

# Mehrwert von bibliographischen Datenbanken

Urs Schoepflin

MPI für Bildungsforschung  
Lentzeallee 94  
D-14195 Berlin, Deutschland  
e-mail: schoepflin@mpib-berlin.mpg.d400.de

Wolfgang Glänzel

Bibliothek der UAW  
Forschungsinstitut für Wissenschaftsinformation und Szientometrie  
H-1361 Budapest Pf. 7, Ungarn  
email: h4324gla@ella.hu

---

## Inhalt

- 1 Einleitung
- 2 Datenbanken und ihre Elemente
- 3 Retrieval und Datenanalyse im Überblick
- 4 Mehrwert bibliographischer Datenbanken durch bibliometrische Nutzung
- 5 Anforderungen an Struktur und Nutzung von Datenbanken
- 6 Schlußfolgerungen

## Zusammenfassung

Betrachtet werden die Anforderungen, die sich aus der veränderten Nutzungsperspektive konventioneller bibliographischer Datenbanken durch bibliometrisches Retrieval ergeben und die Rückwirkungen auf Konstruktion, Struktur und Nutzung von Datenbanken, die diesen neuen Ansprüchen genügen sollen. Der eigentliche Mehrwert besteht in der Möglichkeit der Darstellung von Strukturen, Zusammenhängen und deren Veränderungen, die mit anderen Verfahren nicht oder nur zum Teil sichtbar gemacht werden können.

Der informationelle Mehrwert, der heute in Form von wissenschaftsmetrischen Strukturanalysen aus den online und auf CD-ROM verfügbaren Massendatenbasen geschöpft wird, läßt sich deshalb zunächst auch als Zielkonflikt darstellen zwischen den traditionellen Datenbanken und den Ansprüchen, die mit ihrer bibliometrischen Nutzung verbunden sind. Bei der bibliometrischen Nutzung von Datenbanken machen sich eine Reihe von Grenzen und Schranken geltend. Gleichzeitig wirken die Ansprüche der Bibliometrie bei der Benutzung der Datenbanken auch auf ihre Herstellung zurück. Dies hat zur teilweisen Verbesserung der Datenbasen und schließlich zu eigentlichen bibliometrischen "Sekundärdatenbanken" geführt. Auf der Seite des Retrieval lassen sich ebenfalls wesentliche Verschiebungen feststellen. Bibliometrisches Retrieval bedeutet ein qualitativ und quantitativ neues Niveau der Datenanalyse, die durch gleichzeitigen Zugriff auf bis vor wenigen Jahren unvorstellbar große Datenmengen gegeben ist. Die Charakteristika des bibliographischen und bibliometrischen Retrieval werden systematisch beschrieben und gegenübergestellt.

Anhand von Beispielen wird die Bedeutung des bibliometrischen Retrieval in bibliographischen Datenbanken zur Gewinnung des "informationellen Mehrwerts" deutlich gemacht und auch die Rückwirkung der beschriebenen Verfahren auf die konventionellen Abfragetechniken mitreflektiert.

Schließlich werden die neuen Anforderungen an die Datenbanken und an deren Nutzung auf zwei Ebenen beschrieben: Diese betreffen sowohl die Seite der strukturellen Aufbereitung der Datenbasen selbst als auch die Seite des eigentlichen Retrieval und der Weiterverarbeitung bzw. Interpretation der Ergebnisse.

## Abstract

The requirements for conventional bibliographic databases resulting from the changed user perspectives of bibliometric retrieval are considered as well as the retroactive effects on the construction, structure and usage made of databases which meet these new requirements. The actual added value originates in the possibility to representing structures, linkages and their dynamics which could not or only in part be made apparent with other methods. The value added information - in form of structural scientometric analyses in large online and CD-ROM-available databases - can be described as a target conflict between conventional databases and the requirements connected to their bibliometric use.

The characteristics of bibliographic vs. bibliometric retrieval are systematically described and compared.

