige dieser Journale informieren Subskribenten per E-mail über neue Artikel, augenblicklich, sobald diese akzeptiert sind, wie z.B. das *Electronic Journal of Combinatorics*[3] (ElJC). Die technische Bereitstellung solcher Veröffentlichungen in einem E-Journal nimmt beim ElJC nur wenige Minuten in Anspruch, sagt N. Calkin, einer seiner Herausgeber. Und, der Prozeß der Begutachtung wurde zuvor mittels E-mail beschleunigt. In Deutschland bereitet die *Deutsche Mathematiker-Vereinigung* (DMV) ein universelles, mathematisches E-Journal vor, die *Documenta Mathematica*[4] - es wird voraussichtlich Anfang 1996 erscheinen [Rehmann 95].

Fachgruppen verschiedener mathematischer Teildisziplinen (in Deutschland die Fachgruppen Computeralgebra[5], Optimierung[6], Scientific Computing[7] und Stochastik[8] der DMV) betreiben eigene elektronische Informationssysteme, mit moderierten *Diskussionsforen* und *Web-Servern*, mit denen sie - weltweit - aktuelle Nachrichten verteilen und "Datenbanken" für ihre Fachgebiete führen. Sie verwenden dabei *Hypertext-Systeme*, die sehr einfach zu benutzen sind und keine aufwendige Schulung verlangen; siehe [Nielsen 95], [Flynn 95] und [Graham 95] für eine Einführung. Ihre Informationsangebote enthalten oft nicht nur einfache Kommentare, sondern neben Artikeln und Bibliographien auch umfangreiche Software- und Datensammlungen aus der aktuellen Forschung. Die bekannte Netzbibliothek für Numerik *netlib*[9] enthält mehrere solcher Sammlungen [DongarraG 87]. Sie umfaßt heute über ein halbes *Gigabyte* (mehr als eine halbe Milliarde Zeichen) an Programmen und Dokumenten. Über 10.000 Programmquellen sind dort zu finden, und sie können - es sind viele mehrfach verwendbare Routinen darunter -, nach den verschiedensten Gesichtspunkten kombiniert, abgefragt und abgerufen werden [BrowneDGGMRW 94].

Das Web hat die Wissenschaften im Sturm genommen. Jede Institution, die etwas auf sich hält, "erscheint" mit einem eigenen Server im Internet. Praktisch alle mathematischen Fachbereiche in Deutschland[10] [11] (natürlich auch in den USA) haben inzwischen Web-Server errichtet, in denen sie ihre Projekte und Aufgaben in Forschung und Lehre darstellen. Viele Wissenschaftler bieten innerhalb dieser Server nicht nur einfache Adressenangaben an, mancher auch seine Veröffentlichungen und sogar sein *Curriculum Vitae*. Die großen wissenschaftlichen Fachgesellschaften, wie die *American Mathematical Society*[12] (AMS), die DMV[13], die *European Mathematical Society*[14] (EMS) oder die *International Mathematical*

[3] http://ejc.math.gatech.edu:8080/Journal/journalhome.html - ElJC
[4] http://www.mathematik.uni-bielefeld.de/DMV-J/ - Documenta Mathematica der DMV
[5] http://www.uni-karlsruhe.de/~CAIS/ - CAIS
[6] http://moa.math.nat.tu-bs.de/opt-net/opt-net.html - Opt-Net Archiv
[7] http://www5.informatik.tu-muenchen.de/sci-comp/home.html - GSCI-Archiv
[8] http://elib.zib-berlin.de:8000/Ceconf.st-net - ST-Net Archiv
[9] ftp://netlib.att.com/netlib/master/readme.html.Z - netlib
[10] http://esther.mathematik.uni-osnabrueck.de/imagemaps/karte.html - Mathematik in Deutschland
[11] http://www.rz.uni-karlsruhe.de/Outerspace/VirtualLibrary/51.de.html - German Mathematics
[12] http://e-math.ams.org/ - AMS (e-Math)
[13] http://www_dmv.math.tu-berlin.de/ - DMV
[14] http://www.EMIS.de/ - EMS

Union[15] (IMU) stehen ihnen nicht nach. Wissenschaftliche Kongresse, wie der *International Congress of Mathematicians*[16] (ICM'98), kündigen sich ebenfalls im Web an. Interessenten können sich elektronisch informieren und zur Konferenz anmelden.

1.2 Von der Bedeutung des Internet für die Wissenschaften

Das Internet wächst mit unglaublicher Geschwindigkeit. Die Fachzeitschriften und sogar die allgemeine Tages- und Wochenpresse berichten ausführlich darüber. Was hat das Internet so attraktiv und populär gemacht? Es ist nicht nur der einfache Zugang über das Web, das heute synonym für das Internet steht. Es ist auch nicht die besondere Attraktivität allein, die mit dem Mosaic-*Browser* in das Internet kam [BernersCGP 92] [AndreessenB 94], durch die Verwendung von Hypertexten mit ihren *Hyperlinks (URLs)* und die bunten Bilder und Videos *(Multi-Media-Information)*. Für die Wissenschaftler ist es vor allem die Möglichkeit, mit elektronischen Mitteln schnell und effizient über große Entfernungen hinweg miteinander zu kommunizieren, sowie auf die vorhandene Fülle an Informationen[17], die schon seit einiger Zeit im Internet gewachsen ist, einfach zugreifen zu können. Daß diese Art der Information, die mit bisherigen Internet-Werkzeugen etwas umständlich zu handhaben war (verschiedene Kommandos, Protokolle und Zugriffstechniken), von Anfang an in das Web integriert und dadurch ebenfalls mittels einfacher *Mausclicks* erreichbar wurde, ist einer der wichtigsten Gründe für das anhaltende dynamische Wachstum des Web.

Ein bedeutender Faktor für die weltweite Verbreitung des Internet ist auch sein spezielles Nachrichtensystem, die *UseNet News* (kurz *News* genannt), das über Neuigkeiten informiert und - durch elektronische Diskussionen - einer großen, über viele Länder verstreuten Gemeinschaft die aktive Mitarbeit ermöglicht. An der Entwicklung des Internet und seiner Werkzeuge haben mit Hilfe der News und den vielen Informationsquellen, auf die sie verweisen, - durch das Internet selbst - nicht nur Kommunikationsspezialisten teilnehmen können, sondern gerade auch seine Nutzer mit ihren besonderen Bedürfnissen.

Die News haben in den Wissenschaften allgemein nicht die Bedeutung wie in der Kommunikationstechnik. Sie werden vielfach wegen ihrer "Geschwätzigkeit" und mangelnden Qualität kritisiert. Trotzdem sind sie das Vorbild für elektronische Diskussionsforen, die jetzt - auf der Basis von E-mail - diszipliniert und moderiert eingesetzt, auch in die Gemeinschaft der Mathematiker Eingang finden [DongarraR 91] [DalitzL 93]. Wer ein spezielles Problem hat oder eine besondere Frage diskutieren möchte, schickt sie in die News oder in ein Forum, das weltweit gelesen werden kann - und er wird von irgendwo einen Kommentar oder eine Antwort bekommen. Es ist diese *Interaktivität des Netzes*, die ebenfalls zu seiner breiten Akzeptanz geführt hat, und zunehmend der nun in Erscheinung tretende Reichtum an Informationen und Materialien.

Das Internet hat eine weltweite Massenbewegung ausgelöst, bei der die Information, die Dokumentation, die Archivierung und die Kommunikation jetzt miteinander verschmelzen und dabei auch völlig unvorhergesehene Formen annehmen. Die Medien können fast wöchentlich

[15] http://elib.zib-berlin.de:8000/Cimu - IMU
[16] http://elib.zib-berlin.de:8000/CICM98 - ICM'98
[17] http://www.w3.org/hypertext/DataSources/bySubject/Overview.html - WWW Virtual Library

über spektakuläre neuartige Anwendungen berichten, zum Beispiel über *Live*-Übertragungen von wissenschaftlichen Tagungen und Workshops im *MBone* des Internet, einem weltweiten Träger für Videoübertragungen [Erikson 94] [Schult 95]. Für die Wissenschaften sind vor allem die vielfältigen Einzelinitiativen von Bedeutung, die über alle geographischen Grenzen hinaus wirken, und die Möglichkeiten, auf den Schatz von Erfahrungen und Wissen einer großen Anzahl von im Internet aktiven Personen und Institutionen zuzugreifen. Hierzu ein Beispiel aus dem Bereich der bibliographischen Arbeit.

Viele Wissenschaftler der Informatik und der Mathematik führen im Rahmen ihrer eigenen Forschungsarbeiten Bibliographien, häufig im *BibTEX*-Format [Lamport 86]. Etliche Wissenschaftler haben diese "persönlichen" Dokumentationen inzwischen im Internet zur Nutzung durch Fachkollegen bereitgestellt. A.-Ch. Achilles, ein wissenschaftlicher Assistent der Informatik an der Universität Karlsruhe, hat nun "im Alleingang" und quasi nebenher circa 640 dieser lokal gespeicherten Bibliographien zu einer *Collection of Computer Science Bibliographies*[18] zusammengeführt, die jetzt über 360.000 Referenzen auf aktuelle Konferenz-Beiträge, Artikel in Journalen und technische Reports enthält. Sie enthält 165 Megabyte an BibTEX-Daten, auch *Hyperlinks* auf andere bibliographische Informationssysteme, insgesamt über 1.100 Verweise (URLs). Diese bibliographische "Datenbank" ist nun im Internet kostenfrei recherchierbar, und sie wird gegenwärtig etwa 1.000 mal am Tag genutzt (jede Abfrage, wie im Web üblich, einzeln gezählt). Inzwischen spiegeln 12 weitere Server in verschiedenen Regionen diese Informationen und erzielen ähnliche Nutzungsfrequenzen.

Das alles beruht auf persönlichen Aktivitäten und Initiativen auf breiter Front, ohne spezifische Förderung durch den Staat, allerdings im wissenschaftlichen Interesse der betreffenden Institute, aber ohne die Möglichkeit (und die Absicht), damit Einnahmen zu erzielen. Unter der Voraussetzung, daß diese Daten kommerziell verwendet werden sollten, hätte Achilles wohl kaum die beteiligten Wissenschaftler dafür gewinnen können, ihm ihre Daten zu überlassen. Diese Einwilligung liegt ihm von allen (circa 600) Personen vor, deren persönliche Bibliographien in dieses Informationssystem integriert sind. Für die laufende Pflege benötigt Achilles nach eigenen Angaben ungefähr einen Tag in der Woche (die Zeit für die Entwicklung seiner spezifischen Software nicht mitgerechnet).

Die Daten haben nun möglicherweise nicht immer die technische Qualität, die ein professioneller Dokumentar erwarten würde. Jedoch stammen sie direkt aus der aktuellen Forschung und sind vermutlich auf dem neuesten Stand, denn sie werden laufend aktualisiert, von den beteiligten Wissenschaftlern selbst und durch turnusmäßiges Spiegeln der örtlich verteilten Information. Und, sie sind durch die Verwendung von BibTEX in hoher technischer Qualität ausdruckbar sowie von den Nutzern in eigenen Literaturlisten direkt fast ohne weitere Bearbeitung verwendbar. Wer seine Zeit mit dem (aufgrund verschiedener Konventionen und Formate notwendigen) häufig erforderlichen erneuten Eintippen von Referenzen vergeudet hat, weiß das zu schätzen.

Initiativen dieser Art sind typisch für das Internet. Sie sind erfreulich für die wissenschaftliche Welt, sind aber gleichzeitig von manchem professionellen Service-Anbieter gefürchtet, machen sie doch auch mangelnde Innovationsbereitschaft und monopolistische Preisgestaltung deutlich.

[18] http://liinwww.ira.uka.de/bibliography/index.html - Computer Science Bibliographies

1.3 Von den Anfängen des Internet bis zu den Datenautobahnen

Das Internet kommt aus den Vereinigten Staaten von Amerika. Seine Entwicklung wurde ursprünglich ins Auge gefaßt, um die Folgen eines atomaren Schlagabtausches abzumildern. Die militärischen Nachrichtenverbindungen mit ihren großen Computern und Kommunikationsnetzen sollten noch funktionsfähig sein, wenn größere Teile im Katastrophenfalle ausfielen. So entstand die Idee, Information in kleine Pakete zu zerlegen, die sich ihren Weg durch das Netz der noch funktionsfähigen Teilnetze selbst suchen können, wobei sie die noch existierenden Leitungen gemeinsam benutzen *(packet switching)*. Die *Advanced Research Project Agency*[19] (ARPA) des amerikanischen Verteidigungsministeriums vergab den Entwicklungsauftrag an Forschungslabors im Wissenschaftsbereich, wobei sie keine nähere militärische Zweckbindung verfügte. So kam es, daß der Vorläufer des Internet, das ARPANET, im Umfeld der "freien Forschung" (wie wir sagen würden) entstand. Es verband - als Netz der Netze - die damals nur von Rechnerherstellern gelieferten Netze [Kahn 94].

Zwei Eigenschaften des Internet sind für unser Thema von ganz besonderem Interesse.

1. Es hat die Entwicklung von rechnerunabhängigen Standards mit enormer Kraft vorangetrieben, nicht nur im Bereich der Kommunikationsprotokolle. Wo man früher beim Austausch von digitaler Information durch eine immense Vielfalt der unterschiedlichsten Speicherformate behindert wurde (Bänder, Platten, Floppydiscs, mit jeweils wieder verschiedenen Formaten für Programme, Texte, Bilder und Graphiken), verfügen wir heute über eine Fülle von rechnerunabhängigen Datenformaten für digitale Information fast jeder Art, bis hin zu Musik- und Videoinformation. Im Internet sind die Unterschiede fast verschwunden.
2. Durch das *packet switching* teilen sich die einzelnen Informationspakete auch die Kosten für die Nutzung der teuren Leitungen. Fest geschaltete Verbindungen, die von nur zwei Kommunikationspartnern benutzt werden können, gibt es in dieser Welt nicht mehr. Das führt langfristig gesehen zu einer enormen Kostenreduktion, denn wenn eine Leitung (Glasfaser oder Kupferkabel) erst einmal gelegt ist, hängt ihre Übertragungskapazität (auch *Bandbreite* genannt) fast nur noch von der Leistungsfähigkeit der Kommunikationsrechner ab, die sie benutzen. Diese hat sich im Zuge der elektronischen Revolution dramatisch gesteigert, und mit ihr die möglichen Bandbreiten - bei nahezu gleichbleibenden Kosten.

Die *National Science Foundation*[20] (NSF) hat das Internet ebenfalls vehement gefördert. Es trägt noch heute alle Charakteristiken eines Wissenschaftsnetzes. Wegen seiner technischen Eigenschaften war es nicht nur für die gemeinsame Nutzung von teuren Großgeräten geeignet, sondern auch für die Kollaboration größerer Gruppen von Wissenschaftlern der verschiedensten Institutionen über große Entfernungen hinweg - und für die effiziente und kostengünstige Verteilung wissenschaftlicher Information.

Das Internet ist etwas überspitzt gesagt ein "Äther für Informationsquanten", die in diesem Träger weltweit "gesendet" werden können. Jeder, der über einen Anschluß verfügt, kann

[19] http://ftp.arpa.mil/ - ARPA
[20] http://www.nsf.gov/ - NSF

solche Nachrichten "empfangen". Wir wollen mit dieser Formulierung zum Ausdruck bringen, daß es dem Rundfunk technisch sehr nahe steht - mit Konsequenzen für kommerzielle Anwender. Viele Informationsanbieter beginnen sich derzeit zu fragen, ob und wie sie im Internet Einnahmen erzielen können, denn für solch ein Medium gelten andere Gesetze als für den traditionellen Informationsmarkt.

Die besondere Vorgeschichte und die amerikanische Tradition hatten zur Folge, daß das Internet nicht auf den Wissenschaftsbereich begrenzt blieb. Bald begannen andere große Bundesbehörden der USA das Internet zu nutzen, wie die *National Aeronautics* and *Space Administration*[21] (NASA), das *Department of Energy*[22] (DOE), auch das *Weiße Haus*[23] selbst und nationale Bibliotheken[24] und Dokumentationszentren[25]. Das Militär machte bald den Anschluß an das *Milnet*, den militärischen Teil des Internet, zur Bedingung bei der Auftragsvergabe. So schlossen sich erste Teile der Industrie an. Heute durchdringt das Internet große Bereiche der amerikanischen Gesellschaft. Es ist das Vorbild für die vieldiskutierten Datenautobahnen - jetzt auch in Europa[26].

Die amerikanische Regierung hat sich mit der *National Information Infrastructure*[27] (NII) Initiative das Ziel gesetzt, die gesamte amerikanische Gesellschaft miteinander zu vernetzen, nicht nur die Unternehmen der freien Wirtschaft und die kommerziellen Dienstleister, sondern auch die kulturellen Einrichtungen (Museen, Bibliotheken, Archive), die kleineren Colleges und Schulen, die Behörden in den großen Städten und auf dem Lande, die Gesundheitseinrichtungen und die privaten Haushalte. Aus ihren Verheißungen greifen wir diejenigen heraus, die in bezug auf unser Thema besonders relevant sind:

* Die umfangreichen Schätze aus Kunst, Literatur und Wissenschaft könnten überall verfügbar sein, nicht nur in großen Institutionen oder Museen und Bibliotheken der großen Städte.
* Die besten Lehrer und Kurse könnten allen Studenten und Schülern zur Verfügung stehen, unabhängig von der geographischen Lage, räumlichen Entfernungen, ihren Mitteln, oder einer möglichen Behinderung.
* Regierungsstellen und Behörden könnten von Bürgern elektronisch direkt erreichbar sein, ihre Informationen könnten direkt oder durch örtliche Organisationen, wie Bibliotheken, bereitgestellt und elektronisch erhältlich sein.
* Man könnte an vielen Orten leben, ohne Nachteile bezüglich seiner Beschäftigung in Kauf nehmen zu müssen und ohne sich bietende Chancen zu verpassen.
* Die verschiedenen Regierungsstellen und öffentlichen Einrichtungen, die private Wirtschaft und andere organisatorische Einheiten könnten Information elektronisch austauschen und dadurch den Umgang mit Papier reduzieren und ihre Dienstleistung verbessern.

[21] http://www.gsfc.nasa.gov/NASA_homepage.html - NASA
[22] http://www.doe.gov/ - DOE
[23] http://www.whitehouse.gov/ - Das Weiße Haus
[24] http://lcweb.loc.gov/homepage/lchp.html - Library of Congress
[25] http://www.oclc.org/ - OCLC
[26] http://www.ispo.cec.be/infosoc/backg/bangeman.html - Bangemann Report
[27] http://www.whitehouse.gov/White_House/Publications/html/WH-Major-Docs.html#nii - NII Dok.

