



Martin Grötschel

Joachim Lügger

**Wissenschaftliche Information und Kommunikation
im Umbruch**

**über das Internet zu neuen wissenschaftlichen
Informationsinfrastrukturen**

ersienen in:
Forschung & Lehre, Heft 1/1996, S. 194-198

Wissenschaftliche Information und Kommunikation im Umbruch

über das Internet zu neuen wissenschaftlichen Informationsinfrastrukturen

Steht uns das Ende des traditionellen wissenschaftlichen Publizierens bevor? Neue Technologien eröffnen neue Chancen zur Bewältigung der Informationsflut und des gleichzeitigen Informationsmangels. Doch der Einsatz von Technik allein reicht nicht aus.

Martin Grötschel, Dr. rer. pol., Univ.-Professor für Mathematik, TU Berlin, Vizepräsident des Konrad-Zuse-Zentrums für Informationstechnik Berlin (ZIB), Vorsitzender der gemeinsamen IuK-Kommission der DMV, DPG, GDCh, GI

Joachim Lügger, Dipl.-Mathematiker, Leiter der Abteilung Software-Information des ZIB

Expansion der Wissenschaften

Keine Periode unserer Geschichte ist stärker von der Expansion der Bildung geprägt und kein Zeitalter mehr von Wissenschaft und Technik durchdrungen als das 20. Jahrhundert. Einige Schlaglichter: Vor vierzig Jahren machten 4% der Personen eines Geburtsjahrgangs in Deutschland Abitur. Heute sind dies 30%. In den USA schließen über 75% eines Jahrgangs die Highschool ab. In Japan erreichen 50% einen College-Abschluß. Im Jahre 1910 gab es insgesamt etwa 8000 deutsche und britische Physiker und Chemiker. In den späten achtziger Jahren wurde die Anzahl der aktiv in Forschung und Entwicklung tätigen Naturwissenschaftler und Ingenieure auf etwa 5 Millionen geschätzt.

Überall sehen Eltern bessere Chancen für ihre Kinder darin, daß sie ihnen den Besuch von Universitäten ermöglichen. Schwellenländer hatten durch Aktivitäten im Ausbildungsbereich Erfolg. Den Einsichtigen in Entwicklungsländern ist bewußt, daß nur Ausbildung aus der Armut herausführen kann. Diese Trends werden zu einem weiteren starken Anwachsen des Bildungs- und Wissenschaftsbereiches führen, selbst dann, wenn in den hochentwickelten Ländern, die ja nur einen kleinen Anteil an der Weltbevölkerung haben, eine Stagnation eintritt. Die Bildungsrevolution vollzieht sich langanhaltend und weltweit.

Kapazitätsgrenzen: Informationsflut und -mangel

Mehr Wissenschaftler publizieren mehr. Seit über 100 Jahren ist ein enormes Wachstum der wissenschaftlichen Literatur sichtbar. Die Produktion verdoppelt sich alle 16 Jahre, in der Mathematik und in den Naturwissenschaften schon alle 10 Jahre. Wenn sich dieser Trend fortsetzt, wird von heute bis zum Jahre 2010 ebensoviel publiziert, wie in der Geschichte der Wissenschaften insgesamt. Die großen Bibliotheken, die den Anspruch haben, die gesamte relevante Literatur vor Ort zur Verfügung zu stellen, müßten ihre Archivkapazität verdoppeln – und das nach weiteren 15 Jahren erneut. Aber welche Bibliothek kann das leisten?

Die Wissensproduktion schlägt sich nicht mehr allein in Publikationen auf Papier nieder. Gerade im technischen Bereich und der Industrieforschung entstehen leistungsfähige Forschungssoftware, komplexe Datensammlungen und ihre Visualisierungen, Computermodelle und Animationen, kurz, digitale Informationsprodukte und multimediale Information. Das explosionsartige Wachstum des Wissens ereignet sich nicht nur innerhalb der traditionellen Fachdisziplinen, sondern besonders im interdisziplinären Bereich und mehr und mehr an den "Kontaktpunkten zur Außenwelt". Die Fähigkeit, wissenschaftliche Information in jeder Form weltweit und kostengünstig zu vermitteln, wird damit zu einer Schlüsselfunktion nicht nur für ein einzelnes wissenschaftliches Fach, sondern für das Zusammenspiel der Fächer und ihr Wirken in der Gesellschaft. Aus ökonomischen Gründen kommt es heute zunehmend auf hohe Geschwindigkeit an.

