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

M. Grötschel, J. Lügger und U. Zimmermann

# Wissenschaftliche Information am Wendepunkt?
# Zwänge, Krisen und Chancen
# aus Sicht der Mathematik
# Kurzfassung

Technical Report TR 95–08 (Juni 1995)

# Wissenschaftliche Information am Wendepunkt? Zwänge, Krisen und Chancen aus Sicht der Mathematik Kurzfassung

M. Grötschel*
J. Lügger*
U. Zimmermann†

20. Juni 1995

## 1 Einführung

Unmittelbares Ziel der wissenschaftlichen Forschung ist die Gewinnung neuer Erkenntnisse, also originärer Information zum jeweiligen Forschungsobjekt. Der Wissenschaftler muß dazu auf den bisherigen Erkenntnissen aufbauen, er benötigt einen entsprechend effektiven Zugang zu den Informationen, die andere Wissenschaftler bereits produziert haben. Der Wissenschaftler ist Produzent und erster Nutzer der wissenschaftlichen Information. Produktion und Kommunikation der wissenschaftlichen Information sind unauflösbar miteinander verwoben und voneinander abhängig. Durch verschiedene schnell und gleichzeitig ablaufende Prozesse ist dieses sensible System derzeit gefährdet. Mit der "explosiv anschwellenden Informationsmenge" trifft eine "finanzielle Budgetkrise an Bibliotheken" zusammen, die durch "ständig wachsende Zeitschriftenpreise" verschärft wird. Kann man durch Steigerung der Effizienz des bisherigen Publikationssystems der offenbaren Krise Herr werden oder ist es notwendig, den traditionellen Publikationsweg unter Einbeziehung aller Mittel der entstehenden Informationstechnologie vollständig zu überdenken?

Alle am Prozeß Beteiligten sind sich der Krise mehr oder minder bewußt. Wissenschaft, Bibliotheken, Verlage, Fachinformationszentren, jeder sucht seinen eigenen Ausweg, um die Entwicklung in den Griff zu bekommen. Auch die Politik, die die Wissenschaft und damit insbesondere den gesamten Publikationsprozeß zum Nutzen der Gesellschaft finanziert, beginnt, Informationen und Konzepte zu sammeln, und versucht, eine "Informationspolitik" für ein weiterhin finanzierbares System zu formulieren.

## 2 Der Produktionszyklus wissenschaftlicher Information

Wenn man zunächst einmal alle Möglichkeiten der Veränderung offenläßt, gilt es, Klarheit über das Ziel und eine korrekte Darstellung der Produktion wissenschaftlicher Information zu erhalten. Das unmittelbare Ziel, die Gewinnung originärer Information zum Forschungsobjekt, stimmt offenbar mit der Aufgabenstellung des Wissenschaftlers überein.

Das mittelbare Ziel, die kurz- oder langfristige Umsetzung der gewonnenen Erkenntnisse zum Nutzen der Gesellschaft, kann erst nach der Produktion der Information anvisiert werden - natürlich

*Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Heilbronner Str. 10, D-10711 Berlin-Wilmersdorf
†TU Braunschweig, Abteilung für Mathematische Optimierung, Pockelsstraße 14, D-38106 Braunschweig, Germany

kann der Staat durch Förderung Schwerpunkte in den Forschungsaktivitäten setzen und die Umsetzung beschleunigen. Die hierbei notwendigen und wichtigen Filter- und Übersetzungsprozesse in Form von Sekundärliteratur, wissenschaftlichen Büchern bis hin zu populärwissenschaftlichen Artikeln, etc. gehören aber nicht zum inneren Kern des wissenschaftlichen Produktionsprozesses.

Wir beschränken uns hier auf die Primärliteratur, wie sie traditionell in Form von Zeitschriftenartikeln und in ähnlicher Weise in Proceedings vorliegt. Unsere Sicht ist dabei geprägt von unseren Erfahrungen im eigenen Fach, der Mathematik, läßt sich aber in vielerlei Hinsicht mit entsprechenden Modifikationen auf andere wissenschaftliche Fächer übertragen.