Hohe Mobilität der Bürger, lebenslanges Lernen, Annahme der großen Herausforderungen in der Forschung und eine leistungsfähigere Administration, die aber weniger kostet, sind Schlüsselworte der amerikanischen Gesellschaft - und jetzt dieses Plans. Regierungsstellen und öffentliche Einrichtungen sollen zum Beispiel Informationen, über die sie verfügen und die sie den Bürgern im Rahmen ihrer Aufgaben ohnehin vermitteln müssen, nicht mit zusätzlichen Kosten belegen, wenn sie diese elektronisch im Netz bereitstellen.

Die neuen Technologien der elektronischen Information und Kommunikation machen nicht nur das gezielte und schnelle Auffinden relevanter Information in riesigen Datenbeständen möglich, sondern sie eröffnen auch völlig neue Chancen zur Modernisierung und zur Reduktion von Kosten, wie wir später auch noch an konkreten Beispielen aus Wissenschaft und Verwaltung zeigen werden.

2. Dokumentation und Archivierung, Bibliotheken

Ein wesentlicher Bestandteil der NII-Initiative ist die Entwicklung von *digital libraries*[28]. Sie werden als nationale Herausforderung betrachtet, der im Rahmen von Forschungsarbeiten des *High Performance Computing and Communication*[29] (HPCC) Programms zu begegnen ist. Bei digitalen Bibliotheken handelt es sich nicht, wie vielleicht der Name suggeriert, nur um die elektronische Bereitstellung von Büchern oder Journalen. Es geht um weit mehr. Dieses Programm soll die enormen Mengen an Wissen und innovativer Information erschließen, die bei den großen nationalen Forschungseinrichtungen und Archiven vorliegt, für Forscher, Studenten und Lehrer im ganzen Land - und für die allgemeine Öffentlichkeit.

2.1 Digitale Bibliotheken in der Forschung

Die NASA verfügt über große Schätze an digitalen Aufnahmen aus ihren verschiedenen Raumfahrtexpeditionen, und ihre die Erde überwachenden Satelliten erzeugen ständig neue und außerordentlich umfangreiche Mengen an wertvollen Daten und Bildern. Diese sollen jedermann zur Verfügung stehen. Dafür benötigt die NASA neuartige Technologien. Das *Digital Library* Projekt der NASA[30] soll zum Beispiel auch das langfristig angelegte *Earth Observation System Program* (EOS) mit Pilotprojekten unterstützen, die demonstrieren, wie Datensammlungen, Dokumente, Broschüren, Newsletter, Bilder und Videofilme *online* im Internet zur Verfügung gestellt werden können. Wissenschaftler, Lehrer und Schüler auf der ganzen Welt erhalten dadurch einen freizügigen, kostenlosen und unmittelbaren Zugriff auf EOS-Information.

Informationsquellen digitaler Bibliotheken sind nicht nur Texte, sondern auch numerische Daten, Visualisierungen, Bilder und Symbole, Ton und Musik, gesprochene Worte und Videos. Diese liegen auf den unterschiedlichsten Trägern vor, auf Papier, Filmen, magnetischen und optischen Speichern. Die Informationsquellen sind möglicherweise geographisch sehr weit

[28]http://farside.gsfc.nasa.gov/ISTO/DLT/nsf_info.html - Digital Libraries (NSF, ARPA, NASA)
[29]http://www.hpcc.gov/ - HPCC Office
[30]http://farside.gsfc.nasa.gov/ISTO/DLT/ - Digital Library Project (NASA)

voneinander entfernt, auch von den Personen, die sie vielleicht nutzen wollen.

Quellen dieser Art nicht nur in den USA, sondern weltweit verfügbar zu machen, ist das Ziel dieses Forschungsprogramms. Hierfür sollen geeignete Technologien zum einfachen benutzerfreundlichen Suchen in verteilten Datenbeständen entwickelt werden, sowie entsprechende Verfahren der (verteilten) elektronischen Archivierung und Dokumentation (Indexierungs-, Klassifizierungs-, Abstrahierungs- und Lokalisierungsverfahren). Um das zu erreichen, haben sich NSF, ARPA und NASA abgestimmt und vergeben dafür insgesamt 24,4 Millionen Dollar an Forschungsgeldern, die auch für den Aufbau von Pilotanwendungen gedacht sind. Wir nennen einige Beispiele (Stand: September 1994).

* Die Universität von Michigan entwickelt ein *testbed* (4 Millionen Dollar) für eine große, sich ständig erweiternde digitale Multimedia-Bibliothek für Geowissenschaften und Raumfahrt, die potentiell Tausende von Benutzern versorgen soll.
* Die Carnegie Mellon Universität entwickelt eine *online digital video library* (4,8 Millionen Dollar), die den Zugriff auf, sowie die Explorierung und das Retrieval von Materialien in Videoarchiven aus Naturwissenschaft und Mathematik ermöglicht. Dieses Archiv wird 1.000 Stunden an Video-Information enthalten.
* Die Universität von Kalifornien in Berkeley entwickelt den Prototyp einer digitalen Bibliothek (4 Millionen Dollar), die auf Umweltdaten spezialisiert und weltweit mit anderen Umweltdatenbanken vernetzt ist. Der Prototyp dieser digitalen Bibliothek soll in das bereits bestehende *California Environmental Resources Evaluation System*[31] (CERES) des Staates Kalifornien, integriert werden.
* Die Universität von Kalifornien in Santa Barbara entwirft eine digitale Bibliothek (4 Millionen Dollar), die Landkarten, Luftaufnahmen und anderes digitalisiertes Bildmaterial aus den Geowissenschaften erschließt. Sie verfügt selbst über eine der größten Sammlungen auf diesem Gebiet.
* Die Stanford Universität wird neue Technologien für eine Integrierte Digitale Bibliothek entwickeln (3,6 Millionen Dollar), die eine große Zahl individueller Informationskollektionen miteinander vernetzt, von personenbezogenen Daten über solche, wie man sie in Bibliotheken findet, bis hin zu großen Datensammlungen im Wissenschaftsbereich.
* Die Universität von Illinois in Urbana-Champaign konzentriert sich auf Journale und Magazine aus Ingenieur- und Naturwissenschaften (4 Millionen Dollar). Ihr *digital library testbed* wird auf der Basis von Mosaic das Modell einer universitären *production facility* (Print-Zentrum) implementieren.

Uns in Deutschland erscheint dies alles wie Zukunftsmusik. Es ist ja nur Forschung, mag sich mancher sagen - und wie steht es mit der Realität? Müssen denn die Datenautobahnen nicht erst noch gebaut werden? Aber, die Entwicklung digitaler Bibliotheken wird in den USA schon seit langem vorangetrieben, auch im Rahmen anderer Förderprogramme und eigenständiger Forschungs- und Entwicklungsvorhaben. Einen knappen Überblick über das breite Spektrum dieser Aktivitäten vermittelt die Spezialausgabe der *Communications of the ACM* der *Association for Computing Machinery*[32] [ACM 95] [FoxAFL 95].

[31] http://resources.agency.ca.gov/index.html - CERES
[32] http://www.acm.org/ - ACM

2.2 Digitale Bibliotheken im öffentlichen Bereich

Die Hauptmasse von Archivalien besteht aus Urkunden und Akten, die der Tätigkeit von Regierungsstellen, Behörden und Gerichten entstammen, so sagt es der Wegweiser für Berufe im Archiv-, Bibliotheks-, Informations- und Dokumentationswesen [Gaus 94]. Hier in Deutschland kommt es uns zur Zeit abenteuerlich vor, bei Behörden an E-mail-Korrespondenz und den Einsatz anderer elektronischer Mittel zu denken. Das wird sich mit dem Aufbau der Datenautobahnen in einigen Bundesländern[33] (Baden-Württemberg, Bayern, Berlin, Norddeutschland, Nordrhein-Westfalen) vielleicht bald ändern. Zur Zeit ist es jedoch noch nicht soweit, denn in Deutschland verfügt außer den Forschungsinstitutionen kaum eine öffentliche Einrichtung über einen Internet-Anschluß.

Wenn wir die Fragen der Zukunft diskutieren wollen, müssen wir also (leider) wieder in die Vereinigten Staaten schauen. Bevor wir zum Thema der Archivierung kommen, nennen wir einige Kennzahlen für die Durchdringung der öffentlichen Einrichtungen und Behörden in den USA durch das Internet. Sie können die auch hier bald zu erwartende Wirklichkeit vielleicht besser widerspiegeln, als die Diskussion eines Forschungsprogramms.

Von den rund 400 in der *WWW Virtual Library* genannten U.S. Departments, Regierungsstellen und Agenturen hatten im August 1995 über 330 eigene Web-Server aufgebaut[34], bis auf das *Department of Education*, alle anderen großen Departments, jeweils mit fast allen ihren Abteilungen (in Deutschland ist immerhin das BMBF[35] im Web vertreten). Ein für das Rechtswesen besonders interessantes Beispiel: Die signifikanten Beschlüsse des U.S. Supreme Courts sind 15 Minuten nach der Urteilsverkündigung für jeden Bürger einsehbar im Web zu finden[36]. Entscheidungen von 1989 bis zur Gegenwart stehen hier bereit.

Praktisch alle wissenschaftlichen Universitätsbibliotheken (die *Association of Research Libraries*[37] Nordamerikas zählt rund 110 Mitglieder) haben eigene Server im Web errichtet[38]. Die meisten von ihnen bieten ihre Kataloge online recherchierbar im Internet an, wie z.B. die *Library of Congress*[39] (LOC). Sie führt einige ihrer Ausstellungen auch online im Internet durch. *Rome Reborn: The Vatican Library and Renaissance Culture* ist nur das bekannteste Beispiel. Sie zeigt historische Dokumente aus der Bibliothek des Vatikans[40] . Die Mathematik ist dabei durch so bekannte Klassiker wie Euklid, Archimedes und Ptolemäus vertreten. Die Ausstellung selbst ist schon lange geschlossen (April 93), die online-Ausstellung ist auf unbestimmte Zeit im Internet zu sehen. Inzwischen ist auch die Vatikanstadt mit einigen Glanzstükken im Netz zu finden - vertreten durch eine private Organisation[41].

Die WWW Virtual Library nennt in den USA 130 Museen[42] mit eigenen Web-Servern,

[33] http://www.dfn.de/~ent/home.html - DFN Entwicklungsbereich
[34] http://www.lib.lsu.edu/gov/fedgov.html - U.S. Federal Government Agencies
[35] http://www.dfn.de/bmbf/home.html - BMBF
[36] http://www.law.cornell.edu/supct/ - Entscheidungen des U.S. Supreme Court
[37] http://arl.cni.org/ - ARL
[38] http://www.lib.washington.edu/~tdowling/arl.html - Mitglieder der ARL
[39] gopher://marvel.loc.gov/11/global/math - LOC Mathematics (Marvel)
[40] ftp://ftp.loc.gov/pub/exhibit.images/vatican.exhibit - Vatican Exhibit (LOC)
[41] http://www.christusrex.org/ - Vatikanstadt Museum
[42] http://www.comlab.ox.ac.uk/archive/other/museums.html - WWW Virtual Library: Museums

wie z.B. die Bibliothek der Universität von Michigan mit ihrem *Papyrus Digitization Project*[43] (in Deutschland sind immerhin 10 Bibliotheken vertreten, darunter die der Universität Augsburg mit einer Dokumentation der wunderschönen Öttingen-Wallerstein-Kollektion[44]). Fast täglich kommen neue hinzu, und solche Zusammenstellungen sind bekanntlich niemals vollständig.

Die NASA stellt schon seit einiger Zeit größere Mengen ihres Bildmaterials[45] [46] [47] im Internet bereit, darunter brilliante Photoaufnahmen des Hubble-Teleskops[48]. Die Kollision des Shoemaker-Levy9-Kometen mit dem Jupiter konnten Tausende Fachleute und interessierte Laien online verfolgen[49]. Für Lehrer und Schulen stellt die NASA in ihrem SpaceLink-Projekt[50] ausführliche Unterrichtsmaterialien bereit, so zu den aktuellen Planungen und Phasen ihres Erdbeobachtungsprogramms (EOS). Auch einige europäische Astronomie- und Raumfahrtorganisationen bieten inzwischen ähnliche Dokumente und Bilder an, darunter die *European Space Agency*[51] (ESA) und die *Space Research Organization*[52] (SRON) in Holland.

Die *National Archive and Record Administration* (NARA) betreibt nicht nur ein online-Informationssystem[53] (CLIO) im Internet, sondern hat dort auch eine virtuelle Ausstellung im Web aufgebaut, die Schlüsseldokumente aus der amerikanischen Geschichte zeigt, wie die *Declaration of Independence* in Original- und in der in Stein gravierten Version. Natürlich erhält man nur Abbildungen, jedoch solche von hoher Qualität, und man kann sie auf seinem eigenen Computer weiterverarbeiten. Die Originale sind für den amerikanischen Staat derart wertvoll, daß - bis auf einige auserlesene Personen - niemand sie berühren darf. Über das Web stehen sie weltweit zum Kopieren bereit. Die Bereitstellung solcher Informationen ist nicht nur für Schulen von großem Wert.

"Information ist die Währung einer freien Gesellschaft", stellt die NARA in ihrem *mission statement* fest, das ebenfalls im Internet erhältlich ist. Sie steht mit ihren Aufgaben der Alltagspraxis des Archivars und Dokumentars, wie wir ihn in Deutschland kennen, vielleicht noch am nächsten. Sehen wir, wie sich die NARA auf die Zukunft vorbereitet. Denn daß die Zukunft der Informationsgesellschaft bereits begonnen hat, ist inzwischen klar, zumindest in den Vereinigten Staaten von Amerika.

Die NARA archiviert und konserviert Dokumente des U.S. Government und organisiert Aufzeichnungen, die Handlungen der amerikanischen Regierung dokumentieren. Sie stellt sicher, daß diese Information dem amerikanischen Bürger jederzeit zur Verfügung steht. Das Spektrum des dokumentarischen Materials ist außerordentlich breit. Es umfaßt von Beratungen und Entscheidungen der höchsten Gremien alles bis hin zu einer Vielfalt von individuellen dienstlichen Informationen, auch medizinische und finanzielle Daten. Die NARA archiviert

[43] http://www.lib.umich.edu/pap/HomePage.html - Papyrology (Univ. Michigan)
[44] http://hpwibas2.bibliothek.uni-augsburg.de:8001/ - Oettingen-Wallerstein Kollektion
[45] ftp://explorer.arc.nasa.gov/pub/SPACE/ - NASA File Explorer
[46] http://southport.jpl.nasa.gov/ - Radar Images (NASA, JPL)
[47] http://skview.gsfc.nasa.gov/skyview.html - SkyView (NASA)
[48] http://www.stsci.edu/pubinfo/BestOfHST95.html - Hubble Space Telescope Gallery
[49] http://seds.lpl.arizona.edu/sl9/sl9.html - Shoemaker-Levy 9
[50] http://spacelink.msfc.nasa.gov/ - NASA Spacelink
[51] http://www.esrin.esa.it/ - ESA
[52] http://ruurq2.sron.ruu.nl/ - SRON
[53] http://www.nara.gov/ - NARA (CLIO)

Loseblattsammlungen und gebundene Dokumente (Milliarden Einzelexemplare), Fotografien und Landkarten (Millionen), Filme, Videobänder, Tonaufnahmen und Schallplatten (Tausende) - und elektronisches Material; Computerbänder erst ab 1993.

Die personellen und finanziellen Ressourcen der NARA werden kontinuierlich knapper, ihre Aufgaben und Verantwortlichkeiten nehmen hingegen zu. Sie wird höhere Leistungen zu geringeren Kosten erbringen müssen - einem Leitsatz der NII-Initiative gemäß. Durch Reorganisation will sie Einsparungen erbringen (auch durch Reduzierung des administrativen Overheads). Diese will sie zum Teil reinvestieren, in die Ausbildung des Personal, ihres wertvollsten "Kapitals", und in Modernisierung durch Einsatz elektronischer Mittel. Dabei will sie insbesondere die Chancen nutzen, die sich durch die jetzt in Erscheinung tretenden neuen Technologien eröffnen.

Die Art und Weise, in der Dokumente geschaffen und gebraucht werden, ändert sich in dieser Zeit. Die erste und grundlegende Frage, stellt die NARA fest, besteht darin, welche Information zu konservieren ist. Ihre zweite Herausforderung liegt darin, einen besseren Zugriff auf ihre Sammlungen zu ermöglichen. Das sind komplexe Fragen, die wir hier nicht im Detail erörtern können. Die NARA will sich darauf spezialisieren, ihre Daten elektronisch zu führen.

Es geht ihr um die Entwicklung eines elektronischen NARA Informationssystems (NIS). Dieses soll nach dem strategischen Plan der NARA den Zugriff auf ihre Daten und Dokumente *(records)* ermöglichen. Zu den Bestandteilen des NIS gehören Elemente wie ein nutzerfreundliches öffentliches Zugriffssystem, ein *online agency records schedule*, das den Zugriff auf Informationen über die Aktualisierung der Bestände erlaubt, und auch über die am häufigsten über das Internet angefragten Dokumente. Das NIS soll im Rahmen des *Government Information Locator Service*-Projekts (GILS) mit entsprechenden Informationssystemen der amerikanischen Regierung vernetzt werden. Die NARA plant auch weitere Internet-Ausstellungen.

Government Information ist für moderne Gesellschaften fundamental, besagt das *mission statement* des GILS-Projekts. Es entwickelt eine kollektive Vision für alle Stellen und Agenturen des Government. Als Teil der besonderen Rolle der amerikanischen Administration in der NII Initiative - die Regierung übernimmt eine Führungsposition - wird GILS die öffentlichen Informationsressourcen der gesamten föderativen Struktur ermitteln und beschreiben - und der Öffentlichkeit über das Internet zur Verfügung stellen. Informationen der Regierung sollen dem Bürger zu fairen Preisen zur Verfügung stehen. Den Internet-Zugriff auf den Kern des GILS-Systems, also auf die Daten eines Systems von Verweisen *(Locator)*, das den Zugriff auf die bei vielen Regierungsstellen vorliegende Information organisiert, soll keine Regierungsstelle oder Agentur mit Gebühren belegen.

Statt eines zentralen Designs verfolgt GILS[54] einen dezentralisierten Ansatz, bei dem die verschiedenen Institutionen unabhängig voneinander arbeiten und dabei die unterschiedlichsten Systeme einsetzen, die dennoch miteinander kommunizieren und kooperieren können *(Interoperabilität)*. Der Erfolg von GILS wird weniger von einer massiv zentralisierten Steuerung abhängen, als vielmehr von der freiwilligen Beteiligung talentierter und engagierter

[54] http://www.usgs.gov/public/gils/gilsdoc.html - GILS Dokumentation

Mitarbeiter in den Verwaltungen. Der dafür notwendige Mehraufwand wird als gering eingeschätzt, denn der hauptsächliche Teil der Arbeit muß ohnehin innerhalb der normalen Routinetätigkeit erledigt werden. Für diese hat der Steuerzahler aber bereits bezahlt.