Das traditionelle wissenschaftliche Publikationswesen, auf gedruckten Zeitschriften basierend, stößt hier an seine Kapazitätsgrenzen, nicht nur, weil ihm digitale Information verschlossen bleibt. Es ist vor allem schwerfällig und teuer. In der Mathematik, unserem eigenen Fachgebiet, auf das wir später noch häufiger verweisen werden, treten z.B. exorbitant lange Wartezeiten (2 bis 3 Jahre) und auch beklemmend hohe Ablehnungsraten (bis zu 80%) auf, und das oft nur, weil Zeitschriften aus Kosten- und Marketinggründen die Seitenzahlen beschränken. Die Mathematik, die Physik und die Informatik behelfen sich seit einiger Zeit mit dem Austausch von Preprints, weil gedruckte Journale für die Forschung viel zu spät kommen. Die wissenschaftlichen Bibliotheken stehen vor einem ernstesten Problem. Die Kosten für Abonnements wissenschaftlicher Journale steigen, oft weit über die Inflationsrate hinaus. Ihre Etats hingegen stagnieren oder gehen sogar zurück. Diese müßten aber mindestens mit demselben Tempo wachsen wie die Produktion von Literatur.

Die Bibliotheken sehen derzeit noch den Ausweg, sich mittels elektronischer Dokumentenlieferung zu behelfen. Diese kommt aber für aktuelle Forschung ebenfalls viel zu spät, weil sie im traditionellen Rahmen – dem Papier – verhaftet bleibt. Und, wenn die Mehrzahl der Bibliotheken ein nicht mehr bezahlbares Journal im Vertrauen auf elektronische Dokumentenlieferung einfach abbestellt, schließt sich der Teufelskreis, denn die verbleibenden Abonnements müßten, vereinfacht gesagt, mindestens doppelt so teuer werden, weil die Kosten für das erste Exemplar einer Zeitschrift unabhängig von ihrer Auflage sind.

Paradoxerweise bewirkt die Informationsflut einen gleichzeitigen Informationsmangel. Er stellt sich dadurch ein, daß der einzelne – u. a. durch Abbestellungen der Bibliotheken – auf immer weniger Information direkten Zugriff hat.

Chancen und Risiken neuer Technologien

Das zentrale Problem der wissenschaftlichen Informationsversorgung ist: "Wie können Informationsflut und gleichzeitiger Informationsmangel bewältigt werden?". Unsere These ist, daß die Antwort darauf in einem "Übergang in die elektronische Welt" liegt. Überall verfügbare preiswerte Laserdrucker, hochleistungsfähige PCs und Workstations mit Lichtsatz- und Grafiksoftware haben dazu geführt, daß heute fast die gesamte neue Literatur elektronisch vorliegt und mit hoher Qualität gedruckt werden kann, lange bevor sie

in "offiziellen Journalen" erscheint. Immer preiswertere Server mit Speichern im Bereich von Gigabytes und das weltumspannende Internet verwandeln unsere Welt von einer, in der nur einige Verlage wissenschaftliche Arbeiten drucken und vertreiben konnten, in eine, in der praktisch jeder selbst die Verteilung von Artikeln, Büchern, Software etc. organisieren und offerieren kann - zu geringen Kosten, weltweit und effizient. Das *World Wide Web* hat die Wissenschaft im Sturm genommen, weil es so einfach zu benutzen ist und den Austausch eines sehr breiten Spektrums digitaler Information ermöglicht. Es unterstützt das Geben und Nehmen im Internet in fast idealer Weise, im Wissenschaftsbereich auf der Basis der Selbstorganisation.

Ein berühmtes, nicht unumstrittenes Beispiel hierzu ist der Preprint-Server der Theoretischen Hochenergiephysik. Paul Ginsparg hatte dieses System am Los Alamos National Laboratory ursprünglich im Alleingang nur eingerichtet, um einigen Fachkollegen dadurch zu helfen, daß er mittels E-mail die von ihnen eingeschickten Preprints elektronisch an alle Interessenten kostenfrei verteilt. Innerhalb eines halben Jahres nahmen Tausende Physiker aktiv und passiv an diesem Dienst teil. Heute speichert das System eingeschickte Artikel in einem absuchbaren Internet-Server, und verschickt Abstracts an Abonnenten, inzwischen an über 30.000 Personen. Es versorgt nicht nur die gesamte Gemeinschaft der Hochenergiephysiker mit aktuellen Forschungsergebnissen, sondern macht auch in anderen Bereichen der Physik Schule, ebenso in Mathematik und Informatik.