The limitations and restrictions of the bibliometric use of databases are summarised. The requirements of bibliometrics for database usage effect retroactively the production of the databases themselves which has led to partial enhancements but also to the design of actual "secondary databases" for bibliometric purposes only.

Several examples are given of the significance of bibliometric retrieval for the information added value and the retroactive effects on conventional retrieval techniques are reflected. Finally, the new requirements for databases and their usage are summed up on two levels: first on the level of database structures and second on the level of the retrieval and the problems of interpretation of the findings themselves.

## 1 Einleitung

Aus der veränderten Nutzungsperspektive konventioneller bibliographischer Datenbanken durch bibliometrisches Retrieval ergeben sich Anforderungen auf Konstruktion, Struktur und Nutzung von Datenbanken, die diesen neuen Ansprüchen genügen sollen. Der eigentliche Mehrwert entsteht in der Möglichkeit der Darstellung von Strukturen, Zusammenhängen und deren Veränderungen, die mit anderen Verfahren nicht oder nur zum Teil sichtbar gemacht werden können.

Im Zusammenhang mit der bibliometrischen Nutzung von Datenbanken kann von einer eigentlichen "Zweckentfremdung" insofern gesprochen werden, als die gängigen Datenbanken nach ihrer Entstehungsgeschichte (z.B. disziplinentrierte Bibliographien), ihrer Datenstruktur und nach ihrem primären Verwendungszweck, nämlich auf das Einzeldokument bezogenes Retrieval, nicht dafür vorgesehen sind.

Der "informationelle Mehrwert", welcher heute in Form von wissenschaftsmetrischen Strukturanalysen aus den online und auf CD-ROM verfügbaren Massendatenbasen gewonnen wird, läßt sich deshalb zunächst auch als ein Zielkonflikt darstellen zwischen den traditionellen Datenbanken und den Ansprüchen, die mit ihrer bibliometrischen Nutzung verbunden sind. Dieser Konflikt äußert sich z.B. in der Bedeutungsverschiebung einzelner Datenfelder: bibliographische Kernfelder wie "Autor" oder "Titel" interessieren nicht mehr als einzelne, relevant werden sie erst indem mit ihrer Hilfe Verbindungen zwischen den Datenelementen hergestellt werden können, also etwa zwischen Adressen, Topoi etc. Die Angaben zum Einzeldokument dienen letztlich nur noch als quantitativer Hintergrund zur Darstellung von strukturellen Beziehungen und deren Dynamik. Bevor die verschiedenen Ebenen beim bibliographischen und bibliometrischen Retrieval gegenübergestellt und detaillierter untersucht werden, sollen die inhaltlichen und technischen Möglichkeiten und Beschränkungen bei der bibliometrischen Nutzung von bibliographischen Datenbanken umrissen werden.

## 2 Datenbanken und ihre Elemente

Bei der bibliometrischen Nutzung von Datenbanken machen sich eine Reihe von Grenzen und Schranken geltend. Diese Beschränkungen lassen sich auf drei Ebenen zusammenfassen:

1. Technische Schranken auf der Dokumentenebene durch die Quellen selbst (fehlende oder unvollständige Angaben - z.B. Adressen, Referenzen - in den Zeitschriften); geopolitische Beschränkungen (geringe absolute Publikationszahlen in einzelnen Disziplinen für einzelne Länder oder Regionen), deren Ursache ebenfalls bei den nachgewiesenen Quellen liegen (Winterhager 1993).
2. (Inhaltliche) Schranken in der Repräsentativität, also bei der Dokumentenauswahl; inwiefern sind die Kommunikationsgewohnheiten und -Präferenzen einzelner *scientific communities* wirklich in den Datenbanken abgebildet? (Schoepflin 1992)
3. (Technische) Schranken durch fehlerhafte Datenerfassung und -aufbereitung (ungenügende Normierung, Auslassung relevanten Datenmaterials, Schreibfehler, aber auch zeitliche Lücken etc.) (Jacsó, 1993a,b).