Die amerikanische Gesellschaft entwirft eine Vision, die für uns noch in weiter Ferne zu liegen scheint. Doch dieser Eindruck täuscht. Die Europäische Union hat ein ähnliches Programm beschlossen, und hierzulande entstehen in bestimmten Regionen ebenfalls Datenautobahnen, die in ihrem Anspruch den amerikanischen durchaus vergleichbar sind. Wenn wir auch in den öffentlichen und in vielen anderen Bereichen noch weit zurückliegen, ist in Deutschland und in ganz Europa doch ein funktionsfähiges Wissenschaftsnetz entstanden. Und, auch bei uns ist der allgemeine Kostendruck so hoch, daß uns der Übergang zu modernen Technologien und damit zu effizienteren und gleichzeitig kostengünstigeren Verfahren geradezu zwingend erscheint. Wir diskutieren diese Frage in den nächsten Kapiteln im Bereich des Publizierens und der wissenschaftlichen Kommunikation. Mit dem Übergang in die elektronische Welt treten aber völlig neuartige Fragen auf, die wir zunächst am Beispiel der langfristigen Archivierung darstellen.

2.3 Archivierung und Konservierung in der elektronischen Welt

Schon die Verwendung von E-mail und die elektronische Diskussion werfen gewisse Probleme auf. Wenn Wissenschaftler nur noch elektronisch miteinander kommunizieren, wird es nur noch sehr wenige Dokumente im traditionellen Sinne geben. Wir hätten dann wohl auch keine Handschriften mehr, vermutlich auch nicht vom "Gauß unserer Zeit". Auch die ersten Entwürfe von Artikeln und Diskussionen mit Fachkollegen sind dann nur noch in irgendwelchen Textsystemen und Mailboxen gespeichert. Welcher Mathematiker wagt denn heute daran zu denken, daß seine Arbeit vielleicht einmal unsterblich sein könnte - und sorgt beizeiten für die langfristige Dokumentation seiner (ersten) Ideen und Entwürfe? Bei Gauß fielen diese Papiere quasi nebenher an. Auch er dachte vermutlich in erster Linie an seine Mathematik und nicht an spätere Historiker, denen er die Arbeit leichter machen könnte. Langfristige Archivierung ist, wie wir meinen, keine technische Frage, denn die technischen Probleme sind lösbar, wenn man sich richtig verhält. Personen, die an die Zukunft denken, sollten z.B. keine isolierten Text- oder Datensysteme verwenden, sondern langfristig tragfähige Speichermethoden.

Die Kernfragen hat die NARA gestellt: Wer entscheidet, was langfristig zu bewahren ist? Wie könnte man solche Dinge identifizieren? Was ist in der elektronischen Welt überhaupt ein Original? Die mit der Hand niedergelegte Schrift hilft ja nicht mehr beim Identifizieren, und eine E-mail wird auch nicht mit der Hand unterschrieben. Diese Überlegungen gelten nicht nur für einzelne Personen, sondern für ganze Institutionen. Wenn die amerikanische Administration bald (nur noch?) elektronisch kommunizieren will, welche Rechtskraft haben solche Dokumente? Wäre es dann nicht bald widersinnig, die Rechtsverbindlichkeit nur dadurch herzustellen, daß man parallel Briefe auf Papier verschickt? Bereits wir kennen ja halbwegs elektronische Dokumente mit Rechtsverbindlichkeit (Bewilligungen von Fördervorhaben, Steuerbescheide vom Finanzamt). Bedeutet das alles den langsamen Tod der (späteren) Quellenforschung der Historiker? Vermutlich ist das der Fall, wenn wir nicht rechtzeitig Vorsorge treffen. Sind sich die Historiker dessen bewußt? Wir haben es mit der zentralen

Frage der langfristigen Dokumentation zu tun, einer Aufgabe für Archivare und Dokumentationsfachleute. Damit kommen wir zur zweiten Kernfrage der NARA. Auf welche Weise stellt man Information überhaupt langfristig bereit?

In der Welt der Computer entwickeln sich die Dinge mit unglaublicher Geschwindigkeit. Was gestern noch als tragbares Mittel erschien, ist morgen schon überholt. Wer Dauer und Langlebigkeit in der Hardware sucht, ist bald genarrt. Ein Beispiel aus der jüngsten Vergangenheit sind die CD-ROMs. Aus dem Bereich der Bibliotheken hört man oft die bange Frage, ob denn eine CD-ROM in 50 Jahren noch lesbar sein wird? Man denkt dabei an Chemikalien (des aufgedruckten Labels), die vielleicht in die Silberscheibe einwandern. Niemand kann heute genaueres dazu sagen. Wir meinen aber, daß die Frage falsch gestellt ist. Schon sehr bald, vielleicht bereits morgen, wird es leistungsfähigere CD-ROM-Speicher geben, die die heutigen Technologien auf Dauer ablösen. Schon in wenigen Jahren wird man vermutlich Laufwerke der heutigen Bauart überhaupt nicht einmal mehr kaufen können - und, wer kann denn sagen, wie lange die heutigen noch laufen werden? Wie lange wird man sie überhaupt reparieren können? Und, wenn man sie langfristig erhalten könnte, lassen sie sich dann noch an modernere Hardware anschließen? Modernisieren muß man aber beizeiten, wenn man nicht auf alten, überholten und immer zu teuren Geräten sitzen bleiben will.

Informationsträger haben unterschiedliche Lebensdauern, aber sie sind niemals von Dauer. Gebrannte Tontafeln können tausende von Jahren überstehen, aber bereits die Bibliothek von Alexandria mußte Abschreiber beschäftigen, um dem Verfall von Papyrus zuvorzukommen. Über die Lebensdauer moderner elektronischer Informationsträger wissen wir nicht viel. Wir müssen daher dafür sorgen, daß die Information selbst von Dauer ist. Die Stimme von Caruso kann man heute noch hören, obwohl die alten Schellackplatten längst nicht mehr (allgemein) zur Verfügung stehen. Seine Stimme aber lebt auf modernen Tonträgern weiter, in vielfältiger Form und an vielen Orten der Welt.

Sprach-, Musik- und Videoinformation (*Live*-Aufnahmen) haben - im Gegensatz zu geschriebenen Dokumenten, ob in Stein gemeißelt oder Ton gebrannt, ob auf Papyrus, Pergament oder Papier geschrieben, - eine weitere Charakteristik, die erst bei Dokumenten unserer Zeit auftritt. Zu ihnen gibt es kein Original im traditionellen Sinn. Die Künstler unserer Zeit wissen schon seit langem darum. Viele ihrer Werke sind “nur” noch bloße Ideen. Nur die Information selbst “ist” das Original (nichts als Schallwellen, elektronische Schwingungen oder digitale Sequenzen). Aber, wenn sie uns nur wertvoll genug erscheint, wird sie auf ewig lebendig sein. Es kommt darauf an, sie beizeiten auf die neuen Träger zu kopieren, und die wichtigsten Dinge möglichst weltweit zu verbreiten.

Langfristige Archivierung bedeutet, rein technisch gesehen, daß man immer darauf achten muß, jederzeit zur neuesten Speichertechnologie wechseln können, nicht nur aus Kostengründen. Das sollte - manchem mag das paradox erscheinen - durch die Verwendung von “flüchtigen”, wiederbeschreibbaren Medien immer möglich sein. Wer dabei zwischenzeitlich auch “dauerhafte” Träger einsetzen will, hat damit in der Regel weniger ein technisches Problem, solange er nicht den Träger selbst als “die Lösung” ansieht (auch eine CD-ROM ist kopierbar), sondern vielmehr ein rechtliches. Er muß darauf achten, daß er sich das Recht zum Kopieren verschafft. Bei CD-ROMs zum Beispiel ist das normalerweise nicht gegeben.

Die Verwendung von wiederbeschreibbaren Speichermedien allein löst allerdings noch nicht das Problem. Wenn eine Bibliothek elektronische Dokumente (Journale oder ähnliches) als Bestandteil eines *document delivery* Systems ankauft, wird sie vermutlich ähnliche Sorgen bekommen. Im ungünstigsten Fall ist sie dann nicht einmal mehr in der Auswahl ihrer Hardware frei, denn *document delivery* Systeme sind zumeist hoch spezialisiert und in der Anwendung nicht offen konzipiert, sondern eher abgeschlossen. Sie liefern, wenn man die Information als integrierten Bestandteil kauft, auch keine konvertierbaren Texte aus, sondern Grafiken oder Pixelbilder in speziellen Formaten, die nur mit Hilfe des jeweiligen Systems selbst inspiziert und verwaltet werden können. Welcher Verleger achtet denn nicht auf "sein" Copyright. Und, wie lange hat solch ein Dokumentensystem Bestand? Die Entwicklung von Software verläuft mit noch größerer Dynamik als die der Hardware. Wer wollte denn garantieren, daß Entwickler solcher Software - sie sind selbst keine Verleger - besonders lange "überleben".

Offenheit und Kompatibilität bezüglich der Daten sind die Schlüsselkriterien für Archivare, die an langfristige Bestandserhaltung denken. Wenn sie diese sichern können, sind sie unter bestimmten Bedingungen auch in der Wahl ihrer Software und Systeme frei (Datenbank, Retrievalsystem oder ein System zur Dokumentenverwaltung). Hier ist das entscheidende Stichwort *Interoperabilität*. Solche Software muß mit vielen anderen Komponenten zusammenarbeiten können, auch über größere Entfernungen hinweg. Sie muß heute insbesondere bezüglich des World Wide Web offen sein. Sofern diese Art der Flexibilität gegeben ist, reduziert sich die langfristige Archivierung auf die Verwendung bestimmter Datenformate.

Diese Frage ist, wenn man sie langfristig betrachtet, vielleicht noch ein Problem. Man denke nur an mathematischen Satz, also Formeln oder Grafiken. Wer würde denn behaupten wollen, daß LaTeX- oder Postscript-Dokumente von heute sich noch in 100 Jahren von Software und Laserdruckern verarbeiten lassen? Gibt es dann überhaupt noch Laserdrucker, die Postscript verarbeiten können? Schon heute kommt *PDF* ins Gespräch. Für die langfristige Speicherung strukturierter Dokumente kommen aus heutiger Sicht (der Fachleute in den Verlagen) eigentlich nur *SGML* [Goldfarb 90] oder eine seiner Varianten für multimediale Information in Betracht. In der Mathematik wird jedoch überwiegend nur TeX oder LaTeX eingesetzt, ebenso in der Physik und in der Informatik. Was ist zu tun?

Auch die Frage nach "dem universellen Datenformat", das langfristig Bestand haben soll, erscheint uns falsch gestellt. Man kann sie nämlich in dieser sich ständig verändernden elektronischen Welt niemals definitiv beantworten, heute nicht und auch in Zukunft nicht. Zu jedem bestimmten Datenformat gehört nicht nur ein ganzer Komplex von Software (bis hin zum Betriebssystem und gewissen Treibern), sondern immer auch eine bestimmte Hardware, die es abbilden kann (Laserdrucker, Bildschirm etc.). Entscheidend ist, daß man - auch hier - jederzeit in der Lage sein kann und muß, auf die neueren Techniken zu wechseln.

Wir gehen davon aus, daß im Zuge der Akzeptanz und der weltweiten Verbreitung der im Internet verwendeten Formate, zu denen es ja für jedes offene - kommunikationsfähige - System Abbildungen gibt, immer auch entsprechende Konvertierungsprogramme entstehen, die es gestatten, Daten aus überholten Formaten in aktuellere umzusetzen. Daran sind im Internet immer viele Menschen interessiert, die darauf achten, nicht von speziellen Hardwareherstellern und Softwarehäusern abhängig zu werden. So betrachtet, wird das "Problem", wenn man wachsam ist und rechtzeitig zu neueren Verfahren wechselt, immer lösbar sein - durch einfaches Umsetzen und Umkopieren. Die Stimme von Caruso liefert den Beweis.

2.4 Wachstum wissenschaftlicher Literatur, Krise der Bibliotheken

Die wissenschaftliche Literaturproduktion verdoppelt sich alle 16 Jahre [AAU/ARL 94] [CummingsWBLE 92], in den Naturwissenschaften und in der Mathematik schon alle 10 bis 12 Jahre [Odlyzko 95]. Dieses Wachstum ist schon seit über 100 Jahren sichtbar [Wiiërs 94], und es wird natürlich einmal enden, allerdings - so meinen wir - noch nicht in naher Zukunft. Die Gründe hierfür sind erhöhte "Produktivität" der einzelnen Forscher und ein stetiges Anwachsen der Zahl der Wissenschaftler. Bei den Mathematikern liegt es auch an der zunehmenden Ausstrahlung in andere Wissenschaftsbereiche und den verstärkten Aktivitäten interdisziplinärer Art.

Für die großen wissenschaftlichen Bibliotheken, die den Anspruch haben, die gesamte relevante Literatur vor Ort auf Papier zur Verfügung zu stellen, stellt dieser Umstand ein ernstes Problem dar, denn ihre Etats stagnieren oder gehen sogar zurück. Die Kosten für Abonnements wissenschaftlicher Journale hingegen steigen, sogar weit über die Inflationsrate hinaus [AAU/ARL 94]; für die Situation in der Mathematik in Deutschland siehe [WinterZ 95]. Die Etats der Bibliotheken müßten aber mit demselben Tempo wachsen wie die Produktion von Literatur und wie die Kosten für ihren Ankauf.

Die wissenschaftlichen Bibliotheken sehen derzeit noch den Ausweg, sich mittels elektronischer Dokumentenlieferung zu behelfen, in Deutschland im Rahmen des SUBITO-Projekts [DBI 95]. Dieses scheint uns - langfristig gesehen - kein wirklicher Ausweg zu sein. Wie wir in der Arbeit "Wissenschaftliche Kommunikation am Wendepunkt - Bibliotheken im Zeitalter globaler elektronischer Netze" ausführlicher dargelegt haben [GrötschelL 95b], ist Dokumentenlieferung nicht nur um Größenordnungen teurer, sondern sie kommt für die Forschung ebenfalls viel zu spät, weil sie dem traditionellen Rahmen - dem Papier - verhaftet bleibt. Es können nur Dokumente geliefert werden, die auf Papier vorliegen. Das mag für ältere oder seltene Dokumente noch vertretbar zu sein. Sieht man aber elektronische Dokumentenlieferung als das Mittel der Wahl für die Bereitstellung der großen Menge an Literatur an, verbirgt sich darin ein Teufelskreis.

Wenn die Hälfte der Bibliotheken ein für sie nicht mehr bezahlbares Journal einfach abbestellt und darauf vertraut, daß man einzelne Artikel bei Bedarf mittels elektronischer Dokumentenlieferung besorgen kann, wird der Verleger, der dieses Journal herstellt, nur noch mit der Hälfte der Abonnenten rechnen können. Auf diese muß er nun seine Herstellungskosten umlegen. Da aber ein großer Teil der Kosten unabhängig von der Auflage ist, müßten die verbleibenden Abonnements nun, grob gesagt, das Doppelte kosten, immer vorausgesetzt, der Verleger will sein Einkommen wahren. Wäre es anders, dann würde es ja genügen, wenn die Gemeinschaft der Bibliotheken insgesamt nur noch ein einziges Exemplar bestellt. Was müßte dieses Exemplar dann kosten?

A. Odlyzko sieht in einem richtungsweisenden Artikel "Tragic Loss or Good Riddance? The Impending Demise of Traditional Schorlarly Journals", der auch uns erheblich beeinflußt hat, den nahenden Tod der traditionellen wissenschaftlichen Journale voraus [Odlyzko 95], irgendwann in den nächsten 10 bis 20 Jahren. Das gesamte System des bisherigen traditionellen Publizierens wissenschaftlicher Ergebnisse (auf Papier) ist seiner Meinung nach äußerst schwerfällig und nicht mehr zeitgemäß, und es wird in Kürze vollständig zusammenbrechen. Schon heute treten in der Mathematik beklemmend hohe Ablehnungsraten (bis zu 80%) und exorbitant lange Wartezeiten (2 bis 3 Jahre) auf, und das oft nur, weil Zeitschriften (aus Kosten- und Marketinggründen) die Seitenzahlen beschränken. Die ACM sieht das gegenwärtige Verfahren des Publizierens durch eine Reihe von Zusammenbrüchen in Frage gestellt [DenningR 95]. Sie plant deshalb - mit allen ihren Journalen und Reihen - den vollständigen Übergang in die elektronische Welt.

Das enorme Wachstum der wissenschaftlichen Literatur wird - jedenfalls in der Mathematik - dazu führen, daß in den nächsten 10 bis 12 Jahren etwa so viele Reports geschrieben werden, wie in der Geschichte der Mathematik bisher insgesamt. In 20 bis 24 Jahren wird sich die Anzahl der mathematischen Veröffentlichungen bereits vervierfacht haben. Wir sind mit Odlyzko der Meinung, daß dieser Umstand den Übergang zum elektronischen Publikationswesen zur notwendigen Folge hat und haben muß [GrötschelL 95b], [GrötschelLZ 95]. Wohin denn sonst mit dem Papier? Und, wie anders sollte die entstehende "Publikationslücke" zu schließen sein? Der Markt des traditionellen wissenschaftlichen Publikationswesens ist schon heute nicht mehr in der Lage, sich den Notwendigkeiten entsprechend weiter zu entwickeln. Ständig werden zwar neue Journale gegründet, diese finden jedoch keine Abnehmer mehr.

Wir schlagen, kurz gesagt, vor, daß die Wissenschaften ihre Angelegenheiten selbst in die Hände nehmen und sich dabei auch der Unterstützung durch Bibliotheken bedienen, ihrer ureigensten Infrastruktur. Auch Fachinformationszentren können hier eine wichtige Rolle spielen, wenn sie den Wandel begreifen, flexibel reagieren und wenn die Politik mitspielt.

Die wissenschaftlichen Veröffentlichungen liegen heute immer schon elektronisch vor, in der einen oder anderen Form, und man könnte und sollte sie elektronisch verteilen, sobald sie erstellt sind. Lange Wartezeiten sind heute nicht mehr notwendig. Mit elektronischen Mitteln läßt sich auch der fast in weite Ferne gerückte "Traum" wieder realisieren, alles vor Ort zur Verfügung zu haben, wie wir im folgenden zeigen wollen. Dieses mag den Verfechtern des traditionellen Publizierens auf Papier heute noch wie eine bittere Ironie erscheinen. Vielleicht schon morgen werden auch sie darin ihre eigene Chance sehen.