Solche Preprint-Server sind, so spektakulär ihr Erfolg in der Physik auch ist, natürlich nicht ohne Probleme. Jeder kann sie mit Artikeln beschicken, wenn er nur über E-mail verfügt. Diese werden ohne jede inhaltliche Prüfung an Tausende von Fachkollegen verteilt. Wer schützt diese vor Preprints minderer Qualität und wer bewahrt Autoren vor unüberlegten Sendungen? Wer garantiert die langfristige Stabilität dieses kostenfreien Dienstes und wie steht es mit der Archivierung? Trotz all dieser Fragen ist das System in der Physik gut etabliert und für die Community der Hochenergiephysiker unverzichtbar. Die Paper des Ginsparg-Servers werden intensiv gelesen; Revisionen erscheinen oft schon nach wenigen Tagen. Die meisten Spitzenwissenschaftlern nutzen diesen Dienst; so hat z. B. allein E. Witten, der 1990 als erster Physiker die höchste Auszeichnung der Mathematik, die Fields-Medaille, erhielt, seit 1991 mehr als 40 Paper über den Server geschickt. Paul Ginsparg erhält Dankeschreiben von Fachkollegen aus Entwicklungsländern, die ohne diesen Service von der Versorgung mit Fachliteratur fast vollständig abgeschnitten wären.

Technische Bewältigung der Informationsflut

Die für alle Beteiligten sichtbare Publikationsflut ist kein gesteuerter Prozeß. Sie entsteht dezentral und unkoordiniert dadurch, daß sich immer mehr Menschen mit wissenschaftlichen Fragen auseinandersetzen. Die Flut ist nicht aufzuhalten, auch nicht von denjenigen, die sie beklagen. Wollen wir wissenschaftlich relevant arbeiten und zu neuen Ergebnissen gelangen, so bleibt uns nichts anderes übrig, als die Resultate anderer im Auge zu behalten. Die Beherrschung der immer komplexer werdenden Informationsmenge ist mehr und mehr ein zentraler Bestandteil wissenschaftlicher Arbeit. Dabei ist es erfreulich, daß sich die digitalen Informations- und Kommunikationstechnologien besonders dynamisch

entwickeln. Sie können dabei helfen, die Informationsexplosion zumindest rein technisch zu bewältigen.

Das Fassungsvermögen digitaler Speichermedien wächst erheblich schneller als die Literaturproduktion. Es verdoppelt sich zur Zeit alle 18 Monate, bei konstanten Kosten. In 10 Jahren werden digitale Speicher das 100fache der heutigen fassen. Ein Gigabyte Plattenspeicher kostet heute um die 600 DM. Nach einer Modellrechnung für die Mathematik würden alle publizierten Artikel der mathematischen Weltproduktion eines Jahres etwa 10 Gigabyte Plattenspeicher benötigen, und die der nächsten 10 Jahre insgesamt etwa 150 Gigabyte. Solch ein Speicher – er könnte die Hälfte aller jemals publizierten mathematischen Artikel fassen – wird dann für etwa 1000 DM zu haben sein, und er wird vermutlich in einen Laptop passen. Die kommenden digitalen Speichermedien werden auch ausreichen, wenn die einfache Verfügbarkeit neuer Technologien die Mathematiker dazu verführen wird, mehr Multimediale Produkte (qualitativ hochwertige Bilder, Videosequenzen, etc.) zu erstellen. Gleichzeitig werden in 10 Jahren auch die PCs und Workstations 100mal leistungsfähiger sein als die heutigen. Statt die Besteigung immer größer werdender Papierberge zu üben, brauchen wir uns nur mit der Handhabung von Retrieval-Software zu beschäftigen, falls wir konsequent den Weg in die elektronische Welt beschreiten.