Gleichzeitig wirken die Ansprüche der Bibliometrie bei der Benutzung der Datenbanken auch auf ihre Herstellung zurück. Um den gestiegenen Anforderungen zu genügen muß ein erhöhter Aufwand für die Fehlerkorrektur geleistet werden, die Normierung wird verbessert, eine Vervollständigung der Angaben in den Quellen wird angestrebt (z.B. Adressenangaben in Zeitschriften). Es muß hier jedoch einschränkend vermerkt werden, daß dieser Aufwand als Reaktion auf die bibliometrische Nutzung zwar eine willkommene graduelle aber keine grundsätzliche Verbesserung bringt: im Prinzip werden die technischen Toleranzwerte für Einzeldokumente für die Gesamtdatenbank kumuliert. In der Folge führt dies einerseits zu systematischen Fehlern, die nur schwer zu kontrollieren sind. Andererseits können diese Fehler aber auch zu einem gewissen Grad vernachlässigt werden durch eine Aggregation der Daten.

Diese Entwicklung hat schließlich zur Generierung von eigentlichen "Sekundärdatenbanken" geführt, in welchen nicht mehr alle Einzelelemente eines Dokuments vollständig abgebildet werden, sondern nur noch eine auf den spezifischen Verwendungszweck hin bereinigte und reduzierte Information (*data reduction*), z.B. in Form von *cluster keys* (Schubert et al., 1989) oder geeigneter Zeiger (z.B. *location number*). Solche ausschließlich für die bibliometrische Nutzung konzipierten Datenbanken sind dann selbstverständlich für das bibliographische Retrieval nicht mehr oder nur beschränkt geeignet.

## 3 Retrieval und Datenanalyse im Überblick

Der augenscheinlichste Differenz zwischen herkömmlichem und bibliometrischem Retrieval liegt in der unterschiedlichen Zielsetzung. Während bibliographisches Retrieval darauf abzielt, Fachliteratur dem wissenschaftlichen Kommunikationsprozeß zugänglich zu machen, dient das bibliometrische Retrieval der Analyse dieser Kommunikationsprozesse. Obwohl die Aufgaben der Bibliometrie mittlerweile sehr vielfältig geworden sind, haben alle die gleiche Konsequenz für die Datenbehandlung, egal, ob es sich um recht einfache Input-Output-Analysen für das Science Management oder um hoch theoretisierte Strukturanalysen der Wissenschaft unter Einbeziehung neuester mathematischer Methoden handelt: Die wissenschaftliche Publikation wird als kleinstes meßbares Element dokumentierter Wissenschaftskommunikation, sozusagen als bibliometrische Maßeinheit betrachtet. Die verlorene Individualität des Dokuments wird durch Standardattribute wie Fachgebiets- oder geopolitische Zuordnung ersetzt. Genau hierin liegt aber auch ein wesentlicher Aspekt möglichen Mehrwerts bibliographischer Datenbanken mittels bibliometrischer Nutzung. Im Detail lassen sich auf der Seite des Retrievals folgende Verschiebungen feststellen: Bei der klassischen Suche (sowohl in gedruckten Bibliographien als auch in Datenbanken) mit der überwiegenden Orientierung auf einzelne Dokumente kann der Benutzer jeweils individuell über die Relevanz entscheiden (Interaktivität). Bibliometrisches Retrieval kann mithin als Fortsetzung des herkömmlichen Retrievals auf einer höheren

Abstraktionsebene angesehen werden. An der Stelle des Einzeldokuments steht hier die Datenmenge einer beliebigen, aber im allgemeinen vorgegebenen Aggregationsebene, in der das Einzeldokument im Bedarfsfall noch identifiziert werden kann. Eine individuelle inhaltliche, formale etc. Bewertung der Dokumente bzw. eine Korrekturmöglichkeit der Suchstrategie auf der sog. Mikroebene ist allerdings nicht mehr gegeben. Die Relevanz verlagert sich von den Einzelfällen in der gefundene Datenmenge auf die Menge als Ganzes. Relevant ist nurmehr die gesamte gefundene Menge, und diese wird weiter auf ihre Strukturmerkmale untersucht.