3. Verteilung mathematischer Information

In den guten alten Zeiten war nicht daran zu denken, die teuren Großrechner für den Versand von Reports einzusetzen, man hätte sie mißbraucht. Diese Art zu denken, ist leider auch heute noch recht weit verbreitet, obwohl die Zeiten sich drastisch geändert haben. Heute sind Computer im Zuge der elektronischen Revolution außerordentlich leistungsfähig und preiswert geworden, ebenso die digitalen Netze, die jetzt mit damals kaum vorstellbaren Bandbreiten arbeiten. Heute wird hingegen die menschliche Arbeitskraft als wertvoller betrachtet, und es gilt, diese optimal einzusetzen, zum Beispiel dadurch, daß man lange Wartezeiten vermeidet.

Wer in der internationalen Spitzenforschung mithalten will, muß sich dem hier vorherrschenden Tempo anpassen, sonst läuft er Gefahr, daß seine Arbeit entwertet wird, weil ihn die Konkurrenz überholt. Langes Warten auf die Veröffentlichungen der konkurrierenden Wissenschaftler in traditionellen Journalen kommt nicht in Betracht. In der Mathematik dauert es nicht selten zwei bis drei Jahre, bis ein neues Ergebnis endlich gedruckt erscheint. Das ist für die Forschung viel zu spät. In der Mathematik behelfen sich die Wissenschaftler deshalb schon seit Jahren mit anderen Mitteln, vor allem auch dadurch, daß sie ihre Manuskripte und Preprints schon vorher austauschen. Welche Rolle hier das Internet und insbesondere E-mail spielt, haben wir bereits skizziert. Daß sich durch die neuen elektronischen Mittel obendrein erhebliche Kosteneinsparungen ergeben können und wie man das technisch gestalten kann, ist Gegenstand der folgenden Abschnitte.

3.1 Elektronische Verteilung wissenschaftlicher Reports, Kostenvergleich

Ein mathematisches Institut wie das Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB) produziert im Mittel 50 wissenschaftliche Reports im Jahr und verschickt diese jeweils an etwa 300 Kooperationspartner. Für die konventionelle Vervielfältigung und den Versand gibt das ZIB dafür jährlich über 125.000,- DM aus (Exemplare für eigene Zwecke nicht mitgerechnet).

Elektronisch nehmen diese Reports nur sehr wenige Ressourcen in Anspruch. Wir gehen bei unseren Überlegungen der Einfachheit halber davon aus, daß ein mathematischer Bericht im Mittel 200 *Kilobyte* belegt, etwa 60 Kilobyte für die LaTeX- und 140 Kilobyte für Grafiken und Bilder in Postscript-Form *(EPS-Files)*. Auf eine handelsübliche Platte von 4 *Gigabyte* Speicherkapazität (Anfang 1995 kostete sie im Workstationbereich 4.000,- DM) passen also 20.000 Reports. Die Speicherung eines Reports kostet damit überschlagsmäßig 0,20 DM. Wenn man eine leistungsfähige Workstation mit Plattenspeicher, Software und Wartung ansetzt (sie kostet insgesamt circa 24.000,- DM) und über drei Jahre (also mit 8.000,- DM pro Jahr) anteilig abschreibt, kommt man auf einen Betrag von 0,40 DM pro Exemplar und Jahr, also auf 20,- DM anteilige Kosten im Jahr für die Speicherung der Reportproduktion eines Jahres und auf 120,- DM für einen Zeitraum von drei Jahren (20,- DM im ersten, 40,- DM in zweiten und 60,- DM im dritten Jahr); vergleichbare Kostenrechnungen finden sich in [Lügger 95] und [Rehmann 95].

Solch eine Abschätzung ist natürlich etwas "über den Daumen gepeilt", dennoch gibt sie die Größenordnung korrekt wieder. Die Rechnung macht auch nur dann Sinn, wenn man die Workstation wie im ZIB routinemäßig anderweitig einsetzt, denn sie wird durch die Speicherung von Reports ja nur zu einem ganz kleinen Anteil in Anspruch genommen. Was kosten schon ein paar Kopierbefehle? Wir haben deshalb die Workstation mit demselben Kostenverhältnis wie den Plattenspeicher angesetzt. Wir werden aber in den nächsten Kapiteln ein Modell für die Verteilung von Preprints diskutieren, bei dem das Rechnersystem voll veranschlagt wird, sowohl für die Verteilung als auch für die Bereitstellung und das Retrieval.

Wie hoch ist nun der Kostenanteil für den Versand über das Netz? Ist er nicht zu hoch? Der Anschluß des ZIB an das Wissenschaftsnetz (WiN) und das Internet überträgt *2 Megabit* pro Sekunde. Rechnet man, um den Netzanschluß nicht hundertprozentig auszulasten und wegen eventueller Ausfallzeiten, mit nur einem *Megabit* echter Übertragungsleistung und zieht obendrein noch 20% für *Overhead* im Netz ab (Verbindungsaufbau, Kommunikationsprotokoll und auch mögliche Netzausfälle), dann liefert der Anschluß eine echte Übertragungsleistung von 100 Kilobyte pro Sekunde. Das ist sicherlich sehr konservativ kalkuliert. Wir liegen damit aber auf der sicheren Seite. Die wirkliche Übertragung wird natürlich länger dauern, weil das Netz auch anderweitig belastet ist. Uns kommt es hier also nicht auf die Wartezeit des Benutzers, sondern nur auf die Inanspruchnahme des Netzes an. Wir sprechen deshalb von Netzsekunden und meinen damit, daß die Übertragung eines Reports den Anschluß des ZIB mit 2 Sekunden echter Netzleistung belastet.

Der 2-Megabit-Anschluß an das WiN und Internet kostet das ZIB gemäß den geltenden Tarifen des *Deutschen Forschungsnetzes* (DFN) zur Zeit 410.000,- DM im Jahr, incl. MwSt. (gemäß der Einstufung des ZIB in die Kapazitätsklasse 3 und Volumenklasse IV des DFN) [DFN 95]. Da das Netz rund um die Uhr zur Verfügung steht (eventuelle Ausfallzeiten sind ja oben bereits verrechnet) und ein Jahr 31.536.000 Sekunden hat, kostet eine Netzsekunde 0,013 DM. Eigentlich kostet sie das ZIB sogar nur die Hälfte, denn mit dem oben genannten Jahresbetrag ist der Transfer in beide Richtungen abgegolten, sowohl (2 Megabit pro Sekunde) aus dem lokalen Netz des ZIB heraus, als auch (ebenfalls 2 Megabit pro Sekunde) in das lokale Netz hinein. Da man aber keinen "halben" Anschluß mieten kann, rechnen wir mit dieser Größe weiter und können außerdem annehmen, daß damit die beim Empfänger anfallenden Netzkosten ebenfalls berücksichtigt sind.

Der Versand eines ganzen Reports in LaTeX- und Postscript-Form kostet damit weniger als 3 Pfennige. Alle 50 Reports (10 Megabyte) lassen sich in nur 100 Netzsekunden übermitteln. Der Versand an alle 300 Kooperationspartner (3 Gigabyte) nimmt damit 30.000 Netzsekunden in Anspruch. Der Versand aller Reports schlägt also insgesamt mit 390,- DM zu Buche.

In der Summe kosten die elektronische Speicherung und der Versand (das ZIB oder ein vergleichbares Institut) nur 510,- DM pro Jahr. Ihr stehen die 125.000,- DM gegenüber, die für die traditionelle Vervielfältigung und den Versand von Papier anfallen. Diese Abschätzung gilt nur noch für 1995, denn das DFN hat seine Tarife inzwischen drastisch reduziert. Dadurch werden die Kosten für den elektronischen Versand weiter fallen, was für die Vervielfältigung und den Versand von Papier nicht zu erwarten ist.

Die Kosten für die notwendige Infrastruktur sind in diesen Betrag nicht eingerechnet, weil ein mathematischer Fachbereich oder ein Forschungsinstitut diese heutzutage sowieso vorhalten muß (das ZIB betreibt ein lokales Netz mit Supercomputern, über 160 Workstations und 25 PCs). Auch die Personalkosten für die Abspeicherung der Reports, die von den Sekretariaten druckfertig in LaTeX erstellt werden, sind nicht extra aufgeführt, weil die Archivierung ohnehin ordnungsgemäß vorgenommen werden muß, auch im Falle des Versands von Papier. Die Inanspruchnahme der Workstation durch den elektronischen Versand via E-mail könnte man korrekterweise noch veranschlagen. Wir übergehen sie jedoch an dieser Stelle, weil der Versand so organisiert werden kann, daß er nur nachts oder an Wochenenden stattfindet, wo die Workstation sonst ungenutzt bleibt. In den nächsten Kapiteln werden wir jedoch auch diese Kosten vollständig mit einbeziehen.

Der Trick ist natürlich, daß das ZIB so die Kosten für die Vervielfältigung seiner Reports auf die empfangenden Institute und Leser verlagert - aber ist das nicht auch fair? Es ist auch von der Kostenseite her gesehen korrekt im Vergleich zum traditionellen Versand von Preprints auf Papier. Diese sollten in der Regel in der Bibliothek allgemein zugänglich sein (tatsächlich sind sie oft in Schreibtisch eines Forschers vergraben). Wer nun solch ein Report benötigt, weil er damit arbeiten will, wird ihn für sich kopieren. Ferner, Empfänger elektronischer Reports werden ja nicht immer alles sofort ausdrucken, sondern nur die Reports, die sie wirklich lesen wollen.

Diese Diskussion sollte plausibel machen, welche dramatischen Kosteneinsparungen für "Anbieter" wissenschaftlicher Information möglich sind, wenn sie dafür elektronische Mittel und Netze einsetzen, selbst wenn das Internet heute vielen noch als teuer erscheinen mag. Das ist aber - so gesehen - schon für ein selbständiges mathematisches Forschungsinstitut, wie für das ZIB, nicht der Fall, und innerhalb einer Universität können viele Fachbereiche an einem einzigen Anschluß partizipieren.

3.2 Das Modell eines massiv-verteilten mathematischen Preprint-Archivs

Wir wollen jetzt ein Modell der Verteilung von Preprints diskutieren, das bis vor kurzem noch fast undenkbar erschien. Könnten nicht alle mathematischen Fachbereiche und Forschungsinstitute - wir rechnen in der Mathematik mit etwa 1.000 "funktionierenden" und am Internet angeschlossenen Institutionen - allen anderen ihre Preprints via E-mail zur Verfügung stellen, augenblicklich, sobald diese verfügbar sind?

Dann hätte jede Institution bald eine höchst aktuelle Sammlung der kompletten Forschungsliteratur, ein mathematisches Preprint-Archiv in einem Hypertextsystem oder einer Datenbank, die vor Ort recherchierbar ist. Nicht nur die Mitglieder von *in-groups* hätten dann Zugriff auf die neuesten Forschungsskripte, sondern alle Mitglieder des Fachbereichs und alle Studenten der gesamten Universität. Ja man könnte selbst Lehrern an Schulen und anderen Personen in der Region, auch aus Firmen, den Zugriff auf diese Sammlungen ermöglichen. Den dafür notwendigen Server wird das nicht überlasten.

Jedes Institut könnte die mathematischen Preprints der Welt in der ihm am besten erscheinenden Art und Weise offerieren, mit oder ohne Datenbanken, in Hypertext- oder Dokumentensystemen verschiedenster Bauart, isoliert oder im World Wide Web, integriert oder kompatibel mit anderen Systemen vor Ort. Es könnte dabei immer die ihm vertrauten Verfahren und Systeme verwenden.

Diese Frage erscheint uns für die Akzeptanz des vorgeschlagenen Systems von zentraler Bedeutung zu sein, denn schon über die Gestaltung der Präsentation (des *Nutzerinterfaces*) einer solchen Sammlung kann man heute sehr unterschiedlicher Meinung sein. Es scheint uns viel zu früh zu sein, um hierfür Vorgaben zu machen. Alle diese Systeme sind auch noch Gegenstand der aktuellen Forschung und Entwicklung. Die beteiligten Institute sollten selbst genügend Freiheiten haben und mit den verschiedensten Formen experimentieren, damit sich die besten entwickeln und durchsetzen können. Es ist ferner bei den gegenwärtig stark belasteten internationalen Netzverbindungen nicht ohne Bedeutung, wenn man vor Ort recherchieren und auf Volltexte zugreifen kann, ohne dabei durch Engpässe im Netz behindert zu sein.

Rein technisch gesehen ließe sich der elektronische Versand schon heute organisieren. Wir diskutieren seine Realisierung durch einen simplen E-mail-Verteiler, der 1.000 Einträge enthält.

Das Modell des massiv-verteilten Preprint-Archivs ist ein "Belastungsmodell". Es war ursprünglich nur zum Zwecke gedacht, erste Abschätzungen der notwendigen Investitionskosten und der resultierenden Belastung der Internet-Anschlüsse durch ein *Verteiltes Informationssystem* für die Mathematik zu gewinnen, einem von der DMV geplanten Projekt [DalitzGHLS 95]. Wir halten es jedoch für aufschlußreich, auch seine Realisierung zu diskutieren. Ein massiv-verteiltes Preprint-Archiv wäre sehr viel einfacher zu realisieren, als das in einem anderen Sinne "verteilt" gedachte mathematische Informationssystem, denn es ist auf "Literatur" spezialisiert. Dennoch lassen sich an Hand dieses Modell wichtige Schlüsse für das allgemeinere System ziehen. Und, es liefert Lösungen zu etlichen organisatorischen Problemen, die sonst nicht einfach zu bekommen sind.

Den E-mail-Verteiler braucht man nur zum Kopieren bereitzustellen. Jede Institution, die einen Report verschicken will, besorgt sich dafür die jeweils aktuelle Kopie des Verteilers und nimmt dann den Mailversand selbst vor. Dabei hält es sich an ein einfaches (aber noch genauer zu bestimmendes) standardisiertes Verfahren, so daß Empfänger die relevanten Bestandteile einer Mail leicht identifizieren und in ihre eigene Datenstruktur umsetzen können. Ein zentraler *Mailserver*, über den alle Reports verteilt werden, ist nicht angebracht, weil durch ihn die gesamte "Masse" der Preprints hindurchgeschleust werden müßten. Jeder Report würde doppelt verschickt.

Die in der Theoretischen Hochenergiephysik verwendete Form[55] (gemäß dem Modell von [Ginsparg 94]) wäre eine Möglichkeit. Hierbei kann man den Autor oder die Autoren, den Titel, den Abstract und den Report selbst aus einer E-mail extrahieren, die nur ASCII-Code enthält (einfacher Text oder LaTeX-Code). Aber warum sollte man nicht auch Multimedia-Formate verwenden (MIME-Form, gemäß [BorensteinF 92]), die praktisch alles beinhalten, was heutzutage elektronisch verschickt werden kann, von einfachen ASCII-Texten, über LaTeX- und HTML-Codes bis hin zu (sogar farbigen) Grafiken.

Zentral fällt in diesem Modell nur die Pflege des Mailverteilers an, eine harmlose Aufgabe, die sogar fast selbstorgansierend eingerichtet werden kann. Jede Institution richtet dazu genau eine E-mail-Adresse ein, über die der Versand abzuwickeln ist, also eine Adresse der Form

library@math.tu-berlin.de

wobei der user-name "library" (buchstäblich) fest vorgeben ist. Jede Institution könnte sich in diesen Verteiler selbst eintragen. Zentral genügt eine einfache Gegenkontrolle der *IP-Adresse* (im obigen Beispiel "math.tu-berlin.de"), damit sich nur anerkannte mathematische Forschungsinstitute dieses Verteilers bedienen können. Dadurch soll nicht etwa ausgeschlossen sein, daß die Preprints auch an andere Institute oder Personen weitergegeben werden können, sondern nur, daß dieses nicht mittels des mathematischen Preprint-Verteilers geschieht. Eine weitere Unterverteilung ist Sache der jeweiligen Institute.

Jede Institution benennt einen *Informationsbeauftragten* - der mittels dieser Adresse - die Preprints verschickt. Er trägt die Verantwortung dafür, daß nur von der Institution anerkannte Reports auf diese Weise verteilt werden.

[55]http://xxx.lanl.gov/ - TH-HEP E-Print Archive (LANL)

3.3 Qualität, Authentizität, Archivierung und Urheberrecht

Die Verantwortung für die Qualität eines Reports trägt beim Modell des massiv-verteilten Preprint-Archivs (wie auch beim noch zu diskutierenden Verteilten Informationssystem für die Mathematik) das Institut, und nicht der Informationsbeauftragte und auch nicht ein Autor allein. Damit ist die Frage der Qualität angesprochen, die sonst in Verbindung mit PreprintServern immer kontrovers diskutiert wird; weiteres dazu unten.

In dem *E-Print-Modell* der Physik, den sogenannten Ginsparg-Servern, kann sich jeder beliebige Autor eines solchen Mailverteilers bedienen, wenn er nur über eine elektronische Mailbox verfügt [Ginsparg 94]. Für die Mathematik kommt solch ein Verfahren nicht in Betracht; wir diskutieren diese Frage in [Grötschel 95a]. Es kann aber durchaus möglich sein, daß uns die Wirklichkeit hier überholt, denn einzelne mathematische Fachgruppen experimentieren bereits mit solchen Servern[56] [57] [58] [59] [60].

Auch die Frage der Authentizität von Reports wäre auf diese Weise lösbar. Ein Informationsbeauftragter, der von irgendwoher einen Report empfängt, kann - mittels der Liste der zugelassenen E-mail-Adressen - gegenprüfen, ob diese Sendung "legitimiert" ist, und alles andere zurückweisen, sogar automatisch. Das ist nun vielleicht etwas simplifiziert gesagt, denn ein "Hacker", der über nur geringe Systemkenntnisse zu verfügen braucht, kann seine E-mail so modifizieren, daß die in ihr mitgelieferte Absenderadresse eine beliebige Form annimmt, also auch die einer "zugelassenen" E-mail-Adresse. Er kann das System also technisch schlagen, aber er hinterläßt dabei Spuren im Netz, und er macht sich strafbar, wenn er auf diese Weise fälschlich datenschutzrelevante Information vorgaukelt. Und, Preprints sind, da sie Namen von Autoren enthalten, immer personenbezogen und damit datenschutzrechtlich relevant, jedenfalls hierzulande. Es gibt ferner inzwischen bereits elektronische Unterschriften. Es wäre zu überlegen, in welcher Weise man den Versand mit ihrer Hilfe absichern kann.