Auch das Problem des Transports von Information wird gelöst. Digitale Netze werden ebenfalls immer leistungsfähiger, denn wenn die Leitungen (Glasfaser oder Kupferkabel) erst einmal gelegt sind, hängt ihre Übertragungsrate im Wesentlichen von der Geschwindigkeit ihrer Endgeräte ab – und das sind Computer. Fachleute sagen für solche Kommunikationsrechner noch höhere Leistungssteigerungen voraus als für den Bereich der PCs. Die Netze der Zukunft sollen in 10 Jahren bis zu 1000mal schneller arbeiten als die heutigen, ebenfalls zu denselben Kosten. Solche Netze könnten dann die Informationsmenge der mathematischen Weltproduktion eines Jahres – in anderen Fächern wird es ähnlich sein – in wenigen Sekunden von jedem Ort der Welt zu praktisch jedem anderen übertragen, und zwar zu recht geringen Kosten, denn diese Menge ist für die kommende digitale Welt relativ klein. 10 Gigabyte sind heute ungefähr der Platz, den drei ganz normale Videos in digitaler Form in Anspruch nehmen. Und viel teurer als deren elektronische Lieferung dürfte auch die Übertragung mathematischer Information nicht sein.

Neue Formen der wissenschaftlichen Zusammenarbeit

Auch wenn die schiere Informationsmasse und ihre weltweite Distribution technisch beherrscht werden können, so bleibt doch noch viel zu tun. Das Internet beginnt aber bereits auch, die Art und Weise zu verändern, in der Wissenschaftler miteinander kommunizieren und zusammenarbeiten. Wir demonstrieren dieses an einigen Beispielen aus der Mathematik.

Wissenschaftliche Fachgesellschaften, Fachbereiche, Forschungsinstitute und sogar kleinere Mathematikergruppen können jetzt mit geringem Aufwand eigene elektronische Zeitschriften herausgeben – und sie tun dies auch. In der Mathematik gibt es heute über 30 elektronische Journale, alle referiert, die ihre Artikel kostenfrei im *World Wide Web* offerieren. Einige dieser Journale informieren Abonnenten per E-mail über neue Artikel,

augenblicklich, sobald diese akzeptiert sind. Auch der Prozeß der Begutachtung wird mittels E-mail beschleunigt. Initiativen dieser Art sind erste Beiträge zur Beseitigung von Engpässen im Publikationsbereich. Das wirklich Neue am Internet ist jedoch seine Eigenschaft als digitales Kommunikationsmedium “von vielen an viele” Teilnehmer.

Wissenschaftler tauschen ihre Artikel heute weit im Vorfeld der Publikation aus, als elektronisches “Manuskript”. Sie diskutieren und kommentieren ihre Ergebnisse per E-mail-Verteiler in geschlossenen In-Groups oder gleich in öffentlichen Diskussionsforen. Viele arbeiten an Softwarepaketen oder Büchern, die sie gemeinsam erstellen – auf elektronischem Wege, über große Entfernungen hinweg –, ohne sich häufig persönlich besuchen zu müssen. Mathematische Fachgruppen stellen Preprints und in ihrem Kreise entwickelte Softwaresysteme in öffentlichen Archiven bereit, die man auch absuchen kann, i.d.R. kostenfrei. Sie erstellen, verteilen und archivieren elektronische Newsletter, die über aktuelle Tagungen, Projekte und Ereignisse informieren.

Es mag in unserer Argumentation inkonsequent erscheinen, aber wir rechnen damit, daß die Buchproduktion steigt. Gute wissenschaftliche Bücher sind ausgezeichnete Informationsfilter, sie fassen Wichtiges zusammen, machen schwer Verdauliches erfaßbar und geben Orientierung im “Dschungel der Erkenntnisse”; und, was besonders wichtig ist, gedruckte Bücher haben eine ausgezeichnete “Benutzerschnittstelle”. Zusätzlicher Aufwand durch Bedienung von Rechnern und das Erlernen von Retrieval-Techniken ist nicht notwendig. Es entstehen aber bereits neue Formen der Filterung und Wissenskonsolidierung wie z.B. *dynamic* oder *ongoing surveys*, also Übersichtsartikel, die im Internet erscheinen und die regelmäßig von mehreren Autoren aktualisiert werden. Es kündigen sich umfassendere Vorhaben wie *dynamische Handbücher* an. Weltweit verteilte und lose verbundene Forschergruppen bieten kommentierte Bibliographien zu aktuellen Forschungsthemen an, die laufend fortgeschrieben werden.