Dieses Vorgehen bedeutet ein qualitativ und quantitativ neues Niveau der Datenanalyse, die durch gleichzeitigen Zugriff auf bis vor wenigen Jahren unvorstellbar große Datenmengen gegeben ist. Erst dadurch ist die Möglichkeit von statistisch zuverlässigen Analysen eröffnet worden.

Das Vorgehen birgt aber auch gewisse Gefahren. Zwar erfolgt maschinelles Retrieval noch erfahrungsgestützt, es ist aber nicht mehr interaktiv rückgekoppelt an die Relevanz einzelner Dokumente. Daraus können sich Unsicherheiten in der Bewertung der Resultate ergeben, sodaß letztlich Artefaktbildung nicht immer ausgeschlossen werden kann. Unter Anwendung einer ausgereiften Methodik kann jedoch in Kenntnis der inhaltlichen und technischen Schranken der bibliographischen Datenbanken und bei informiertem Umgang mit dem Datenmaterial ein optimales Verhältnis von Zuverlässigkeit und Aussage erreicht werden.

Im folgenden sollen die Charakteristika des bibliographischen und bibliometrischen Retrievals tabellarisch zusammengestellt werden:

<i>Bibliographisches Retrieval</i>	<i>Bibliometrisches Retrieval</i>
Lokalisieren von Einzeldokumenten	Lokalisieren von Dokumentmengen
Verbindung von Einzeldokumenten	Verbindung zwischen Dokumentmengen (Topics, Institutionen, Ländern etc.)
Erweiterung durch paralleles Abfragen verschiedener Datenbanken (Redundanzproblem)	Erweiterung durch Zusammenführung verschiedener Datenquellen

Tab. 1: Charakteristika des bibliographischen und bibliometrischen Retrievals

#### 4 Mehrwert bibliographischer Datenbanken durch bibliometrische Nutzung

Im folgenden sollen die neuen Möglichkeiten der Datenbanknutzung anhand einiger wichtiger Retrievalfelder untersucht und dem bibliographischen Retrieval gegenübergestellt werden. Generell werden beim traditionellen Retrieval die einzelnen Felder einander ergänzend genutzt. Ziel ist dabei die Treffsicherheit der Suche zu erhöhen, oder fachsprachlich ausgedrückt, die Parameter der Recall-Precision-Kurve zu optimieren. Das heißt, durch optimale Auswahl der geeigneten Suchfelder und logisch gut strukturierten Abfrageformulierungen soll der Anteil der Treffer (gefundene und relevante Literatur) möglichst hoch, die Zahl der gefundenen aber nicht relevanten Dokumente möglichst niedrig, der Zeitaufwand und die Kosten für das Retrieval möglichst gering gehalten werden. Auch wenn das Retrieval auf der logischen Verknüpfung von Suchfeldern und -optionen basiert, so sind bei dieser Form alle Abfragen nach Feldern separierbar und auch komponentenweise sinnvoll, nur eben nicht unbedingt optimal. Zur Verdeutlichung soll hier folgendes Beispiel dienen: Es werden Dokumente zu einem vorgegebenen Thema gesucht. Dazu werden die Felder AUTHOR/ADDRESS, SCOURCE, PUBLICATION YEAR sowie eine Kombination von Schlüsselwörtern benutzt. Es ist offensichtlich, daß die gleiche Abfrage jedoch feldweise formuliert und ohne Verknüpfung mit Ausnahme des Feldes PUBLICATION YEAR ebenfalls brauchbare (aber nicht optimale) Ergebnisse liefern würde. Im Gegensatz hierzu kann man sich die bibliometrische Nutzung als komplexe Bearbeitung ausgewählter Felder vorstellen. Auf die Abfrage einzelner Felder reduziertes Retrieval spielt nur in Form einiger statistischer Basisinformationen eine Rolle (z.B. die Zahl der Publikationen ausgewählter Forschungsgruppen). Der