Für Zwecke der eindeutigen Identifizierung (im Zweifelsfalle) von Reports könnten die mathematischen Institutionen auch noch ein Namensschema vereinbaren, etwa in der Art

www.math.tu-berlin.de/library/reports/SC-95-02

wobei auch hier die Zeichenketten "library" und "reports" fest vorgegeben sind. In der Namenssystematik ihrer Preprints oder Reihe(n) wären sie völlig frei (im Beispiel: "SC-95-02"), ebenso in der Wahl der WWW-Adresse (im Beispiel: "www.math.tu-berlin.de"), solange diese nur bekannt gemacht wird (ebenfalls mittels des bereits genannten Verteilers). Die Institute sollten aber garantieren, daß ihre Berichte auf diese Weise auch tatsächlich elektronisch erreichbar sind (in einem Hypertext; man könnte auch *ftp* verwenden). Jeder könnte dann einen Report (die Originalversion) nicht nur adäquat zitieren, sondern sich im Falle irgendwelcher Zweifel oder Fehler von der Identität mit der ihm bekannten Version überzeugen.

Die Frage der Authentizität hat noch eine andere Seite, die im Zusammenhang mit dem elektronischen Angebot von Preprints immer wieder kritisiert wird, zumeist von Verlegern,

[56] http://eprints.math.duke.edu/alg-geom/ - Algebraic Geometry Preprint Archive
[57] http://www.msri.org/preprints/auto-fms.html - Automorphic Forms Preprint Archive
[58] http://www.msri.org/preprints/cd-hg.html - Complex Dynamics and Hyperbolic Geometry
[59] http://www.msri.org/preprints/dg-ga.html - Differential Geometry and Global Analysis
[60] http://eprints.math.duke.edu/q-alg/ - Quantum Algebra and Topology

die das Vorrecht, die Authentizität zu sichern, ganz für sich und das traditionelle Publizieren auf Papier in Anspruch nehmen wollen. Den Autoren wird dabei bisweilen unterstellt, daß sie ihre eigenen Arbeiten nachträglich noch ändern und modifizieren wollen, zum Beispiel, wenn ihnen Fehler nachgewiesen werden. Dabei übersehen sie jedoch, daß sich elektronisch verteilte Nachrichten und Dokumente niemals zurückziehen lassen, auch wenn sie nur an einen kleineren Teilnehmerkreis verschickt werden. Jeder, der solch eine Nachricht erhält, kann sie beliebig leicht weiterverteilen, auch ohne daß der Autor davon erfährt.

Die Mathematik kennt hierfür ein Paradebeispiel: Am 23. Juni 1993 ging - im Anschluß an eine Vortragsserie in Cambridge - die Nachricht vom Beweis der berühmten Fermat'schen Vermutung durch A. Wiles per E-mail blitzschnell um die Welt (DMV-Mitteilungen 3/93, S. 3-6). In nur wenigen Stunden war die Gemeinschaft der Mathematiker wie elektrisiert, vielleicht schneller, als es dem Autor lieb war, denn er hatte seinen Beweis nicht einmal publiziert, sondern nur angekündigt. Dieser Beweis war recht lang (einige hundert Seiten) und nicht unkompliziert. Doch die allgemeine Freude währte nicht lang, denn er war fehlerhaft (DMV-Mitteilungen 1/94, S. 22-23). Die Nachricht von dem Fehler verbreitete sich dann ebenso schnell. Nebenbei gesagt: Dieses Beispiel spricht nicht gegen den Einsatz elektronischer Mittel, denn A. Wiles hatte seinen Beweis ja gar nicht selbst elektronisch verkündet, es demonstriert aber die Wirksamkeit der elektronischen Kommunikation, im positiven wie im negative Sinne. Inzwischen hat Wiles seinen Beweis modifiziert, und er scheint allgemein akzeptiert zu sein (DMV-Mitteilungen 1/95, S. 12-14 und 2/95, S. 6-8).

Elektronisch verschickte Dokumente tragen nicht nur einen Zeitstempel (sogar mit genauer Uhrzeit), sondern sie "sind" gewissermaßen (weit) publiziert. Es wäre nicht im Sinne des Modells, wenn einmal elektronisch angebotene Berichte noch geändert oder zurückgezogen werden könnten. Sie verlören dann in gewisser Weise ihre Identität und wären nicht mehr korrekt zitierbar.

Im Gegensatz zu traditionellen Preprints auf Papier, die aus Kostengründen nur in beschränkter Stückzahl hergestellt werden und bald "vergriffen" sind, können elektronische Preprints in fast beliebig großer Zahl (und noch beliebig lange) kopiert werden. Mathematische Institute, die ihre Reports elektronisch anbieten, tragen deshalb eine besondere Verantwortung, nicht nur im eigenen Interesse, sondern auch dem Autor gegenüber. Diese Frage trat früher nur deshalb nicht in ihrer ganzen Schärfe auf, weil der Verbreitungsgrad von Preprints auf Papier nicht potentiell so hoch war.

Das Beispiel "A. Wiles" zeigt jedoch, daß das Problem der Sicherung der Qualität nichts mit der Frage der Publikation auf elektronischem Wege zu tun hat. Seinen fehlerhaften Beweis hatte er ja nicht publiziert, auch nicht elektronisch. Das "Problem" ist heute einfach durch die Existenz elektronischer Kommunikationsmittel gegeben. Wer es heute verhindern will, daß seine Arbeit Gefahr läuft, zu früh elektronisch verbreitet zu werden, dürfte schon beim Schreiben keinen Computer benutzen, geschweige denn elektronische Post. Aber wer kann das ernsthaft in Betracht ziehen? Wir leben heute "in einer elektronischen Welt" - auch die Verfechter des traditionellen Publizierens.

Selbst die Frage der Archivierung wäre im geschilderten Modell des massiv-verteilten mathematischen Preprint-Archivs lösbar, und zwar ohne daß jede einzelne Institution eigene *Backup*-Verfahren einsetzen muß. Wenn 1.000 über die ganze Welt verstreute Institutionen immer die komplette Sammlung vorhalten, wird es kein großes Problem sein, sich - zum Beispiel durch Spiegeln der Daten (dafür gibt es automatische Verfahren) - immer wieder eine vollständige Kopie zu verschaffen, wenn einmal vor Ort ein Speichermedium ausfällt.

Solch ein einfaches Modell zur Verteilung von Reports und Preprints, bei dem die mathematische Information vollständig bei allen teilnehmenden Institutionen repliziert wird (wie es sich in ähnlicher Weise schon bei den News bewährt hat), hätte noch einige andere Vorteile, so zum Beispiel bezüglich der Frage der Sicherung des Urheberrechts. Gegen die elektronische Distribution von Artikeln wird oft eingewendet, daß diese dann leicht kopiert, verändert und möglicherweise unter falschem Namen weiterverbreitet werden könnten. Dieses gehört zu den wichtigeren Argumenten, die heute von Kolleginnen und Kollegen genannt werden, die der Welt des elektronischen Publizierens skeptisch gegenüberstehen. Rein technisch gesehen ist das auch möglich, jedoch setzt sich ein "Ideendieb" damit einem besonders hohen Risiko aus. Der von ihm im Internet gestohlene Text läge ja bereits bei 1.000 Instituten vor, er wurde ja vorher verteilt - per E-mail und mit Zeitstempel. Es kommt uns außerordentlich unwahrscheinlich vor, daß dies niemand bemerken sollte, falls der Plagiant dann seinerseits "seine" Arbeit verteilt, er müßte es ja später tun.

Es ist also eher umgekehrt: Je früher ein Autor seine Ergebnisse an die Öffentlichkeit bringen kann, desto sicherer kann er sein, daß seine Urheberschaft gewahrt bleibt (und ein Datendieb sich praktisch selbst überführt). Genau dieser Umstand ist eines der Hauptmotive dafür, in der mathematischen Forschung mit Preprints zu arbeiten. Heute kann man elektronische Mittel einsetzen.

3.4 Kostenüberlegungen und Belastung der Netzanschlüsse

Organisation und Technik sind - wie wir gezeigt haben - nicht das Problem. Diese Mittel stehen uns bereits zur Verfügung. Welche Investitionen müßte jedes einzelne Institut vornehmen? Wird die Realisierung eines solchen Modells nicht insgesamt zu teuer? Jede Institution müßte ja eine vollwertige Workstation vorhalten. Und, werden die Netzanschlüsse nicht allzusehr belastet?

Wenn 1.000 mathematische Institutionen ihre wissenschaftlichen Preprints, 50 Reports im Jahre 1995 und jährlich nahezu 10% mehr, an alle anderen Partner verschicken (wir rechnen wieder mit einer Verdoppelung des wissenschaftlichen Outputs in 10 Jahren), dann entstehen, verteilt über 10 Jahre, folgende Kosten (wir führen die Rechnung im Anhang detailliert aus):

Investitionen für insgesamt 145 Gigabyte Plattenspeicher	36.000,- DM
Modernisierung der Plattenspeicher - alle 3 Jahre	13.000,- DM
Beschaffung einer Workstation und ihr Ersatz - alle 3 Jahre	40.000,- DM
Wartung, Betriebs- und Datenbanksoftware	30.000,- DM
Summe über 10 Jahre:	120.000,- DM

Jedem Institut entstehen damit insgesamt weniger Investitionskosten, als etwa das ZIB schon im ersten Jahr für Vervielfältigung und Versand von Reports auf Papier ausgibt.

Der Grund dafür liegt im "Gesetz von Moore", nach dem sich die Kapazitäten der Speicher auf Grund der Fortschritte in der Technologie alle 18 Monate verdoppeln - bei gleichbleibenden Kosten. Das entspricht einer jährlichen Kostenreduktion um ein Drittel (wir gehen von den Verhältnissen im Workstation-Bereich aus). Diese Entwicklung soll nach Angaben der Fachleute noch in den nächsten zehn Jahren anhalten. Man muß aber ständig modernisieren und dabei optimal große Plattenspeicher beschaffen, um den Kostenvorteil auch wahrnehmen zu können. Interessant ist, daß die Gesamtkosten für die Beschaffung des Plattenspeichers (von dem jährlich etwa 10% mehr zu beschaffen sind) mittelfristig selbst dann fallen, wenn alle 3 Jahre eine Ersatzbeschaffung vorzunehmen ist. Das gibt einen Hinweis darauf, wie der allseits beklagten "Informationsexplosion" auf technischer Ebene begegnet werden kann.

Stellt sich in dieser Betrachtungsweise dann überhaupt noch die Frage, ob der Einsatz elektronischer Mittel zu teuer ist, oder ist es nicht eher umgekehrt? Viel teurer scheint uns zu sein, sie nicht einzusetzen. (Vielleicht sollten wir die Rechnungshöfe darüber informieren! So seltsam es erscheinen mag, manchmal sind sie es, die derartigen Veränderungen mehr Schubkraft verleihen.)

Mit einer Rechnung analog zu der in Kapitel 3.1 für den Versand und dem Empfang erhalten wir (für das ZIB) ferner:

anteilige Netzkosten in 1995 (2-Megabit-Anschluß)	3.000,- DM
anteilige Netzkosten in 1996 (34-Megabit-Anschluß)	300,- DM
anteilige Netzkosten in 1997 (34-Megabit-Anschluß)	330,- DM

denn das Deutsche Forschungsnetz wird ab 1996 höhere Übertragungskapazitäten anbieten, die das ZIB sicherlich wahrnehmen wird.

Diese Betrachtungen sind natürlich nur dann korrekt, wenn sich alle mathematischen Fachbereiche und Forschungsinstitute an dem Austausch von Preprints und Reports beteiligen und dabei Anschlüsse einer gewissen Mindestleistung einsetzen. In Deutschland verfügen heute 43 von insgesamt 74 Universitäten und mathematischen Forschungseinrichtungen über 2-Megabit-Anschlüsse. In den USA ist die Entwicklung der Datenautobahnen schon sehr viel weiter vorangeschritten. Dort ist die Mehrzahl der Universitäten bereits mit 34-Megabit-Leitungen versorgt (zu deutlich geringeren Kosten), und etliche testen schon noch höherwertigere Anschlüsse (155 Megabit pro Sekunde und mehr).

Ein 2-Megabit-Anschluß kann, wie wir im Anhang ausführlicher zeigen, monatlich 288 Gigabyte übertragen, auch unter erheblicher Belastung. In 1995 würde der Versand der mathematischen Reports (1,09 Gigabyte im Monat) den Anschluß des ZIB damit nur zu 0,38% belasten und der Empfang (0,84 Gigabyte im Monat) zu einem noch geringeren prozentualen Anteil. Ein 34-Megabit-Anschluß kann monatlich die gewaltige Informationsmenge von 4.896 Gigabyte übertragen. Die Belastung eines solchen Anschlusses (des ZIB) durch den Versand in 1996 (monatlich 1,19 Gigabyte) liegt dann unter 0,025% und durch den Empfang (monatlich 0,92 Gigabyte) nochmals darunter.

Wir möchten heute keine Aussage darüber treffen, wie es ab 1998 mit der Leistungsfähigkeit und den Kosten der Netze aussehen wird. Dann werden auch in Europa die Datenautobahnen dereguliert. Wir rechnen sowohl mit weiteren kräftigen Preisnachlässen als auch mit wesentlich höheren Bandbreiten. Den Grund sehen wir in der Kommerzialisierung des Internet und der globalen Datenautobahnen, die einen noch breiteren Kreis von Anwendern an sich ziehen werden. Dadurch wird sehr viel Geld eingenommen und in die weitere Entwicklung der Netztechnologien fließen - und diese damit noch stärker beschleunigen. Auch die Wissenschaft wird davon profitieren.

Nehmen wir, um eine langfristige Perspektive aufzuzeigen, als Beispiel den Versand von Spielfilmen über das Netz, der unter dem Stichwort "Multimedia-Information" für die Wissenschaften interessant ist. Ein Spielfilm nimmt in digitalisierter Form und komprimiert *(JPEG)* weniger als 3 Gigabyte in Anspruch, also etwa soviel wie der Versand der Jahresproduktion der Reports des ZIB in 1995 an seine 300 Kooperationspartner. Dieses entspricht einem Netzkostenanteil von 390,- DM (wie oben bereits dargelegt). Wenn *video-on-demand* ein Erfolg werden soll, darf die Übermittlung "auf" der *Datenautobahn* nicht mehr kosten als das Video selbst. In den USA rechnet man mit einem Preis von unter 10 Dollar, und das wäre - bezogen auf deutsche Verhältnisse - noch sehr viel. Wir sind mit den neuen 34-Megabit-Anschlüssen bereits ein Stück auf diesem Weg. Bei einem solchen Anschluß würde die Übertragung (im Jahre 1996 und in Deutschland) nur noch mit 33,50 DM zu Buche schlagen. Diese Entwicklung wird aber noch weitergehen.

3.5 Ein verteiltes Informationssystem für die Mathematik

Über die massive Verteilung von Preprints läßt sich trefflich streiten. Werden Mathematiker dann nicht durch die Fülle an verfügbarer Information überlastet? Soll man überhaupt jeden Report allgemein zugänglich machen, auch allen Studenten? Wäre es nicht besser, mehr auf Qualität zu achten und nur eine gewisse Auswahl bereitzustellen [Hoffmann 95]? Wir vertreten in dieser Kontroverse eine klare Position. Uns geht es in erster Linie darum, die mathematische Kommunikation auf breiter Basis wiederherzustellen und dabei die Defizite des traditionellen Publizierens zu überwinden, sowohl bezüglich der Geschwindigkeit der Informationsverbreitung als auch bezüglich der Kosten. Die genannten Gegenargumente sprechen unserer Ansicht nach ganz allgemein gegen die Verwendung von Preprints, doch die Mathematik kommt nicht ohne sie aus, auch in der bisherigen Welt ohne elektronische Mittel nicht.

Wir treten allerdings dafür ein, jetzt - weil uns die elektronische Welt gegeben ist - in besonderem Maße auf Qualität zu achten (bzw. diese Position nicht widerstandslos aufzugeben). Hier appellieren wir an das Verantwortungsbewußtsein der beteiligten Institutionen, denn daß sich der im Bereich der Journale etablierte Standard, also das aufwendige Verfahren der Begutachtung und des *Reviewing*, nicht so ohne weiteres voll auf die wissenschaftlichen Reports übertragen läßt, scheint unbestritten zu sein. Aber warum sollte man nicht auch Studenten mit dem "Problem" der Informationsfülle konfrontieren? Gerade sie wachsen in eine Welt hinein, in der sie damit umgehen können müssen. Und, bei guter Strukturierung, läßt sich das bewußte Problem reduzieren.

Setzen wir voraus, daß alle Autoren ihre Arbeiten nach dem in der Mathematik weltweit verbreiteten *Mathematics Subject Classification* Schema (MSC-Index) klassifizieren [FIZKA 94]). Dieser enthält heute rund 4.000 Rubriken. Bei 50.000 Reports im Jahr kommen auf jede Rubrik - sie repräsentiert jeweils eine spezielle Teildisziplin - 24 bis 36 Beiträge (wir rechnen mit 2 bis 3 MSC-Indices je Report). Durch ein aufwendiges Reviewing-Verfahren ließe sich dieser Anteil vielleicht reduzieren, aber ist diese Zahl denn wirklich zu hoch? Würde sich der personelle Mehraufwand lohnen?

Unser Vorschlag ist, ein System zu schaffen, das jedem, der bei einem bestimmten Sachgebiet subskribiert oder sucht, die für ihn interessanten Artikel zugänglich macht. Im Modell des massiv-verteilten Systems wäre eine solche Unterverteilung Sache der mathematischen Institutionen. Wir gehen aber beim gegenwärtigen Stand der Diskussion davon aus, daß sich ein massiv-verteiltes Preprint-Archiv nicht sofort und in einem Zuge realisieren läßt. Die mathematischen Institutionen müssen erst lernen, ihre Reports überhaupt elektronisch bereitzustellen. Die Schwierigkeiten liegen nicht im technischen Bereich, sie sind anderswo zu suchen (Lernbereitschaft und notwendiges Umdenken).

Wir schlagen deshalb zunächst die Einrichtung eines Nachrichtensystems vor. Jede Institution, die einen Preprint elektronisch im Web bereitstellt, schickt nicht den ganzen Report, sondern nur eine Nachricht über seine Verfügbarkeit über einen Mailverteiler. Diese enthält, neben der Adresse (URL) des Reports etwa seine erste Seite, mit den Angaben zu Autor(en), Titel, Klassifikationen und Stichworten - und den Abstract. Da diese Texte wesentlich kürzer sind als der Report, braucht die Verteilung nicht auf mathematische Institutionen beschränkt zu werden. Jeder Mathematiker soll subskribieren können, zu beliebigen Teilgebieten (oder auch Obermengen davon) und zu jeder Kombination von Stichworten, die sein spezifisches Interesse beschreiben. Das System soll die Nachrichten aber nicht nur verschicken, sondern gleichzeitig elektronisch archivieren, in einer Datenbank, damit man später darin suchen kann, und auch in einem Hypertextsystem, das den einfachen Zugriff auf alle bisherigen Nachrichten erlaubt.