Jede Institution, die etwas auf sich hält, erscheint mit einem eigenen Server im Internet (oft einem kompletten Informationssystem), allen voran die großen mathematischen Fachgesellschaften wie z. B. die International Mathematical Union (<http://elib.zib-berlin.de/IMU>) oder die Deutsche Mathematiker-Vereinigung (<http://www.dmv.math.tu-berlin.de/>). Wissenschaftler sind heute mit persönlichen Seiten – wissenschaftlichen “Visitenkarten” – im Internet präsent, viele mit ihren Veröffentlichungen und beruflichen Lebensläufen. Etliche Institutionen und auch Einzelpersonen sammeln solche Informationen, verdichten sie zu Übersichtslisten oder umfangreichen Informationssystemen und stellen sie der Internet Community zur Verfügung. Dabei setzen sie auch “Informationsroboter” ein, die das Internet automatisch nach neuen Servern und Informationsressourcen absuchen. Es wird bereits ernsthaft über die Einrichtung eines elektronischen weltweiten Archivs für mathematische Literatur diskutiert. Das alles beruht auf persönlichen Aktivitäten und Initiativen auf breiter Front.

Das Internet entstand (initiiert durch ARPA und gefördert durch die NSF der USA) an Forschungslabors im Wissenschaftsbereich. Es trägt, bei starken Kommerzialisierungstendenzen, auch heute noch alle Charakteristiken eines Wissenschaftsnetzes. Der möglichst freizügige Zugang zu jeglicher Information ist fester Bestandteil der Internet-Kultur des

“Gebens und Nehmens”. Die Verheißung des Internet ist die “Globale Bibliothek”, der freie Zugriff auf der “Wissen der Welt”, von jedem Ort der Welt und zu jeder Zeit.

Wandel durch Selbstorganisation und Kooperation

Das Internet ist in den USA das Vorbild für die Datenautobahnen der kommenden Informationsgesellschaft. In einer Sendung des ZDF zu diesem Thema war kürzlich Donald Trump vor einem seiner Bürohochhäuser in Manhattan zu sehen, die er in Wohnsilos umbauen ließ. Etwa ein Drittel der Bürofläche in Manhattan steht leer und ist kaum vermietbar. Viele Firmen sind auf’s Land gezogen, dank Internet. Die Redaktion stellte schon die Frage: “Ist mit den Datenautobahnen das Ende der Hochhäuser gekommen?”. Die Presse berichtet fast wöchentlich von neuen spektakulären Anwendungen des Internet, wie z.B. über weltweite Videosendungen, internationales Telefonieren “zum Ortstarif” etc. Die deutsche Telekom prognostiziert umwälzende gesellschaftliche Veränderungen, z.B. durch Teleworking und Telemedizin. Große Erwartungen werden geweckt. Eli M. Noam sagt in seinem Science-Artikel “Electronics and the Dim Future of the University” (Vol. 270, 13. Oktober 1995, 247–249) den schleichenden Tod der klassischen Universität durch die neuen Informationstechnologien und insbesondere durch Veränderungen im Informationsfluß und in der Informationsverarbeitung voraus.

Jedoch, gesellschaftliche Veränderungen benötigen Zeit. Die Ausbildung entsprechender organisatorischer und rechtlicher Strukturen hinkt hinter der rasanten Entwicklung neuer und immer leistungsfähigerer Technologien hinterher. Hier sind die Wissenschaften vor eine besondere Verantwortung gestellt. Qualität und Zuverlässigkeit wissenschaftlicher Information ergeben sich nicht von selbst. Lassen sich traditionelle Verfahren der Begutachtung und des Referierens in die elektronische Welt übertragen? Nichtreferiertes Material, wie z.B. institutionelle oder persönliche Seiten im *World Wide Web*, oder auch Software, spielen eine immer größere Rolle. Welche organisatorischen Mechanismen sichern ihre Qualität? Wie können Authentizität und Identifizierbarkeit von Publikationen aller Art garantiert werden? Und, wie steht es mit den Urheberrechten der Autoren und der an den Forschungsaufgaben beteiligten Institutionen? Wer entscheidet inhaltlich, was langfristig zu bewahren und zu archivieren ist, und wie ist dieses technisch zu meistern?