eigentliche Mehrwert ergibt sich aus der Verknüpfung der in einer Vielfalt ausgewählter Felder enthaltenen Information, sodaß diese dann nach Feldern nicht mehr separierbar und - wie bereits erwähnt - in Hinsicht auf Dokumente oder Dokumentenmengen nicht mehr disaggregierbar ist. Als typisches Beispiel hierfür kann der relative Anteil der Publikationen eines einzelnen Landes in einen Fachgebiet im Vergleich etwa zum Durchschnitt aller Länder angeführt werden. Obwohl neben dem Feld PUBLICATION YEAR nur zwei weitere Felder (ADDRESS und ein zusätzliches zur Klassifizierung der Dokumente nach Fachgebiet) genutzt werden, ist jeder Rückschluß auf Ergebnisse, die aus dem getrennten Retrieval der einzelnen Elemente resultieren würden, ausgeschlossen. Aus dem beschriebenen Verfahren leiten sich eine Vielzahl der in der Bibliometrie üblichen Indikatoren ab, die zur Beschreibung der unterschiedlichen, auseinander direkt nicht erschließbaren Aggregationsebenen erforderlich sind.

Es soll nun kurz umrissen werden, welcher informationelle Mehrwert sich aus der genannten Verknüpfung von Datenfeldern und aus dem Verlagern des Retrievals auf eine sog. höhere Abstraktionsebene gewinnen läßt. Die folgenden Beispiele mögen das verdeutlichen: Für eine Analyse der Forschungsdynamik in ausgewählten geopolitische Regionen wurden in einem festgesetzten Zeitraum von fünf Jahren lediglich die Felder AUTHOR und ADDRESS genutzt. Abbildung 1 zeigt den Plot des Indikators *Renewal* gegen *Transience* (s. Glänzel, 1992). Als Transienten werden Autoren bezeichnet, die nach Veröffentlichung einer Arbeit (meist ihrer Dissertation) ihre Publikationstätigkeit aufgeben. Für den Anteil solcher Autoren an der gesamten *Scientific community* wird der Begriff *Transience Index* verwendet. Im Gegensatz dazu widerspiegelt der *Renewal Ratio* das Verhältnis von sogenannten *Newcomers* und *Terminators*. In Abbildung 1 wird die Dynamik der wissenschaftlichen Gemeinschaften in den ausgewählten Länder veranschaulicht.

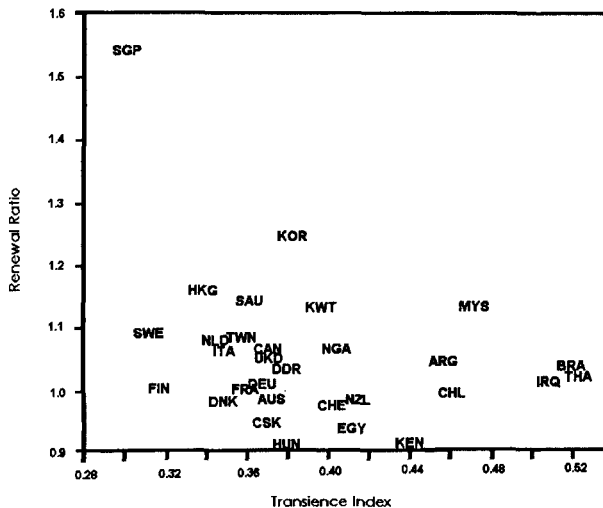


Abb. 1: Graphische Darstellung des Indikatorpaars *Renewal-Transience* für 30 ausgewählte Länder aus 10 geopolitischen Regionen (nach Glänzel, 1992)

Das zweite Beispiel illustriert die bibliometrische Verknüpfung zweier bibliographischer Datenbanken, in diesem Falle den Datenbanken INSPEC und SCI (*Science Citation Index*). In einer Studie von Tijssen et al. (1988) wurden physikalische Fachzeitschriften mittels multivariater statistischer Methoden hinsichtlich ihrer Fachgebietsstruktur und ihrer Rolle im Rahmen der durch Zitierungen dokumentierten wissenschaftlichen Kommunikation untersucht. Die Zeitschriftentitel wurden jeweils dem Feld SOURCE der beiden Datenbanken, der CLASSIFICATION CODE (PACS-code) wurde der Datenbank INSPEC, die

Zitierungen dem SCI entnommen. Abbildung 2 zeigt je ein CA-Diagramm für die Fachgebiets- und Zitierungsklassifikation der untersuchten Physikzeitschriften (PACS-code sind als zweistellige Zahlen, die Abkürzungen der zitierten Zeitschriften sind in Normalschrift, die der zitierenden kursiv dargestellt).