Das ist - in knapper Form - der technische Kerngedanke eines verteilten Informationssystems für die Mathematik, das die DMV[61] realisieren will [DalitzGHLS 95] [Lügger 95]. Dieses System beläßt die Originalinformation bei der jeweiligen Institution, sichert aber - weltweit - sowohl die Information über Neuigkeiten, als auch (für alle, die nicht sofort daran teilnehmen) die Möglichkeit der Lokalisierung relevanter Information. Beides ist notwendig, sowohl die Benachrichtigung als auch das Retrieval, denn nicht jeder kann von vornherein wissen, was ihn einmal interessieren wird. Für das Retrieval kommen ebenfalls zwei unterschiedliche Arten in Betracht, zum einen traditionelle Techniken (klassisches Datenbankretrieval), zum anderen der Einsatz von Hypertextmitteln (zum Browsen in der vollständigen Menge der angebotenen Information). Wir werden im Rahmen des geplanten Projekts der DMV auch untersuchen, inwieweit sich diese sehr verschiedenen Techniken weiter integrieren lassen. Denkbar und hilfreich wäre zum Beispiel für Nutzer von Hypertexten das Browsen auf Teilmengen zu ermöglichen, die durch bestimmte Kombinationen von Klassifikationen und Schlüsselworten gegeben sind.

Das vorgeschlagene Modell des verteilten Informationssystems läßt sich - im Gegensatz zum Modell des massiv-verteilten Preprint-Archivs - auch für Informationen anderer Art generalisieren und einsetzen, zum Beispiel für Software- und Datensammlungen. Mathematische Softwaresysteme sind oft außerordentlich umfangreich, und sie entwickeln und ändern sich auch, solange sie "leben", nicht selten in außerordentlich kurzen Zeitabständen. Sie kommen damit, kurz gesagt, kaum für eine massive Verteilung in Betracht. Da es aber auch im

[61]http://elib.zib-berlin.de:8000/Cmath.org.softinf.pub - Planungspapiere der DMV

Bereich der mathematischen Software ein international anerkanntes Klassifikationssystem gibt, den *Guide to Available Mathematical Software* (GAMS-Index) des *National Institute for Standards and Technology* (NIST) [BoisvertHK 91], kann man mit ihr in ähnlicher Weise verfahren. Die Datenbank des verteilten Informationssystems der Mathematik sollte allerdings dafür die Verwendung unterschiedlicher Klassifikationssysteme und Kategorienschemata erlauben, denn Software ist anders zu beschreiben als Literatur. Wir denken in diesem Zusammenhang an die in der *netlib* verwendeten Schemata [BrowneDGGMRW 94] und - technisch - an ein objektorientiertes Datenbanksystem.

Die Mathematik hat vieles an Informationen zu bieten, das nicht nur für Mathematiker interessant ist, sondern auch für Anwender der Mathematik, für Physiker, Chemiker und andere Naturwissenschaftler, für Ingenieure, Verfahrenstechniker und andere Wissenschaftler aus den technischen Bereichen, bis hin zu Psychologen, Sozialwissenschaftlern und anderen Geisteswissenschaftlern. Sie alle könnten von Informationsangeboten dieser Art profitieren. Für die Forschung in der Mathematik sind auch Beschreibungen aktueller Forschungsprojekte und Informationen zu Institutionen, die sie durchführen, relevant. In diesem Zusammenhang sind auch organisatorische Angaben von Interesse, wie Namen und Adressen von Institutsangehörigen und Projektleitern und vieles andere mehr. Für Studenten und Hochschullehrer sind Skripte und Materialien zur Lehre von Bedeutung, die elektronisch bundesweit zur Verfügung stehen könnten. Alles dieses will die DMV in das verteilte Informationssystem integrieren.

Für den Nutzer solcher heterogenen Datenbestände ergibt sich dadurch ein besonderes Problem. Wie sollen er sich in der auf viele Orte verteilte Information zurechtfinden? In aller Regel sucht er Information zu einem bestimmten Thema und will dafür nicht erst eine große Zahl von Servern absuchen. Und, das oben geschilderte Nachrichtensystem wird ihm auch nicht immer helfen, denn es ist ja nicht sinnvoll, jede einzelne Web-Seite anzukündigen.

Die örtliche Information soll deshalb in Kollektionen desselben Typs gegliedert sein, die mit Datenbankmitteln durchsuchbar sind, und sie soll vielstufig hierarchisch gegliedert und klassifiziert sein. Diese "vertikalen" Informationshierarchien sollen "horizontal" miteinander vernetzt sein, indem alle Kollektionen desselben Typs (also Preprints oder Software oder organisatorische Daten) durch Hyperlinks miteinander verbunden sind. Mit ihrer Hilfe sollen Nutzer "quer durch Deutschland" - auch mit Hypertext-Mitteln allein - nach speziellen Inhalten suchen können. Darüberhinaus wollen wir untersuchen, in welcher Weise sich auch im Internet vorhandene *Robots* und *Indexierer* für diese Zwecke einsetzen lassen, wie z.B. das Harvest-System [BowmanDMS 95] [BowmanDHMSW 95].

Ein Thema, das der DMV besonders am Herzen liegt, ist das elektronische Publizieren, denn sie sieht - wie wir - auch die Schwächen der traditionellen Informationsversorgung. Sie will durch die Gründung eines universellen elektronischen mathematischen Journals, der *Documenta Mathematica* [Rehmann 95], nicht nur einen Beitrag dazu leisten, die drohende Publikationslücke zu schließen, sondern damit auch der explosiven Kostenentwicklung im Bereich der wissenschaftlichen Journale entgegenwirken, die in dieser Zeit zu Abbestellungen bei vielen Bibliotheken führt; siehe z.B. [WinterZ 95]. Sie berührt damit auch das Thema der Qualität, das für das neue elektronische Medium von zentraler Bedeutung ist.

Mathematiker aus aller Welt diskutieren dieses Thema schon seit einiger Zeit und mit besonderer Intensität, denn sie fürchten qualitative Einbußen beim Übergang in die elektronische Welt. Es geht ihnen vor allem um die Zuverlässigkeit der mathematischen Literatur, die über die Jahre einen besonders hohen Standard erreicht hat; siehe [Quinn 95]. Die progressiveren unter ihnen sehen eher Chancen als Gefährdungen und schlagen vor, die bewährten Mechanismen der Qualitätssicherung auch in die elektronische Welt zu übertragen; siehe [Odlyzko 95], ferner [FMC 94]. Die *Documenta Mathematica* werden deshalb ein hochrangig besetztes Gremium von Herausgebern, Redakteuren und Gutachtern.

Die Frage der Qualität ist aber auch für die Akzeptanz des verteilten Informationssystems von Bedeutung. Wenn es auch nicht sinnvoll erscheint, hierfür den Journalen vergleichbare Standards anzustreben, sollte doch eine gewisse Kontrolle der Qualität vorgenommen werden. Unser Vorschlag ist, bestimmte "kritische" Kollektionen und Sachgebiete des Nachrichtensystems (wie Preprints gewisser MSC-Indices oder Software gewisser GAMS-Klassen) durch "Moderatoren" zu betreuen, welche Einsendungen kurz überprüfen, bevor sie verschickt und archiviert werden.

4. Strategien gegen Information Overload

Wie steht es nun um den Wissenschaftlers, der sich einer "überwältigenden" Fülle an wissenschaftlicher Information gegenüber sieht. Für den Mathematiker, der es mit Computern zu tun hat, und das ist zunehmend der Fall, stellt sich diese Frage mit noch größerem Gewicht. Der mathematische "Output" umfaßt heute nicht mehr nur Reports, sondern auch Algorithmen, also Software und die zugehörigen Sammlungen von Test- und Ergebnisdaten, letztere oft in graphischer *(Visualisierungen)* und sogar in dynamischer Form (bewegte Graphiken, Videos, Multimedia-Information). Auch diese Dinge, die vom traditionellen Publikationswesen bisher noch gar nicht erfaßt wurden (welche Bibliothek könnte Software archivieren?), muß er in der Forschung berücksichtigen. Obendrein verfolgt er Tagungen, Workshops, Seminare und nimmt an einer Reihe von weiteren formellen und informellen Diskussionen teil, heute zunehmend unter Einsatz elektronischer Mittel. Aber spricht ein Wissenschaftler dann vom "Information Overload"? Im Gegenteil, wenn er sein Fach beherrscht und liebt, wird er darüber nicht unglücklich sein. Wir kommen jetzt zu Strategien und Mitteln, wie sich auch größere Informationsmengen angemessen behandeln lassen.

4.1 Qualität als Mittel der Selektion

Der menschliche Geist verfügt über eine Reihe von hervorragenden "Mechanismen", die für ihn wesentlichen Informationen auszuwählen, zu bewerten, einzuordnen und in neuen Zusammenhängen zu sehen. Je mehr Information ihm dabei zur Verfügung steht, desto besser sind seine Vergleichsmöglichkeiten. Der menschliche Wissensdurst ist nahezu unbegrenzt, aus welchen Motiven auch immer. Anders wäre der ständige Drang nach immer neuen Erkenntnissen kaum zu verstehen. Wir sind nicht davon überzeugt, daß "Information" (was immer das auch ist) den Menschen jemals zu viel wird, wenn sie nur neu und interessant - und attraktiv gestaltet ist.

Als Last oder "Overload" empfindet man sie nur dann, wenn sie wenig Neues bringt (inhaltsleer, langweilig), zu schwer einzuordnen (verworren, chaotisch) oder ganz einfach falsch (widersprüchlich) ist, also dann, wenn sie nicht eigentlich "Information", sondern nur "Rauschen" ist. Von "Overload" würden wir bei "echter Information" nur sprechen, wenn es nicht gelingt, den eigentlichen Inhalt aus ihr zu gewinnen und dabei den "Non-Sense" effizient zu überspringen, etwa indem man ihn einfach identifizieren und damit ignorieren kann. Wir lassen als Mathematiker den Fall der Falschinformation für unsere weitere Betrachtung jetzt einmal außer acht. Die Mathematik verfügt über gute Kriterien, die in sich widersprüchliche Information von der übrigen zu trennen. In anderen Wissenschaften ist es vielfach nicht ganz so "einfach".

Nach der "Selektion" der einfach falschen Artikel bleibt die Frage, was ist Information guter Qualität, denn die "wertvollen" Artikel aus der Fülle der übrigen effizient herausfinden zu können, ist das eigentliche Heilmittel gegen "Information Overload".

Nun kann das, was für den einen von höchster Bedeutung ist, für den anderen absolut nichtssagend sein. Qualität und Non-Sense sind zunächst also personenbezogen zu sehen. Einem Physiker kann eine Arbeit, die ein Mathematiker vielleicht gerade verworfen hat, den Schlüssel zu seiner eigenen Forschung liefern. Ob ein Artikel "wirkliche" Information ist oder nicht, hängt oft mehr vom Verständnis ab, mit dem er gelesen wird, und dem "Kontext", in den er eingebettet ist, als von seinem schriftlichen Inhalt. Überspitzt könnte man sagen, das Wesentliche steht nicht im Papier (nicht einmal zwischen den Zeilen). Diesen Kontext bildet kein Mensch für sich allein, sondern er wird von der kulturellen Gemeinschaft gebildet, also - im Wissenschaftsbereich - von den verschiedenen Gruppen und Institutionen, in denen ein Forscher arbeitet und lebt.

Was Qualität und was Non-Sense sind "bestimmen" die maßgeblichen, fachspezifischen Gruppen. Dabei orientieren sie sich an den im Fach aktuellen Grundsatzfragen und Problemen (Paradigmen) [Kuhn 76]. Mathematiker sind dabei bekanntlich besonders rigoros. Für sie hat eine Arbeit höchstens dann Qualität, wenn sie wenigstens einen einzigen neuen mathematischen Gedanken enthält. Der Punkt, auf den es uns hier ankommt, ist, daß durch den Kontext (das aktuelle "Wissen" eines Fachs) bestimmt ist, was Qualität ist und was nicht. Dieser Kontext wird von vielen, sehr unterschiedlichen Gruppen (immer aufs Neue) gebildet, in einem lebendigen, hochgradig dynamischen Prozeß der wechselseitigen Information und Kommunikation.

Unsere wichtigste Schlußfolgerung ist, daß das fachliche Urteil (Begutachtung, fachliche Kommentierung und Moderation) unerläßlich ist, wenn man an Information, an Dokumentation und an Kommunikation von Wissen denkt. Der vielschichtige Prozeß der Auswahl, der Bewertung und der Vermittlung von Information kann deshalb unserer Meinung nach auch grundsätzlich nicht automatisiert werden. Das Wesentliche steht nicht explizit im Text. Kein automatisches Information-Retrieval-Verfahren, das ja "nur" auf der Analyse von Texten beruht, wird dieses Kernproblem der Bewältigung von "zu viel" Information lösen können.

Der Prozeß der Bewertung kann auch nicht "in inhaltlicher Distanz" vom Fach in qualifizierter Art und Weise vollzogen werden, im Gegenteil. Wir müssen deshalb damit rechnen, daß der Anteil an Verantwortung, den die Fachwissenschaftler zu übernehmen haben, in Zukunft zusammen mit dem Mehr an Information noch weiter wächst.

4.2 Struktur und Transparenz umfangreicher Information

Wie kann nun "Qualität", die ja für jeden etwas anderes bedeutet, in der zu erwartenden Fülle an Information angemessen dokumentiert und vermittelt (mitgeteilt) werden? Trotz aller Filterung und Selektion wird die insgesamt zur Verfügung stehende Information weiter wachsen, auch wenn man sehr stringente Maßstäbe anlegt. Das ist das Wohl und Wehe der jetzt sich entwickelnden Informationsgesellschaft. Wir möchten diese Fülle erhalten, denn wenn zu viel bereits gewonnenes Wissen durch Ausfiltern wieder verloren geht, kann der resultierende Schaden beträchtlich sein, ganz abgesehen von dem Schicksal der durch übermäßige "Selektionsmaßnahmen" betroffenen Personen.

Welche Hilfsmittel stehen uns zur Verfügung, um den in Zukunft noch größeren Reichtum an Information bewältigen zu können? Es bleibt nur der Weg, sich aller heute zur Verfügung stehenden technischen und organisatorischen Mittel zu bedienen. In erster Linie sind das, nach der qualitativen Bewertung, eine übersichtliche und transparente Strukturierung und eine hochwertige Präsentation der Information. Dafür gibt es gute Vorbilder.

J. Nielsen diskutiert ein Beispiel aus dem traditionellen Pressewesen, das schon seit über hundert Jahren entsprechende Verfahrensweisen entwickelt hat [Nielsen 95]. Die *New York Times* "konfrontiert" ihre Leser in ihren Wochenendausgaben mit einer geschätzten Informationsmenge von 7 bis 8 Megabyte an Text und 180 Megabyte an Bildern (zum Vergleich: das reichlich bebilderte Buch von Nielsen enthält auf knapp 500 Seiten circa 1 Megabyte Text und 13 Megabyte an Bildern). Trotzdem sind die Abonnenten, die Sonntag für Sonntag diese gewaltige Informationsmasse vor ihrer Tür finden, darüber nicht unglücklich. Es würde sie sicherlich außerordentlich viel Zeit kosten, dieses alles Wort für Wort und Bild für Bild zu studieren. Doch die Leute tun das nicht. Statt dessen wählt jeder Leser individuell den Teil aus, der ihn interessiert - und übergeht den Rest.

Das ist möglich, weil die *New York Times* so gestaltet ist, daß der Leser sich nicht nur leicht in der Struktur orientieren kann, sondern sich auch auf jeder Seite sofort zurecht findet. Er kann weniger Interessantes leicht überspringen, denn er kennt seine Zeitung. Sie ist sinnfällig in mehrere Teile gegliedert (Politik, Wirtschaftsteil und so weiter), sogenannte "Bücher". Diese sind wieder themengerecht angeordnet. Jede Seite ist durch *Layout* und *Satz* und durch die Verwendung von geeigneten Stilarten und Schriftbildern übersichtlich gestaltet. Es gibt Überschriften, Zwischentitel, erläuternde Textboxen, Grafiken und vieles anderes mehr. Wer nur sehr wenig Zeit zum Lesen aufbringen kann, findet alle wichtigen Nachrichten auf der ersten Seite - und bleibt dennoch informiert.

Glücklicherweise gibt es also bereits ein umfassendes Repertoire an Gestaltungs- und Ausdrucksmitteln, um auch sehr große Mengen an Information angemessen darstellen zu können; wir haben nur ganz wenige davon genannt. Und, das Medium Zeitung ist "billig", so daß es nicht viel ausmacht, wenn man das meiste davon nicht liest.

Außerdem, und nicht zuletzt, enthält die *New York Times* nicht nur "trockene" (neutrale, unpersönliche) Information, sondern sie ist durchsetzt mit Artikeln, die zum Lesen anregen: Geschichten, Auszüge von Büchern, Kritiken, Kommentare und Berichte (natürlich auch Klatsch), geschrieben von Menschen, die man wiederkennt und deren Urteil man schätzen oder ablehnen kann.

Alles das ist in der "guten alten Welt" der Großrechner und ihrer großen Datenbanken und Retrievalsysteme nicht zu finden. Schöne Schriften, Ausdrucke und bunte Bilder wurden damals bisweilen sogar als unseriös verlacht. Heute gewinnen guter Satz, schöner Druck und die farbige Gestaltung hoher Qualität eine immer größere Bedeutung auch in der Welt der Datenverarbeitung. Das alles ist mehr als nur eine Spielerei. Es hilft bei der Orientierung und macht nicht nur das Leben angenehm.

4.3 Die Ausdrucksmittel traditioneller Retrieval-Systemen sind beschränkt

Machen wir jetzt ein Gedankenexperiment: Stellen wir uns vor, man müßte die New York Times mit Hilfe eines klassischen Retrieval-Systems "lesen". Jede einzelne Nachricht, jeder Kommentar und jede andere Informationseinheit wäre elektronisch erfaßt und stünde nun in einer Datenbasis zur Verfügung, in der man mit traditionellen Mitteln nach Stichworten suchen kann.

Zunächst fällt auf: Die Zeitung ist "atomisiert". Der übergeordnete Zusammenhang ist fast vollständig verschwunden. Auch wenn jedes Dokument zusätzlich deskribiert wäre (aber mit welchen Stichworten oder Klassifikationen?), ist der natürliche Aufbau der Zeitung verloren. Und sollte es wirklich ein normiertes System von Klassifikationen geben, es wäre sehr umfangreich und der Leser müßte es Wort für Wort kennen, und dazu einen "Thesaurus" der Synonyme.