Phase I: Der Wissenschaftler verfaßt in zeitintensiver Arbeit ein Manuskript, wobei er sorgfältig die vorhandene Literatur studiert und in Kommunikation mit seinen Fachkollegen absichert, so daß er ein aus seiner Sicht originäres Resultat gefunden hat. Diese Rohfassung der Information wird entweder durch eine Sekretärin im Institutssekretariat oder durch den Wissenschaftler selbst in eine publikationswürdige Form gebracht. Hierbei werden Standardtextsysteme wie TEX oder LATEX benutzt, die heute bereits im traditionellen System benutzt werden, so daß das fertiggestellte Manuskript in vielen Fällen ohne weitere Kontrolle eines Desk-Editors in Druck gehen könnte. Als Preprint wird es meist den engeren Fachkollegen zugänglich gemacht, um die wissenschaftliche Kommunikation zu beschleunigen.

Phase II: Dasselbe Manuskript geht an den Herausgeber einer wissenschaftlichen Zeitschrift. Dieser ist selbst ein Wissenschaftler, der von einem Verlag oder einer wissenschaftlichen Gesellschaft allein oder in Kooperation mit einem Verlag bestellt wird. Dies erfolgt meist durch einen Beirat aus weiteren Fachwissenschaftlern.

Der Herausgeber startet den Referentenzyklus, in dem das Manuskript unter Zwischenschaltung von Mitherausgebern an mehrere Referenten zur fachlichen Begutachtung gesendet wird. Mit den Gutachten wird die Arbeit an den Autor zurückgereicht. Wenn keine Ablehnung erfolgt, durchläuft die Arbeit in mindestens 50% der Fälle ein weiteres Mal den Zyklus, wobei die Vorschläge der Referenten in das Manuskript eingearbeitet werden. Die angenommenen Manuskripte liegen nahezu vollständig in einer publizierfähigen Form vor. Der Herausgeber stellt sie zu Heften der Zeitschrift zusammen und sendet alles an den Verlag.

Phase III: Der Verlag kann mittels eines Setzprozesses eine optische Aufbereitung vornehmen, sprachliche Korrekturen anbringen oder die Arbeiten drucken, wie sie sind. Eine Aufbereitung muß dem Autor in gedruckter Form nochmals vorgelegt werden. Unabhängig davon liegen die Hefte dann in einer Warteschlange, da die Seitenzahl der Zeitschriften abhängig von der Preisgestaltung begrenzt ist und Sorge für eine kontinuierliche Erscheinungsweise zu tragen ist, u.a. in dem stets ein gewisser Vorrat an fertigen Heften besteht. Für das gedruckte Heft liegen Werbung und Vertrieb beim Verlag, der es den wissenschaftlichen Bibliotheken zum Kauf anbietet.

Phase IV: Was geschieht nun mit der gedruckten Version der Information? Die Bibliotheken entscheiden in Absprache mit den Fachwissenschaftlern, welche Zeitschriften erworben werden und sorgen für die Archivierung und Erschließung der Zeitschriften. Das schwierige Wiederauffinden relevanter Information in den Bibliotheken wird unterstützt durch entsprechende Informationssysteme und die Fachinformationszentren, die dazu die Inhaltsverzeichnisse der Zeitschriften durchgehen und erfassen. Dieses "Textretrieval" wird insbesondere durch kurze Zusammenfassungen von Autoren oder geeigneten Fachwissenschaftlern unterstützt, die zu den einzelnen Artikeln durch die Herausgeber von Referatejournalen angefordert werden. In elektronischer Form sind die so entstehenden Datenbanken prinzipiell eine hervorragende Grundlage beim Literaturstudium des Wissenschaftlers, der ein Manuskript verfaßt. Womit der Produktionszyklus wieder beginnt.

Mit Ausnahme von Druck und Verkauf liegt die Produktion der wissenschaftlichen Information bei den Wissenschaftlern, die in allen Phasen die entscheidenenden Beiträge leisten. Archivierung und Erschließung ist dann Aufgabe der Bibliotheken. Die Unterstützung des Textretrieval erfolgt über die Fachinformationszentren. Der Staat finanziert den ganzen Zyklus, auch den Kauf der gedruckten Version der Information von den Verlagen. Dadurch sichert er, daß wissenschaftliche Information allen, insbesondere den Wissenschaftlern selbst für ihre Aufgaben im wesentlichen frei wieder zugänglich ist.

Wo liegen nun die Wurzeln der Krise des derzeitigen Systems? Kurz und pointiert: die Informationsmenge wächst explosiv, die relativen Kosten wachsen bei fallender oder bestenfalls stagnierender staatlicher Finanzierung, und der Produktzyklus ist viel zu lang.