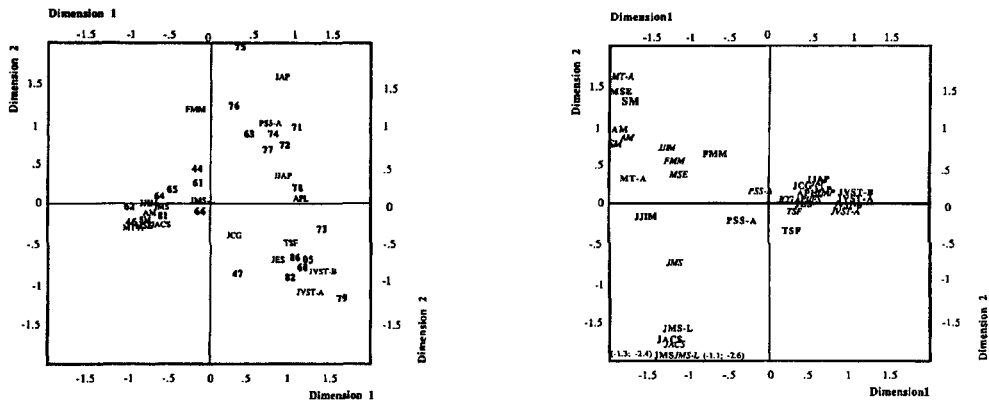


Abb. 2: Fachgebietsklassifikation (links) und Zeitschrift-zu-Zeitschrift-Zitierungen (rechts) in der Physik 1985 mittels Korrespondenzanalyse (nach Tijssen et al., 1988)

Die beiden angeführten Beispiele liegen deutlich außerhalb der Ziele und Möglichkeiten konventioneller Nutzung bibliographischer Datenbanken. Mit einem letzten und zugleich aktuellsten Beispiel soll demonstriert werden, daß trotz der Verarbeitung hochaggregierter Daten in der bibliometrischen Forschung auch Anwendungen auf das Retrieval von Einzeldokumenten nicht ausgeschlossen sind. Die bibliographische Verknüpfung (*bibliographic coupling*) gilt traditionell neben der *co-word* und der *co-citation analysis* zu den bevorzugten Methoden der "Kartographie der Wissenschaft". Mit Hilfe multivariater statistischer Analysen werden bei den genannten Verfahren das gemeinsame Vorkommen von Schlüsselwörtern bzw. Referenzen oder Zitierungen in den entsprechenden Feldern ausgewertet. Bibliographische Verknüpfung kann aber darüber hinaus auch zur Identifizierung von Forschungsfront-Themen und sogenannten Kerndokumenten (Glänzel & Czerwon, 0000) dienen. Einmal identifiziert, können diese Kerndokumente ihrerseits zum Auffinden weiterer mit ihnen verknüpfter Dokumente genutzt werden, um den Retrievalprozeß für diese Themen zu vereinfachen und zu optimieren.

## 5 Anforderungen an Struktur und Nutzung von Datenbanken

Die im vorangegangenen Abschnitt angeführten Beispiele zeigen deutlich, welche hohen Qualitätsanforderungen bibliometrisches Retrieval an die bibliographischen Datenbanken stellt. Diese betreffen sowohl die Seite der strukturellen Aufbereitung der Datenbasen selbst als auch die Seite des eigentlichen Retrieval und der Weiterverarbeitung bzw. Interpretation der Ergebnisse.