Fast noch wichtiger ist: Ein System dieser Art verschließt die einzelnen Dokumente wie in einer *Black Box*. Diese muß der Leser mit Stichworten "befragen" (vielleicht in einer formalisierten Retrievalsprache). Jedoch, ist das, was er erhält auch das was er (nach Ansicht eines Redakteurs) finden sollte? Was bei der *New York Times* durch *Layout-* und *Satzgestaltung* ins Auge springt, ist in solch einem Modell nicht vorhanden. Wie soll der Leser dann sofort "sehen" können, was wesentlich ist? Wie soll er die weniger wichtigen Dinge überspringen? Und, hat er das wesentliche mit seiner Anfrage überhaupt getroffen?

Es fehlt der "Zusammenhang". Den stellt der Leser mit seinen Anfragen erst her, eine nichttriviale Aufgabe, selbst für einen Retrievalspezialisten. Doch was er dabei "konstruiert", entspricht nicht dem, was der Redakteur und der Layouter mit ihrer besonderen Art der "Buch-", der Seiten- und der Textgestaltung mitteilen können. Hilfestellungen zum Überlesen gibt es dabei nicht. Die heute gängigen Retrievalsysteme liefern keine typographische Information, sondern nur ASCII-Texte. Überschriften, Zwischentitel und Textboxen können somit nicht dadurch helfen, daß sie sofort ins Auge fallen.

Nicht zuletzt gibt es Beziehungen, die sich überhaupt nicht in Worte fassen lassen. Auch die Mathematik kennt dafür Beispiele. Wie soll man sich ausdrücken, wenn man nach einer mathematischen Formel sucht? Wie sollte man etwa ein Bild beschreiben, wenn man nach Visualisierungen von Rechenergebnissen sucht?

Wir geben zu, unser Beispiel ist nicht ganz fair. Traditionelle Retrievalsysteme sind nicht zum Lesen von Zeitungen gedacht, sondern für andere Zwecke. Unser Gedankenexperiment, das etwas überzeichnet ist, sollte nur demonstrieren, was ihnen vielleicht fehlt.

4.4 Ausdrucksmittel von Hypertexten, Semantische Assoziationen

Hypertextsysteme bieten völlig anders geartete Ausdrucksmittel an. Sie integrieren eine große Vielfalt an verschiedenartigster Information, nicht nur Texte in verschiedenen Größen, Stil- und Schriftarten, sondern auch Grafiken und farbige Bilder, Töne und Musik, ja sogar vollständige Videoinformation. Man spricht deshalb von Multimediasystemen (wir verwenden jedoch den gebräuchlicheren Ausdruck “Hypertext”). Selbst Spezialisten, wie J. Nielsen fällt es schwer, zu sagen, was Hypertexte eigentlich genau sind. Sie sind noch Gegenstand der aktuellen Forschung und Entwicklung.

Viele Hypertextsysteme enthalten obendrein integrierte Datenbanken oder einfache Retrieval-Systeme. Wir wollen uns aber darauf konzentrieren, den Unterschied zwischen Hypertexten und Datenbanken darzustellen, denn wir halten Hypertexte für ein völlig neuartiges Medium.

Hypertexte gestatten es, die Information vor dem Leser in strukturierter Weise zu entfalten, sie verbergen sie nicht. Man liest sie am Bildschirm mit dem Browser, so wie ein langes Dokument. Autoren von Texten können ihre Dokumente schön und übersichtlich gestalten, wobei sie Mittel verwenden, die denen in der Satz- und Drucktechnik verwandt sind. Wer das World Wide Web kennt, weiß, daß wir jetzt nicht nur von Hypertexten im Internet reden. Das Web ist noch sehr jung. Es enthält noch nicht alle Ausdruckmittel, die denkbar oder notwendig sind, wie zum Beispiel zur Darstellung mathematischer Formeln. Dennoch sind die Gestaltungsmöglichkeiten von WWW-Dokumenten, die in der “Hypertextsprache” (HTML) geschrieben sind, schon recht vielfältig, sie werden sich auch noch weiter entwickeln [Graham 95].

Autoren können zwischen einzelnen Worten und Bildern einen Dokuments und irgend einem anderen Hypertextdokument Querbezüge herstellen, mittels sogenannter Hyperlinks. Wer beim Lesen eines solchen Dokuments auf einen Hyperlink stößt (das betreffende Wort oder Bild ist entsprechend gekennzeichnet, meist unterstrichen oder umrahmt), kann ihn anklicken, und bekommt dann das zugehörige Dokument gezeigt. Dieser “Mechanismus” bildet den Kern aller Hypertextsysteme. Hyperlinks können auch auf Musik oder Videos “zeigen”, diese werden dann abgespielt. Aber wir wollen unsere Darstellung vereinfachen und konzentrieren uns jetzt auf die Behandlung von strukturierten Texten.

Schon ein einziger, gut gesetzter Hypertext-Link drückt etwas aus, was sich nicht mit sprachlichen Mitteln formulieren läßt. Die durch den Link gegebene “Beziehung” ist ja nicht mit Worten definiert. Ted Nelson, einer der Erfinder von Hypertextsystemen, spricht lieber von Gedankenassoziationen. Bei komplex vernetzten Mengen von Hypertextdokumenten spricht man von *belief*-Netzen, weil sie semantische Zusammenhänge repräsentieren.

Hyperlinks sind für Leser das hauptsächliche, technische Mittel, “neue” Information zu erschließen. Dabei bleibt der vom Autor gedachte Zusammenhang im Prinzip gewahrt. Dieses funktioniert bei den meisten Hypertextsystemen allerdings nur “in eine Richtung” (ihre Links sind *unidirektional*), vom Ausgangsdokument zum Zieldokument. Die meisten Browser merken sich deshalb bereits gelesene Seiten und gestatten es, zu ihnen zurückzublättern. Einige wenige Systeme wie Hyper-G verwenden *bidirektionale* Hyperlinks [KappeM 93]. Sie gestatten Assoziationen in beide Richtungen, was manchmal von Vorteil ist.

Das World Wide Web und einige andere, globale Hypertextsysteme, enthalten Hyperlinks, die auf andere Rechner verweisen, genauer, auf eine beliebige Stelle der Information, die ein anderer Web-Server bereitstellt. Solche Hyperlinks "zeigen" gewissermaßen dem Leser den Platz einer elektronischen Ressource, die er in gleicher Weise nutzen kann. Man spricht von deshalb einem *Uniform Resource Locator*, kurz URL [BernersCGP 92]. Die gefundenen Information bietet sich dem Leser immer in derselben, uniformen Weise dar. Leser des Web haben es deshalb besonders einfach, zwischen den verschiedensten Web-Ressourcen hin- und herzuwechseln und dabei die unterschiedlichsten Rechner zu nutzen.

Autoren können Sie dabei unterstützen, indem sie - wenn auch auf völlig anderer Weise, als ein Redakteur einer Zeitung - ihre Texte gliedern, strukturieren, und im Aussehen übersichtlich gestalten. Allerdings sind Autoren und Leser durch die Technik etwas eingeschränkt, denn ein Browser kann immer nur einen recht kleinen Ausschnitt der Information am Bildschirm zeigen. Es ist für Leser deshalb nicht immer leicht, den Überblick über die gesamte Information zu behalten, wenn er zwischen den verschiedenen Dokumenten eines Hypertextes hin- und herspringt. Man sagt, er kann dabei "im Hypertext verloren gehen".

Gute geschriebene Hypertexte sind jedoch klar, oft hierarchisch, strukturiert, und ein gut gebauter Server bietet eine Reihe von Hilfen zum Blättern in seinen jeweiligen Textdokumenten: zurück, vorwärts, in der Hierarchie nach oben, in ein Inhaltsverzeichnis hinein und von dort wieder - per Hypertextlink - an eine beliebige andere Stellen seines Informationsangebots.

Mit der richtigen "Schreibtechnik" können sogar sehr große Dokumentenmengen angemessen strukturiert werden. Das bekannteste Beispiel ist das World Wide Web selbst. Es umfaßt heute viele Millionen Dokumente. Von einigen wenigen, guten Startdokumenten[62] [63] [64] ausgehend, die hierarchisch nach Sachgebieten oder nach Orten gegliedert sind, kann sich ein Leser für ihn relevante Teile des Web erschließen. Inzwischen gibt es bereits eine Reihe von Suchmaschinen, also auf das Web spezialisierte Retrieval-Systeme[65] [66] [67], die ihn dabei unterstützen. Hypertexte sehen wir als das neue Paradigma für die Information und Dokumentation an.

5. Neue Perspektiven für das Informations- und Dokumentationswesen

In der heutigen Zeit eröffnen sich völlig neue Perspektiven für das gesamte Informations- und Dokumentationswesen. Durch die sich entwickelnden Datenautobahnen sind wir dem alten Traum, auf das gesamte gespeicherte Wissen der Welt zugreifen zu können, direkt von jedem Arbeitsplatz und sogar von zu Hause aus, einen Schritt näher gerückt. Der globale *Cyberspace*, die *Infobahn*, die *Information Super Highways*, die *globale Informationsinfrastruktur* oder wie man es auch nennen mag, weist uns den Weg, ihn zu realisieren. Begriffe wie die

[62] http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/MetaIndex.html - Internet Resources Meta-Index
[63] http://www.yahoo.com - Yahoo (Stanford Univ.)
[64] http://www.w3.org/hypertext/DataSources/bySubject/Overview.html - WWW Virtual Library
[65] http://lycos.cs.cmu.edu/ - Lycos Suchmaschine
[66] http://webcrawler.cs.washington.edu/WebCrawler/WebQuery.html - WebCrawler Suchmaschine
[67] http://www.infoseek.com/ - InfoSeek Suchmaschine

globale Bibliothek und die Informationsgesellschaft beginnen in greifbare Nähe zu rücken. Wir fassen all dieses jetzt unter dem Stichwort "Digitale Bibliotheken" zusammen. Durch Investitionen in die Forschung und in die Ausbildung haben wir die Chance, die Welt etwas näher zusammenrücken zu lassen und unsere Umwelt dadurch zu verbessern [FoxAFL 95].

Digitale Bibliotheken existieren bereits in vielfältigen Formen und in den verschiedensten Ausprägungen, vor allem in der Forschung, aber auch im traditionellen Bibliotheks- und Dokumentationswesen, und ebenfalls in den öffentlichen Einrichtungen und Archiven. Jeder mag etwas anderes darunter verstehen - wir haben nur einige wenige Facetten zeigen können. Für den Geowissenschaftler und den ökologisch orientierten Forscher sind es vielleicht die großen Datenarchive, die bei der Beobachtung unserer Erde anfallen, für den Historiker sind vielleicht die Datensammlungen der großen Museen interessant, ein Mathematiker sieht vielleicht eher den Schatz der mathematischen Forschungsliteratur, der Bibliothekar und Dokumentar versteht möglicherweise darunter seine klassische Datenbank, der eher technisch interessierte Informatiker sieht vielleicht "nur" ein automatisches Retrieval- oder ein (globales) Hypertextsystem, der Verleger vielleicht eine Dokumentenbank seiner Literatur, der Archivar oder Museologe denkt vielleicht an Technologien zur elektronischen Konservierung.

Wie auch immer, durch das Internet und die neuen Möglichkeiten der globalen Kommunikation erhalten digitale Bibliotheken eine neue Dimension, deren Tragweite erst langsam in das Bewußtsein dringt - gerade in unserem Lande. Allenthalben sind auch Ängste zu spüren. Wird man die sich abzeichnenden Veränderungen bestehen können, wenn nun praktisch jeder mit eigenen Informationsangeboten konkurrieren kann? Gehen nicht Arbeitsplätze verloren, wenn effizientere Methoden zum Einsatz kommen? Kann man sich seinen Lebensunterhalt noch verdienen, wenn alles das auch noch die Kosten reduziert? Was ist, wenn die alten Strukturen zusammenbrechen? Wird das Rollenspiel völlig neu definiert?

Die bange Frage ist immer: Wie sieht die Zukunft aus? Unsere These ist, wir können sie nicht dadurch meistern, daß wir die Augen verschließen - und uns verweigern. Wir leben in einer offenen Gesellschaft. Viele Mitspieler sind in diesem Umfeld tätig, oft mit hohen Einsätzen. Niemand kann vorhersagen, welcher Strategie der Vorzug zu geben ist. Und, selbst mit großem finanziellen Einsatz könnte niemand durchsetzen, daß das alte Rollenspiel bestehen bleibt. Wer könnte denn z.B. isoliert entscheiden, daß die traditionellen Informationsanbieter auch in Zukunft das ganze Spiel allein beherrschen, wenn schon einzelne und Personen in der Lage sind, ihre Leistungen zu erbringen, wie etwa Paul Ginsparg [Ginsparg 94]? Selbst wenn wir in Deutschland solch eine Maßnahme durchsetzen wollten, wer könnte denn Wissenschaftler in anderen Ländern davon abhalten, ihre eigenen Interessen wahrzunehmen?

Wir können zwar nicht in die Zukunft sehen, jedoch haben wir die Möglichkeit sie mitzugestalten, indem wir uns ihr stellen und die neuen Möglichkeiten erproben. Unser Vorschlag ist, das Experiment Zukunft zu wagen und entsprechende Initiativen zu starten. Wir nennen nur einige aus einem breiten Spektrum von Möglichkeiten im Bereich der Archivierung, Dokumentation, Information und Kommunikation.

1. Im Bereich der Wissenschaften:

   * Aufbau und Pilotanwendungen verteilter Informationssysteme in ausgewählten Wissenschaftsbereichen auf der Basis globaler Hypertextsysteme; Einsatz insbesondere für

Informationen über laufende Forschungsvorhaben (Projektdarstellungen) und deren organisatorische Einbettung.

* Aufbau von verteilten und globalen Archiven für wissenschaftliche Preprints und spezielle Forschungsliteratur (wie zum Beispiel Dissertationen) in ausgewählten Fachgebieten; Entwicklung eines Nachweissystems für diese Art der verteilten Information.

* Entwicklung und Erprobung von Modellen elektronischer Journale sowohl im Bereich der Wissenschaften allein als auch im Umfeld des Verlagswesens; Schaffung elektronischer Organisationsformen zur Unterstützung und Beschleunigung der Begutachtung und des Abstrahierungs- und des Referatewesens; Übertragung von Verfahren der Qualitätskontrolle in den elektronischen Bereich.

* Entwicklung von Systemen und Standards für verteilte Software- und Testdatensammlungen, die auch Visualisierungen in Bild und Video beinhalten können; Entwicklung entsprechender Nachweisverfahren; Fokussierung auf Fragen der Wiederverwendbarkeit von Forschungssoftware bei verschiedenen Rechnerarchitekturen und Rechnerplattformen insbesondere im Bereich des Höchstleistungsrechnens.

* Erschließung wissenschaftlicher Datenarchive, zum Beispiel aus den Geowissenschaften, der Luft- und Raumfahrt, der Meeresforschung oder der Ökologie; Vernetzung solcher Archive innerhalb des jeweiligen Fachgebiets; Einsatz benutzerfreundlicher Zugriffsverfahren über das Internet, insbesondere auf Daten, die für den Umweltschutz von Bedeutung sind, und Öffnung solcher Archive für jeden interessierten Bürger.

* Elektronische Konservierung und Erschließung historischer Dokumente und Bücher, einschließlich Copyright-freier Materialien; langfristige Bewahrung wertvollen Kulturguts, zum Beispiel durch elektronische Bereitstellung ausgewählter, seltener und wertvoller Dokumente und Materialien auch aus Handschriftenarchiven und aus dem Museumsbereich.

2. Im Bereich der öffentlichen Einrichtungen und Verwaltungen:

* Klärung von Fragen der langfristigen Archivierung, sowohl im organisatorischen (wer entscheidet, was archiviert wird) als auch im technischen Sinn (welche Datenformate sind für die langfristige Speicherung geeignet); Erprobung digitaler Archivierungsverfahren in ausgewählten Pilotprojekten im Bereich der Bibliotheken und der Archive der öffentlichen Verwaltungen.

* Offnung von Online-Informationssystemen für den Remote-Zugriff auf aktuelle ökonomische und statistische Daten, welche zum Beispiel bei statistischen Landesämtern oder Forschungsinstituten für die Wirtschaft vorliegen; Erprobung des Remote-Zugriffs auf solche Datensammlungen nicht nur im Bereich öffentlicher Verwaltungen, sondern ebenso durch Lehrer, Schulen und interessierte Bürger.

* Verteilte Bereitstellung von Informationen der Bundesregierung, wie zum Beispiel zu Fördermaßnahmen und ihren organisatorischen Randbedingungen; Verteiltes Angebot institutioneller und organisatorischer Informationen der verschiedenen Projektträger und von Organisationen, die im Bereich der Forschungsförderung tätig sind, sowie der von ihnen betreuten Förderprojekte.

Unser Vorschlag ist nicht, alle diese Daten jetzt erst zu erheben. Er zielt auf bereits vorhandene Informationen ab. Jahr für Jahr geben Bund und Länder Millionensummen aus, um solche Daten zu erheben und zu sammeln, nicht nur in der Forschung und im öffentlichen Bereich. Diese liegen in der einen oder anderen Form heute zwar elektronisch vor, sie sind aber oft nur Spezialisten zugänglich, weil sie in technisch isolierten Systemen "verschlossen" sind. In der Forschung spricht man des öfteren sogar von "Datenfriedhöfen". Ihr "Inhalt" ist oft nicht einmal den betroffenen Fachleuten mehr zugänglich, sie müßten die Daten erneut erheben.

Für diese Daten hat der Steuerzahler aber teuer bezahlt. Warum sollte er sie nicht ebenfalls nutzen können? Die neuen Technologien im Bereich der Kommunikation machen dieses jetzt möglich, und zwar zu einem geringen Preis. Die wertvollen Daten aus der Wissenschaft und dem öffentlichen Bereich müssen ohnehin archiviert und bereitgestellt werden, damit Steuergelder nicht verschwendet werden. Wenn jede betroffene Institution dieser Pflicht schon im eigenen Interesse genügt, kann sie das - ohne besonderen Mehraufwand - auch auf solche Art und Weise tun, daß der Zugriff über das Internet möglich ist.

Die Institutionen sind selbst für ihre Daten verantwortlich. Das Stichwort heißt Dezentralisierung. An eine zentrale Zusammenführung und Speicherung solcher Information denken wir dabei nicht. Für die Wissenschaftler und die Bürger sollten aber Möglichkeiten geschaffen werden, sich auf einfache Art und Weise in den ja von ihnen entfernten und auf mehrere Institutionen verteilten Angeboten zu orientieren. Zu diesem Zweck sind entsprechende Technologien zu entwickeln. Die Verantwortung hierfür sollte aber nicht bei den technologischen Spezialisten allein liegen. Unser Vorschlag ist, sie den Anwendern dieser Technologien zu übertragen, also den am verteilten Informationsangebot aktiv beteiligten Institutionen.