Die Wissenschaftler spüren die resultierenden Schwierigkeiten bei der Informationsbeschaffung inzwischen sehr deutlich. Auf der Grundlage einer Umfrage an den mathematischen Fachbibliotheken (cf. [Winter 95]) wird in Hinblick auf die aus den Budgetproblemen resultierenden inzwischen drastischen jährlichen Abbestellungen gefragt, wie lange die wissenschaftlichen Verlage die Herausgabe von Zeitschriften in der bisherigen Form aufrechterhalten können. Für die aktuelle Forschung kommen die Zeitschriftenartikel, die nach Eingang beim Herausgeber meist 2-3 Jahre bis zum Stellplatz in den Bibliotheken benötigen, in jedem Fall zu spät. Die Aufnahme der Artikel in die Referatejournale liegt noch ein Jahr später, so daß ein aktueller Überblick zu einem Thema auf dem Weg über die gedruckten Veröffentlichungen überhaupt nicht möglich ist.

Das gedruckte Exemplar dient eigentlich im wesentlichen als Publikationsnachweis mit Qualitätskontrolle und als traditionelle Grundlage für die Langzeitarchivierung. Wenn es dann von einem Wissenschaftler nach geglückter Recherche in einer Datenbank schließlich in einer fernen Bibliothek im Archiv nachgewiesen wird, muß erst eine elektronische oder gedruckte Kopie erstellt und auf den Weg zu ihm gebracht werden, was mit Zeitaufwand und teilweise erheblichen Kosten pro Kopie verbunden ist. Je weniger gedruckte Exemplare in den Bibliotheken zu finden sind, um so aufwendiger wird diese Art der Literaturversorgung, d.h. auf dem Wege der Dokumentenlieferung, ist langfristig weder das Zeitproblem noch das Finanzproblem zu lösen.

In der aktuellen Forschung ist die Bereitschaft der Wissenschaftler zum Einsatz der zu geringen Kosten nutzbaren elektronischen Netze unzweifelhaft gestiegen, obwohl die zur Verfügung stehenden Hilfsmittel kaum als nutzerfreundlich eingestuft werden können. Das elektronische Netz (Internet) bietet einen attraktiven, kurzfristigen Ausweg aus der Krise, der Kommunikation mit Kollegen, Verfügbarkeit von Preprints unmittelbar nach Manuskripterstellung und die Nutzung von Informationsservern erlaubt, ohne daß damit bereits ein sicherer, auf Dauer tragfähiger Weg aus der Krise bekannt ist. Zwar können sich damit die in Diskusssionszirkeln befindlichen Wissenschaftler der Universitäten helfen, aber Forschern in Industrie und Wirtschaft und dem jungen wissenschaftlichen Nachwuchs sind diese Quellen kaum zugänglich.

Verschiedenene Fachgesellschaften, in Deutschland die Chemiker, Informatiker, Mathematiker und Physiker, haben sich daher entschlossen, gemeinsam solche Wege zu suchen, die zu einer langfristigen Sicherung eines schnellen und zuverlässigen Informationszugangs führen. Sie tun das weder mit besonderer Freude noch mit wissenschaftlicher Motivation, sie sehen aber die Notwendigkeit, essentielle Grundlagen für die zukünftige Forschung zu sichern. Dabei sind sie offen für jede zweckdienliche Kooperation mit den bisher am Produktionszyklus Beteiligten, inbesondere den Bibliotheken, Fachinformationszentren und Verlagen.

## 3 Elektronische Journale und Proceedings

Wenn man sich überlegt, ob mit der vorhandenen und sich derzeit entwickelnden Informationstechnologie Auswege aus der Krise möglich sind, tut man gut daran, einmal zu sehen, was es alles schon gibt. Auf die seit langem bekannten, traditionellen und bislang bewährten gedruckten Medien muß hier nicht genauer eingegangen werden.

Bei den sich abzeichnenden Neuerungen sind die elektronischen Journale und Preprintserver am auffälligsten. Ein Preprintserver ermöglicht die schnelle Verfügbarkeit neuer Informationen über den Kreis der Insider hinaus. Allerdings birgt die fehlende Qualitätskontrolle für den Nichtspezialisten ein weit größeres Risiko. Die elektronischen Journale dagegen unterscheiden sich von den traditionellen Zeitschriften nur in der Form der Information nicht in Ihrer Qualität. Die Wahl der elektronischen Form verbietet auch keineswegs eine gedruckte Version. Wie aus dem obenbeschriebenen Produktionszyklus ersichtlich, liegt heutzutage dem gedruckten Artikel meist ein elektronisches Manuskript zugrunde. Die elektronische Form bietet also eine Alternative, deren Möglichkeiten es auszuschöpfen gilt. Für die Erstellung von Tagungsberichten (Proceedings) gelten weitgehend analoge Überlegungen, so daß wir hier stets nur von Zeitschriften reden werden.