Zunächst ist natürlich die Vollständigkeit der Daten zu nennen. Die Felder AUTHOR und ADRESS sollten sämtliche Autorennamen und alle Adressen enthalten. Einige Datenbanken beschränken sich immer noch auf die Angabe der Adresse des ersten Autors bzw. der Korrespondenzadresse. Die lapidare Angabe "et al" im Autorenfeld ist für viele bibliometrische Untersuchungen unzulänglich. Die zweite Forderung betrifft die Genauigkeit der Angaben, vor allem in Hinsicht auf Orthographie und Transliteration. Hierher gehört aber auch die einheitliche Handhabung der Umlaute in deutschen Namen und die Reihenfolge bei z.B. flämischen (*van*) oder japanischen Namen. Wichtige Elemente der Felder sollten wenigstens innerhalb einer Datenbank vereinheitlicht sein (z.B. Ländernamen, Postleitzahlen, Zeitschriftentitel usw.). Neben der Vereinheitlichung der Datensätze kommt

einwichtiges weiteres Desiderat hinzu: die Erfahrung mit den in der Bibliometrie am häufigsten benutzten Datenbanken des ISI zeigen, daß an einer Verbesserung der Struktur der Referenzfelder dringend gearbeitet werden muß. Dies betrifft sowohl die Erweiterung der Angaben um Ko-Autoren als auch die Präzisierung der Quellenangaben. Schließlich und letztlich wäre auch eine gewisse Kompatibilität der größeren Datenbanken untereinander erwünscht. Die konsequente Umsetzung dieses Konzeptes wäre dann neben der größeren Zuverlässigkeit der bibliometrischen Statistiken auch durch eine gewisse Rückkopplung zur konventionellen Nutzung (s. drittes Beispiel des letzten Abschnitts) dem bibliographischen Retrieval förderlich.

Selbstverständlich stellt der Anspruch auf Wissenschaftlichkeit auch an die bibliometrischen Nutzer der Datenbanken einige Forderungen. Dem uninformierten Gebrauch ist mit verstärkter Aufklärung über Retrievalmöglichkeiten und -Grenzen sowie mit Hilfen zur Ergebnisinterpretation zu begegnen (Schoepflin, 1994). Eine größere Transparenz und die genaue Dokumentation der angewendeten Methoden würden die Reproduzierbarkeit der bibliometrischen Ergebnisse verbessern und potentiellen Anwendern im konventionellen Bereich des Retrieval deren Anwendung erleichtern.

## 6 Schlußfolgerungen

Die Eingangs als "zweckentfremdet" bezeichnete Nutzung bibliographischer Datenbanken eröffnet neue Dimensionen. Einerseits wird eine Darstellung von Strukturen und Trends ermöglicht, wie sie mit den einfachen Retrievalverfahren, aber auch mit anderen traditionellen Methoden (z.B. Gutachten), nicht sichtbar wären. So läßt sich das Entstehen, das Auseinanderdriften oder das Zusammenwachsen von Fachdisziplinen beobachten, teilweise noch bevor diese Tendenzen von Fachwissenschaftlern wahrgenommen werden.

Das Lokalisieren von sogenannten *invisible colleges* gehört ebenso zu den Möglichkeiten bibliometrischen Retrievals wie die Auswertung der Auswirkungen politischer Veränderungen in einzelnen Ländern und Regionen (z.B. Osteuropa ab 1990) auf Forschungspotential, -profil und -leistung der Wissenschaft in diesen Ländern. Die Zusammenführung verschiedener Datenbanken (z.B. von Fachdatenbanken und multi-disziplinären Datenbanken) kann dabei die Zuverlässigkeit der getroffenen Aussagen wesentlich verbessern (Glänzel & al., 0000).

Auch das Erfassen und die Modellierung wissenschaftlicher Informationsprozesse gehört zu den Einsatzgebieten bibliometrischen Retrievals. Die Analyse von Alterungsprozessen wissenschaftlicher Literatur (Glänzel & Schoepflin, 1994a) ermöglicht die detaillierte Charakterisierung der dokumentierten wissenschaftlichen Kommunikation und sogar mittelfristige Vorhersagen bei der Rezeption wissenschaftlicher Literatur (Glänzel & Schubert, 1994).