Selbst einfache und eher technisch erscheinende Probleme harren noch ihrer Lösung. Auch wenn es bereits recht akzeptable Suchmaschinen und -mechanismen im Internet gibt, ist die gezielte Lokalisierung bestimmter Informationen, wie etwa neuer Artikel zu einem spezifischen Sachgebiet, nach wie vor schwierig. Sowohl die Fülle als auch die Heterogenität der im Internet vorhandenen Information verhindern oft eine genauere Eingrenzung bei der Suche. Es gibt inzwischen Informationsroboter, die den Anspruch erheben, das ganze *World Wide Web* nach Schlüsselworten zu indizieren; jedoch wird keiner dieser Roboter solche Information automatisch klassifizieren oder gar “verbessern” können. Das Mittel gegen Informationsüberflutung ist wirksame Filterung, z. B. durch qualitative Selektion und Einbringung von Strukturinformation. Kann ein Automatismus dieses jemals leisten? Nach wie vor ist mühsame Handarbeit notwendig. Schon das Auffinden eines speziellen elektronischen Preprints in einem einfachen Internet-Server kann ein Problem sein, wenn man diesen nicht kennt und jeder Server anders aufgebaut ist.

Hier öffnet sich ein weites Feld für wissenschaftliche Fachgesellschaften und Institutionen. Sie sollten nicht nur die durch die neuen Technologien gegebenen Chancen wahrnehmen und nutzen, sondern diese Entwicklung und ihre Anwendung selbst mitbestimmen und vorantreiben. Insbesondere sollte das Zusammenspiel zwischen den Wissenschaften mit ihren Bibliotheken, den Verlagen und Softwarehäusern, dem Buchhandel und den Zentren der Fachinformation neu gestaltet und den zukünftigen Verhältnissen gerecht werden. Es gilt – gemeinsam – Kompetenz zu entwickeln und zu wahren.

Vier wissenschaftliche Fachgesellschaften, die Deutsche Mathematiker-Vereinigung (DMV), die Deutsche Physikalische Gesellschaft (DPG), die Gesellschaft Deutscher Chemiker (GDCh) und die Gesellschaft für Informatik (GI) haben sich zur Förderung aller Belange der elektronischen Information und Kommunikation (IuK) in Deutschland entschlossen und eine Kooperationsvereinigung abgeschlossen. Es geht ihnen um die Neugestaltung der Informationslandschaft im Dialog mit allen Beteiligten auf lokaler, nationaler und auf internationaler Ebene. Den Gesellschaften ist natürlich klar, daß effiziente Informationsversorgung und -verarbeitung nicht automatisch bessere Forschung implizieren; gute Wissenschaft ist allerdings ohne sie selten möglich. DMV, DPG, GDCh und GI sind davon überzeugt, daß auf der Basis der in diesem Artikel skizzierten Technologien und damit verbundener Software neue Informationsinfrastrukturen für die Wissenschaften geschaffen werden können, die die oben angesprochenen Fragen zufriedenstellend lösen.

Allerdings wird die globale, effizient absuchbare, digitale Bibliothek der Wissenschaften nicht in wenigen Jahren errichtet werden können. Strukturwandel stößt immer auf Widerstand, selbst dann, wenn er als unabwendbar erkannt wird (siehe Kohlebergbau, Werften). Politische Unterstützung des Wandels wäre wünschenswert, damit sich der notwendige Veränderungsprozeß nicht zu lange hinzieht. Auch wenn die Bundesregierung in ihrem (noch nicht verabschiedeten) Programm “Wissenschaftlich-technische Information für das 21. Jahrhundert” die Situation richtig analysiert, bleibt doch abzuwarten, ob daraus auch die angemessenen Konsequenzen gezogen werden.

Dieser Artikel ist ein kurz gefaßter Querschnitt durch einige unserer Arbeiten zu diesem Thema, die ebenso wie die o.g. Kooperationsvereinbarung elektronisch verfügbar sind (<http://elib.zib-berlin.de/math.org.softinf.pub>).

Anschrift der Autoren:

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)
Heilbronner Str. 10, 10711 Berlin
E-mail: groetschel@zib-berlin.de