Reicht der Wissenschaftler am Ende der Phase I sein Manuskript bei dem Herausgeber einer elektronischen Zeitschrift ein, so beginnt ein analoger Referentenzyklus. Zusätzlich kann der Artikel in der jeweils aktualisierten Form als Preprint in einem Preprintserver der Zeitschrift stehen und automatisch zum Retrieval registriert werden, wenn der Autor es wünscht. Sobald die Arbeit angenommen ist, steht sie in der Zeitschrift via Netz zur Verfügung und kann wiederum automatisch in einem Referatejournal erfäßt werden.

Im Vergleich zum traditionellen Produktionszyklus entstehen den Wissenschaftlern im wesentlichen keine zusätzlichen Aufgaben, aber die Kooperation mit den anderen Beteiligten muß neu organisiert werden. Hierbei können insbesondere den wissenschaftlichen Bibliotheken sowohl bei der Langzeitarchivierung als auch bei der Sicherung der Verfügbarkeit weitere Aufgaben zuwachsen (cf. [Grötschel 95-1]). Welche Unterschiede entstehen für den Wissenschaftler als Nutzer, der auf die Informationen zugreifen will? Er kann sofort nach erfolgter Recherche, die mit dem Namen des Journals auch den elektronischen Standort des Journals liefert, eine elektronische Kopie über das Netz erhalten. Damit wird der Produktionszyklus entscheidend verkürzt.

Das allen Schwierigkeiten zugrundeliegende rasche Anwachsen der Informationsmenge wirkt bei elektronischer Publikation weit weniger gefährlich, da die Speicherung und Archivierung der resultierenden gedruckten Versionen mehr oder minder entfällt. Die Gesamtkosten für den Staat als Geldgeber sinken an verschiedenen Stellen des Produktionszyklus. Wenn man davon ausgeht, daß der Aufbau und Betrieb der notwendigen Netze, durch anderweitige kommerzielle Nutzung getragen wird, werden die Gesamtkosten sicherlich geringer ausfallen. Natürlich kann man über die Verteilung der Kosten nachdenken, was in einem föderativ organisierten Staat wie Deutschland zu einigen Auseinandersetzungen führen kann und wird.

In Hinblick auf diese entscheidenden Vorteile, die elektronische Zeitschriften bieten, sind wir der Meinung, daß parallel zum Ausbau der Netze elektronische Zeitschriften die gedruckten Versionen ablösen werden. Die meisten Zeitschriftenartikel wenden sich an einen sehr kleinen Kreis von potentiellen Lesern (Spezialisten). Nur wenn notwendig wird sich der Nutzer, bei dem dann die Druckkosten anfallen, eine gedruckte Kopie erzeugen. Dies ist im übrigen derselbe Nutzer, der sich heute seine persönliche Kopie des gedruckten Bibliotheksexemplars auf den Schreibtisch holt. Eine Reduktion der vielen heutzutage hergestellten Druckexemplare, die niemals von einem Nutzer angefaßt werden, ohne daß die Zugriffsmöglichkeiten für potentielle Nutzer eingeschränkt werden, trägt u.a. zur Kostenreduktion bei.

Über den prinzipiellen Vorteilen der elektronischen Zeitschriften soll nicht verschwiegen werden, daß hier neben dem unerläßlichen Netzausbau noch viel Entwicklungsarbeit und Neuorganisation durch die am Produktionszyklus Beteiligten notwendig ist. Dies sollte unserer Meinung nach in einem umfassenderen Ansatz erfolgen, der über die Zeitschriftenproblematik weit hinausgeht.