Es muß zum Schluß jedoch nochmals auf die ernstesten Probleme beim Umgang mit hochaggregierten Daten hingewiesen werden. Das sind zum einen unzuverlässige Aussagen und Artefakte durch Nutzung von Daten aus zweiter oder dritter Hand (Glänzel & Schoepflin, 1994b), zum anderen aber die Gefahr der Anonymität der zugrundeliegenden Dokumente bei bibliometrischem Retrieval. Diese "Anonymität der Daten" kann durchaus zu einem "sorglosen" Umgang mit bibliographischer Information verführen und damit vor allem in der wissenschaftspolitischen Anwendung erheblichen Schaden verursachen.

Die obigen Beispiele machen trotz allem deutlich, daß die bibliometrische Nutzung bibliographischer Datenbanken zu einer neuen Qualität führen kann, die jenseits aller Möglichkeiten der traditionellen Nutzung bibliographischer Information liegt.

## Literatur

- P. Jacsó**, Searching for the Skeletons in the Database Cupboard Part I: Errors of Omission. *Database*, 16 (1), 1993a, 38-49.
- P. Jacsó**, Searching for the Skeletons in the Database Cupboard Part II: Errors of Commission. *Database*, 16 (2), 1993b, 30-36.
- W. Glänzel**, Publication Dynamics and Citation Impact: A Multi-Dimensional Approach to Scientometrics Research Evaluation. In P. Weingart, R. Sehringer, & M. Winterhager (Eds.), *Representations of Science and Technology. Proceedings of the International Conference on Science and Technology Indicators, Bielefeld, Federal Republic of Germany, 10-12 June 1990* (pp. 177-188). Leiden: DSWO Press, 1992
- W. Glänzel, A. Schubert**, Predictive Aspects of a Stochastic Model for Citation Processes. *Inf. Proc. & Management*, 1994 (in press)
- W. Glänzel, U. Schoepflin**, A Stochastic Model for the Ageing of Scientific Literature. *Scientometrics* 30 (1), 1994, 49-64
- W. Glänzel, U. Schoepflin**, Little Scientometrics, Big Scientometrics - and Beyond. *Scientometrics* 30 (2), 1994, 375-384
- W. Glänzel, A. Schubert, T. Braun, E. Rinia, M. Brocken**, Physics in the European Union in the 1980-s. (unpublished manuscript)
- U. Schoepflin**, Problems of Representativity in the Social Sciences Citation Index. In P. Weingart, R. Sehringer, & M. Winterhager (Eds.), *Representations of Science and Technology. Proceedings of the International Conference on Science and Technology Indicators, Bielefeld, Federal Republic of Germany, 10-12 June 1990* (pp. 177-188). Leiden: DSWO Press, 1992
- U. Schoepflin**, Scientometrie im Schnittpunkt zwischen Fachinformation und Wissenschaftspolitik. In W. Neubauer (Ed.), *Qualität und Information*. Deutsche Gesellschaft für Dokumentation, Frankfurt a.M., 1994, p. 69-81.
- A. Schubert, W. Glänzel, T. Braun**, Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981-1985. *Scientometrics*, 16(1-6), 1989, 3-478.
- R.J.W. Tijssen, J. de Leeuw, A.F.J. van Raan**, A Method for Mapping Bibliometric Relations Based on Field-Classifications and Citations of Articles. In L. Egghe & R. Rousseau (Eds.), *Informetrics 87/88*, Elsevier Science Publishers B.V., Amsterdam, 1988, p. 279-292
- M. Winterhager**, Möglichkeiten und Grenzen der Anwendung bibliometrischer Methoden in den Sozialwissenschaften. In W. Neubauer & K.-H. Meier (Eds.), *Technik und Information: Markt, Medien und Methoden* (pp. 571-582). Frankfurt a.M.: Deutsche Gesellschaft für Dokumentation, 1993