Wir schlagen vor, daß Wissenschaftler der verschiedensten Fachrichtungen und Verwaltungsfachleute, traditionelle Informationsversorger, wie Archivare, Bibliothekare, Dokumentare und Verleger, Informatiker, Computerspezialisten und Entwickler von Netzen miteinander kooperieren. Wir sind uns darüber im klaren, daß wir ein langfristiges Ziel vorschlagen. Es wird nicht gleich morgen realisierbar sein. Dennoch können wir schon heute erste Schritte wagen und - auch kurzfristig - erste Erfolge erreichen. Die dafür notwendigen Mittel stehen längst bereit. Wir brauchen sie nur anzuwenden.

Literatur

Die mit einem Sternchen gekennzeichneten Artikel stehen in elektronischer Version zur Verfügung:
http://elib.zib-berlin.de:8000/Cmath.org.softinf.pub
Die mit einem Plus-Zeichen gekennzeichneten Artikel erreicht man elektronisch über die Adresse:
http://elib.zib-berlin.de:8000/Cmath.org.softinf.pub.ref

[ACM 94] Special Issue: Internet Technology; Comm. ACM, Vol. 37, No. 8, August 1994

[ACM 95] Special Issue: Digital Libraries; Comm. ACM, Vol. 38, No. 4, April 1995

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

[AAU/ARL 94] + AAU/ARL Task Forces Reports on: (1) Acquisition and Distribution of Foreign Language and Area Studies Materials, (2) A National Strategy for Managing Scientific and Technological Information, (3) Intellectual Property Rights in an Electronic Environment; Association of American Universities (AAU) Research Libraries Project in collaboration with the Association of Research Libraries (ARL); 1994

[BernersCGP 92] T. Berners, R. Cailliau, J.-F- Groff, B. Pollermann; World-Wide-Web: The Information Universe; Electronic Networking, Vol. 2, No. 1, 1992, 52-58

[BoisvertHK 91] R.F. Boisvert, S.E. Howe, D.K. Kahaner; The Guide to Available Mathematical Software problem classification system; Comp. Stat. - Simul. Comp., 20 (4) 1991, 811-842

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

[BowmanDMS 95] C.M. Bowman, P.B. Danzig, U. Manber, M. Schwartz; Scalable Internet Resource Discovery: Research Problems and Approaches; Comm. ACM, Vol. 37, No. 8, August 1994, 98-107,114

[BowmanDHMSW 95] C.M. Bowman, P.B. Danzig, D.R. Hardy, U. Manber, M. Schwartz, D.P. Wessels; Harvest: A Scalable, Customizable Discovery and Access System; Techn. Report CU-CS-732-94, Univ. of Colorado, Boulder, 1995

[BrowneDGGMRW 94] S. Browne, J. Dongarra, E. Grosse, S. Green, K. Moore, T. Rowan and R. Wade; Netlib Services and Resources; Tech. Report UT-CS-94-222, University of Tennessee Comp. Sci. Dept., Feb. 1994

[CummingsWBLE 92] + A.M. Cummings, M.L. Witte, W.G. Bowen, L.O. Lazarus, R.E. Ekmann; University Libraries and Scholarly Communication: A Study Prepared for the Andrew W. Mellon Foundation; Association of Research Libraries, Nov. 1992

[DalitzL 93] W. Dalitz, J. Lügger; Opt-Net: A new communication Facility for Scientists interested in Optimization; GMÖOR Newsletter, No. 2, 1993, 7-9

[DalitzGLS 95] * W. Dalitz, M. Grötschel, J. Lügger, W. Sperber; Verteiltes Informationssystem für die Mathematik - Kurzfassung eines Projektplans der DMV; DMV-Mitteilungen, Heft 1, 1995, 52-54

[DalitzGHLS 95] * W. Dalitz, M. Grötschel, G. Heyer, J. Lügger, W. Sperber; Ein verteiltes Informationssystem für die Mathematik - Beschreibung eines Vorhabens der Deutschen Mathematiker-Vereinigung (DMV); Konrad-Zuse-Zentrum für Informationstechnik Berlin, 22. Juli 1995; die Berichte [Lügger 95] und [Rehmann 95] sind Bestandteil des Projektplans der DMV.

[DenningR 95] P.J. Denning, B. Rous; The ACM Electronic Publishing Plan; Comm. ACM, Vol 38, No. 4, April 1995; 97-103

[DFN 95] Wir im Deutschen Forschungsnetz - Verzeichnis der Anwender des Wissenschaftsnetzes (WiN) und der DFN-Dienste, No. 18, DFN-Verein, Juni 1959

[DBI 95] Deutsches Bibliotheksinstitut; SUBITO: Bund-Länder Initiative zur Beschleunigung der Literatur- und Informationsdienste; vorläufige Projektunterlagen 1.9.1994 bis 31.12.1996 des Deutschen Bibliotheksinstituts, Berlin 1995

[DongarraR 91] J. Dongarra, B. Rosener; NA-Net: Numerical Analysis Net; Computer Science Dept. Univ. of Tennessee, 1991

[DongarraG 87] J.J. Dongarra, E. Grosse; Distribution of mathematical software via electronic mail; Comm. ACM, 30(5), May 1987; 403-407

[Erikson 94] H. Erikson; MBone: The Multicast Backbone; Comm. ACM, Vol. 37, No. 8, August 1994, 54-60

[FIZKA 94] Fachinformationszentrum Karlsruhe, Heidelberger Akademie der Wissenschaften (Hrsg.); 1994 Mathematics Subject Classification; Springer-Verlag, 1994

[Flynn 95] P. Flynn; The World Wide Web Handbook; Intern. Thomson Comp. Press, 1995

[FMC 94] + The Future of Mathematical Communication; Background Material of the Conference at MSRI, Berkeley, Nov. 30 - Dec. 3., 1994

[FoxAFL 95] E.A. Fox, R.M. Akscyn, R.K. Futura, J.J. Leggett; Digital Libraries (Guest Editorial); Comm. ACM, Vol. 38, No. 4, April 1995, 23-28

[Gaus 94] W. Gaus; Berufe im Archiv-, Bibliotheks-, Informations- und Dokumentationswesen - Ein Wegweiser zur Ausbildung, 3. Auflage; Springer-Verlag, 1994

[Ginsparg 94] P. Ginsparg; First Steps Towards Electronic Research Communication; Computers in Physics, Vol. 8, No. 4, Jul/Aug 1994

[Goldfarb 90] C.F. Goldfarb; The SGML Handbook; Clarendon Press 1990

[Graham 95] I.S. Graham; The HTML Sourcebook; Wiley 1995

[Grötschel 95a] * M. Grötschel, J. Lügger; Die Zukunft wissenschaftlicher Kommunikation aus Sicht der Mathematik; Spektrum der Wissenschaft, März 1995; 39-43

[GrötschelL 95b] * M. Grötschel, J. Lügger; Wissenschaftliche Kommunikation am Wendepunkt - Bibliotheken im Zeitalter elektronischer Netze; Zeitschrift für Bibliothekswesen und Bibliographie (ZfBB) 42, Heft 3, Mai/Juni 1995, 287-312

[Hoffmann 95] K.-H. Hoffmann; Die bibliothekarische Versorgung der Hochschulen im Zeitalter der elektronischen Medien; ABI-Technik, Vol. 15, No. 2, 1995, 101-105

[Kahn 94] R.E. Kahn; The Role of Government in the Evolution of the Internet; Comm. ACM, Vol. 37, No. 8, August 1994, 15-19

[KappeM 93] F. Kappe, H. Maurer; Hyper-G: A large Universal Hypermedia System and Some Spinoffs; ACM SIGGRAPH's experimental special online issue of "Computer Graphics", May 1993; auch IICM Report 364, TU-Graz, 1993

[Krol 94] E. Krol; The Whole Internet - Users Guide and Catalog; O'Reilly 1994, 2nd Edition

[Kuhn 76] T.S. Kuhn; Die Struktur wissenschaftlicher Revolutionen; Suhrkamp 1976; 2. rev. Auflage

[Lamport 86] L. Lamport; LaTeX - A Document Preparation System; Addison Wesley, 1986

[Lügger 95] * J. Lügger; Designüberlegungen für ein verteiltes Informationssystem für die Mathematik in Deutschland; Konrad-Zuse-Zentrum für Informationstechnik Berlin; Juni 1995, Technischer Report TR-95-09

[Nielsen 95] J. Nielsen; Multimedia and Hypertext - The Internet and Beyond; Academic Press 1995

[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 in Notices Amer. Math. Soc., Jan. 1995, elektronische Version verfügbar:
ftp://elib.zib-berlin.de/netlib/att/math/odlyzko/tragic.loss.long.ps.Z

[Quinn 95] + F. Quinn; Roadkill on the electronic highway: The Threat to Mathematical Literature; Notices of the AMS; Jan 1995

[Rehmann 95] U. Rehmann; Pilotprojekt zur Führung von elektronischen wissenschaftlichen Zeitschriften durch Bibliotheken; Fachbereich Mathematik; Universität Bielefeld; Juni 1995

[Schult 95] Th. Schult; Auditorium Massimum; Die Zeit, Nr. 33, 11. August 1995, S. 54

[Stix 95] G. Stix; Publizieren mit Lichtgeschwindigkeit; Spektrum der Wissenschaft, März 1995, 34-39

[Wiiërs 94] L. Wiiers; A Vision of the Library of the Future; in Developing the Library of the Future - The Tilburg Experience; H. Geleijnse, C. Grootaers (Eds.); Tilburg University Press, 1994, 1-9

[WinterZ 95] * T. Winter, U. Zimmermann; Zur Entwicklung des Zeitschriftenbestandes an mathematischen Fachbibliotheken; DMV Mitteilungen, Heft 1, 1995, 34-42

Anhang: Kosten des globalen mathematischen Preprint-Archivs

A.1 Investitionskosten je Institution

Wir gehen bei unseren Betrachtungen von zunächst 50.000 Preprints in 1995 aus (1.000 Institutionen mit je 50 Reports im Jahr) und rechnen mit einer Verdoppelung des wissenschaftlichen Outputs alle 10 Jahre. Für 1995 sind also 10 Gigabyte an Plattenspeicher anzusetzen, für 1996 dann 11 Gigabyte etc. Wir rechnen der Einfachheit halber mit 1 Gigabyte Zuwachs pro Jahr. Dieser Progression an notwendigem Speicherplatz steht eine Degression an Kosten gegenüber, die einer Preisreduktion um ein Drittel in jedem Jahr entspricht, denn die Kapazitäten der modernen Plattenspeicher verdoppeln sich erfahrungsgemäß alle 18 Monate bei gleichbleibenden Preisen.

Zur Beschaffung des Plattenspeichers: 1 Gigabyte an Speicher, das Anfang 1995 circa 1.000,- DM kostete, wird also in 1996 bereits für nur 666,66 DM zu haben sein etc. Das führt zu insgesamt fallenden Kosten für den notwendigen Plattenspeicher, obwohl jährlich 10% mehr an Speicherplatz vorzuhalten sind. Die Kosten betragen dann 10.000,- DM in 1995, nur noch 7.333.33 DM in 1996 etc. In der Summe entstehen über 10 Jahre 36.000,- DM an Kosten für die insgesamt notwendigen 145 Gigabyte an Plattenspeicher.

Modernisierung des Plattenspeichers: Platten lassen sich nicht in beliebiger Stückzahl an eine Workstation anschließen. Man muß also modernisieren, um mit der technologischen Entwicklung Schritt halten und um den daraus resultierenden Kostenvorteil wahrnehmen zu können. Wir rechnen deshalb mit einer Ersatzbeschaffung alle 3 Jahre. Das führt zur zusätzlichen Ersatzbeschaffung von 10 Gigabyte in 1998, 11 Gigabyte in 1999, etc. 23 (!) Gigabyte im Jahre 2001, 25 Gigabyte in 2002, und so weiter, und schließlich 39 (!) Gigabyte im Jahre 2004. In der Summe entstehen Kosten in Höhe von 13.000,- DM für Ersatzbeschaffungen.

Beschaffung und Modernisierung der Workstation: Für den Zeitraum von 10 Jahren ziehen wir die Beschaffung von drei Workstations in Betracht, wieder aus Modernisierungsgründen. Insgesamt setzen wir hier 40.000,- DM an und weitere 30.000,- DM für Betriebs- und Datenbanksoftware. Diese steht dann sowohl für die Datenhaltung und den Datentransfer, als auch für die Suche (Browsen und Datenbankretrieval) zur Verfügung.

In der Summe entstehen also 120.000,- DM an Investitionskosten - gestreckt über 10 Jahre.

A.2 Netzkostenanteil je Institution

Der Empfang aller 50.000 mathematischer Reports (10 Gigabyte Information) in 1995 nimmt nach einer Rechnung analog zu der im Kapitel 3.1 (mit Berücksichtigung einer nur 50%igen Belastung des Netzes) 100.000 Netzsekunden eines 2-Megabit-Anschlusses in Anspruch. Das entspricht einem Kostenanteil von 1.300,- DM für ein Institut der DFN-Volumen- und Kapazitätsklasse des ZIB. Derselbe Betrag ist für den Versand von 50 Reports an 1.000 andere Institutionen anzusetzen (ebenfalls 10 Gigabyte). Zählt man noch die 390,- DM für den Versand an persönliche Kooperationspartner hinzu, so kommen wir auf einen Betrag von rund 2.690,- DM Netzkostenanteil in 1995.

Schon 1996 sieht diese Rechnung völlig anders aus. Nach der neuen Preisstruktur des DFN kostet ein 2-Megabit-Anschluß (ohne Internet-Mehrwertdienste) nicht mehr 350.000,- DM, sondern nur noch 220.000,- DM. Wichtiger erscheint uns aber, daß das DFN ab dem nächsten Jahr 34-Megabit-Anschlüsse anbietet. Diese kosten 500.000,- DM im Jahr. Wir rechnen noch rund 100.000,- DM für die Internet-Mehrwertdienste hinzu, wobei wir jetzt die höchste DFN-Volumenklasse ansetzen (wieder mit Bezug auf ein Institut wie das ZIB). In 1996 schlägt ein Internet-Anschluß mit dann rund 600.000,- DM zu Buche (incl. MwSt.).

Das ZIB wird sich sicherlich für einen hochwertigen Anschluß entscheiden. Eine Netzsekunde dieses Anschlusses kostet dann zwar etwas mehr, nämlich 0,019 DM, jedoch lassen sich in dieser Netzsekunde jetzt "echte" 1,7 Megabyte an Information übertragen, denn der 34-Megabit-Anschluß ist 17 mal schneller als der 2-Megabit-Anschluß. Das ZIB benötigt für den Transfer von jetzt 25 Gigabyte in 1996 (11 Gigabyte für Empfang plus 11 Gigabyte für Versand plus 3,3 Gigabyte für die Beschickung von Kooperationspartnern) dann nur noch 14.883 Netzsekunden. Das entspricht einem Kostenanteil von deutlich unter 300,- DM.

Der Netzkostenanteil wird damit für ein vergleichbares Institut im Jahre 1996 dramatisch fallen, obgleich 10% mehr an Information zu übertragen sind. Wir überlassen es dem Leser, die anteiligen Kosten für die Netzbelastung über 10 Jahre zu extrapolieren.

A.3 Belastung der Netzanschlüsse und des Internet

Bei der Abschätzung der Belastung der Netzanschlüsse gehen wir zunächst davon aus, daß in 1995 (vom ZIB) monatlich 1,09 Gigabyte verschickt werden (empfangen werden noch etwas weniger). Der Versand soll - wir wollen eine mögliche Engpaßsitutation abschätzen - während der regulären Arbeitszeit geschehen. Diese setzen wir mit 20 Tagen im Monat und 20 Stunden am Tag an. In dieser Zeit ist der Anschluß im Mittel "regulär" belastet, denn das Internet operiert ja weltweit. Wir sehen mögliche Ausfallzeiten damit als bereits abgegolten an.

Insgesamt stehen im Monat damit 1.440.000 Netzsekunden zur Verfügung. Unter der Voraussetzung, daß aus einem 2-Megabit-Anschluß nur 80% an echter Übertragungsleistung gezogen werden, kommen wir zu einer Transportleistung von 200 Kilobyte pro Netzsekunde. Mit dieser Kapazität kann man die Informationsmenge von 288 Gigabyte im Monat übertragen, jeweils in beide Richtungen, in das lokale Netz ZIB "hinein" und aus dem Netz heraus.

1995 würde der Versand der mathematischen Reports eine 2-Megabit-Anschluß nur zu 0,38% belasten und der Empfang (0,84 Gigabyte) zu einem noch geringeren prozentualen Anteil.

Im Jahre 1996 steht dann dem ZIB wohl schon ein 34-Megabit-Anschluß zur Verfügung, der monatlich die gewaltige Informationsmenge von 4.896 Gigabyte (fast 5 Terabyte) transferieren kann, wieder jeweils in beide Richtungen. Die Belastung eines solchen Netzanschlusses durch den Versand (monatlich 1,19 Gigabyte) liegt dann unter 0,025% und durch den Empfang (monatlich 0,92 Gigabyte) nochmals darunter.

Gegen die Massiv-Verteilung von Preprints könnte man noch den technischen Einwand machen, daß dadurch die wertvollen internationalen Leitungen zu stark belastet werden. Das wäre möglicherweise auch der Fall, wenn die beteiligten Institute allzu sorglos vorgingen und ihre Reports jeweils einzeln verschicken, also jeden Report in tausendfacher Kopie. Wir überlassen es dem Leser, abzuschätzen, welche Informationsmengen zwischen Europa und den USA dann hin- und hergeschickt würden. Unser Berechnungsmodell suggeriert vielleicht ein allzu grobes Bild, denn wir haben für die Abschätzung der maximalen Belastung (nur) der Einfachheit halber angenommen, daß jeder elektronische "Brief" einzeln zählt. Tatsächlich ist das aber nur bei Empfängern der Fall, wenn die elektronische Post beim Absenden richtig gehandhabt wird. Die Belastung beim Absender ist sehr viel niedriger, wenn das Mailsystem richtig bedient wird.

Wenn eine Institution in Deutschland eine E-mail in die USA schickt, etwa an 1.000 Adressaten, muß der elektronische Brief nicht bereits in Deutschland tausendfach kopiert und dann erst verschickt werden. Das wäre nur bei regulärer Post der Fall. Der Vorteil elektronischer Post liegt auch darin, daß sie sehr viel sparsamer sein kann. In unserem Beispiel braucht die E-mail nur in einem Exemplar in die USA geschickt zu werden. Erst dort wird sie unterverteilt, und dabei für die jeweilige Adressaten kopiert - eine Leistung des Mailsystems. Beim Massenversand von E-Mail können auch noch raffiniertere Verfahren zum Einsatz kommen, wie etwa bei den UseNet News, die in großen Mengen durch die ganze Welt geschickt werden. Das sind allerdings Fragen für Netzspezialisten, die wir hier nicht mehr näher erörtern wollen.