# 4 Stand der Entwicklung verteilter Informationssysteme

National und international werden zur Zeit Konzepte entworfen, Methoden entwickelt und getestet, um eine effiziente, zukünftige Informationslandschaft zu gewährleisten und dabei den verschiedenen Optionen Rechnung zu tragen. Für Länder, deren Industrie und Wirtschaft auf innovativen technischen Spitzenprodukten fußt, und für Wissenschaften, die in diesen Bereichen Schlüsseltechnologien vorbereiten, entwickeln oder nutzen, ist es lebensnotwendig, an diesem Gestaltungsprozeß aktiv teilzunehmen. Anderenfalls besteht die Gefahr, daß die Entwicklung an den eigenen Informationsbedürfnissen vorbeigeht und daß die Betroffenen nicht in der Lage sind, effektiv die neuen Werkzeuge und Informationsquellen zu nutzen. Insbesondere die Deutsche Mathematikervereinigung (DMV) hat im Anschluß an ein bereits länger laufendes Fachinformationsprojekt ein Konzept entworfen, das ein verteiltes Informationssystem für die Mathematik beinhaltet und die verschiedenen zukünftigen Möglichkeiten integriert (cf. [Grötschel 95-2]). Für eine mehr technisch geführte Diskussion verweisen wir auf [Dalitz 94]. Eine angemessene Darstellung dieses umfangreichen Vorhabens ist im gegebenen Rahmen nicht möglich. Die Zielsetzung in [Grötschel 95-2] soll aber zitiert werden:

> Das *langfristige Ziel* ist der Aufbau eines verteilten, kostengünstigen und nutzergesteuerten elektronischen Informations– und Kommunikationssystems, das den effizienten Zugang zu qualitativ hochwertiger Volltext–, Fakten–, Literaturhinweis– und Software–Information im Bereich der Mathematik von Arbeitsplatzrechnern aus ermöglicht.

Diese langfristige Zielsetzung läßt sich wie viele der einzelnen Ziele des Vorhabens auch auf andere Wissenschaften übertragen, wobei eine Erweiterung in Richtung graphischer multimedialer Information notwendig sein kann. Wie bereits oben erwähnt, hat daher die DMV frühzeitig mit anderen wissenschaftlichen Fachgesellschaften Kontakte geknüpft, um auf lokaler, nationaler und internationaler Ebene die Entwicklung besser koordinieren und vorantreiben zu können.

Wesentlich neben der Diskussion der strukturellen inhaltlichen Formen und die adaequaten Entwicklung und Bereitstellung der technischen Werkzeuge ist der Aufbau einer angemessenen organisatorischen Infrastruktur. Neben der hierbei in der Übergangsphase erforderlichen finanziellen Unterstützung wird der Erfolg im wesentlichen von der persönlichen lokalen Unterstützung und Akzeptanz in den Wissenschaften abhängen. Wir sind überzeugt, daß gerade verteilte Systeme, die nahe bei den Betroffenen ansetzen, den einzelnen Wissenschaftler weit besser motivieren, als dies durch zentrale Systeme möglich wäre. Die technischen Erfordernisse stehen hier in selten glücklichem Einklang mit dem entscheidendem menschlichen Faktor. Unsere positive Einschätzung wird nicht zuletzt durch die rasant gewachsene Nutzung der internationalen elektronischen Netze gestützt, deren rascher Ausbau aus den verschiedensten Motiven deutliche politische Unterstützung findet. Die dadurch gegebene Chance zur Verbesserung der Informationsmöglichkeiten muß genutzt werden.

**Literatur**

[Dalitz 94] W. Dalitz, M. Grötschel, J. Lügger, W. Sperber; Neue Perspektiven eines verteilten Informationssystems für die Mathematik; GAMM Rundbrief; Sept. 1993; auch Report TR-94-7, ZIB-Berlin, August 1994; elektronisch verfügbar via URL ftp://elib.zib-berlin.de/pub/zib-publications/reports/TR-94-7

[Grötschel 95-1] M. Grötschel, J. Lügger; Wissenschaftliche Kommunikation am Wendepunkt - Bibliotheken im Zeitalter globaler elektronischer Netze -; April 1995; ZIB-Berlin; elektronisch verfügbar via URL ftp://elib.zib-berlin.de/pub/zib-publications/reports/TR-95-01

[Grötschel 95-2] W. Dalitz, M. Grötschel, G. Heyer, J. Lügger, W. Sperber; Ein verteiltes Informationssystem für die Mathematik, Beschreibung eines Vorhabens der Deutschen Mathematiker-Vereinigung (DMV); April 1995; ZIB-Berlin; vorläufige Version eines geplanten Antrags der DMV an das BMBF, Bonn.

[Winter 95] Thomas Winter, Uwe Zimmermann; Zur Entwicklung des Zeitschriftenbestandes an mathematischen Fachbibliotheken DMV-Mitteilungen, Heft 1, 1995, pp 34-42; elektronisch verfügbar über URL http://moa.math.nat.tu-bs.de/service/preprints/auswertung.ps.gz