

Ulrich Herb: Vermessung der Wissenschaft, erschienen in telepolis, 01.11.2008

<http://www.heise.de/tp/r4/artikel/28/28942/1.html>

Vermessung der Wissenschaft

Die Freiheit der Forschung besteht weiter, auch wenn Verwirtschafterlichung und Privatisierung der Hochschulen

[http://www.gew.de/Privatisierungsreport_6_Schoene_neue_Hochschulwelt.html]

voranschreiten. Die voranschreitende Kommerzialisierung lässt sich schon an der Namensgebung erahnen: angefangen beim „Hörsaal Aldi Süd“

[<http://www.sueddeutsche.de/jobkarriere/berufstudium/artikel/775/88687/>] der

Fachhochschule Würzburg-Schweinfurt, über die Jacobs-University Bremen

[<http://www.jacobs-university.de/>] bis zur vom regionalen Tourismusverband gesponserten

Alpen Adria Universität Klagenfurt [<http://www.uni-klu.ac.at/main/index.htm>]. Die

Ökonomisierung beeinflusst aber nicht nur Bezeichnungen, sondern auch die

hochschulinterne Mittelverwaltung, die mehr denn je eine Effizienzprüfung vornehmen will.

Wissenschaftler an Hochschulen, Lehrstühle und Fachbereiche werden auf vielfältige Art

bewertet und die Bewertung hat für sie handfeste Folgen: Sie entscheidet über zukünftige

materielle und personelle Ausstattung und durch Entzug oder Erhöhung von Mitteln auch

über die Wahl von Forschungsschwerpunkten. Unreflektierte Interpretation und

leichtgläubiger Umgang mit den ermittelten Werten dürften allerdings schädlich sein: Zu oft

vermitteln sie falsche Konkretheit und vernachlässigen fachspezifische Unterschiede, die zu

ihrer Ermittlung angewandten Methoden ignorieren weitgehend neuartige Publikations- und

Forschungsmodelle – was durch die ausbleibende Belohnung Innovationen behindert.

Ein Workshop der „Information und Kommunikation IuK-Initiative Wissenschaft e.V.“

[<http://www.iuk-initiative.org/>] bot kürzlich einen Einblick in die Praxis der

Leistungsbewertung in den Wissenschaften. Vor Vertretern von Fachgesellschaften,

Forschungseinrichtungen, Hochschulverwaltungen und Bibliotheken stellte Prof. Stefan

Hornbostel, Leiter des Institut für Forschungsinformation und Qualitätssicherung

IFQ [<http://www.forschungsinform.de/>], das die Deutsche Forschungsgemeinschaft DFG

[<http://www.dfg.de>] in Fragen der Qualitätssicherung in der DFG-geförderten Forschung

berät, die Entwicklung zum integrierten Qualitätsmanagement fest: In Zeiten leerer Kassen

sollen möglichst viele Faktoren zur Qualitätsbestimmung herangezogen werden. Für

Hochschulverwaltungen und Geldgeber ist die *leistungsorientierte Mittelvergabe (LOM)* Instrument der Wahl, für Wissenschaftler ist sie eher Folterwerkzeug der Controller und Technokraten.

Prinzipiell ist LOM nicht gleich LOM: Die Mittelverteiler haben durch die Auswahl der zu bewertenden Faktoren großen Einfluss darauf, welches Engagement belohnt und gefördert werden sollen und welches nicht. Für die LOM im wissenschaftlichen Metier sind aber in erster Linie Promotionen, Drittmittelakquise und Publikationsverhalten relevante Kennziffern. Die Anzahl der Promotionen soll Aufschluss über die Nachwuchsförderung geben, aber leider spielt die Qualität der Arbeiten dabei keine Rolle. Das Volumen der Drittmittelinwerbung bezeichnet das Ausmaß, zu dem Forschungsgelder aus hochschulexternen Quellen akquiriert werden – üblicherweise auf den Töpfen der öffentlichen Wissenschaftsförderer wie DFG oder des Bundesministeriums für Bildung und Forschung BMBF [<http://www.bmbf.de/>], zunehmend aber auch von Wirtschaftsunternehmen wie der Volkswagenstiftung [<http://www.volkswagenstiftung.de/foerderung.html>]. Nun haben die Wirtschaftsunternehmen ihrer Natur nach meist eher Interesse an wirtschaftlich verwertbarer Forschung. Berücksichtigt die LOM zunehmend unternehmensgesponserte Drittmittelforschung werden Fächer und Forschung von rein gesellschaftlicher Relevanz oder mit geringer Anwendungsorientierung, die sich schwierig kommerzialisieren lassen, rasch benachteiligt. Man denke etwa an Migrationsforschung: gelungene Integration von Migranten verringert vielleicht soziale Spannungen, Konflikte und Kriminalität, lässt sich aber nicht vermarkten. Von zentraler Bedeutung für die LOM-Verfahren ist aber das Publikationsverhalten: Anders als bei den Promotionen soll dabei nicht nur die Menge der veröffentlichenden Artikel in wissenschaftlichen Journals oder in Sammelbänden, der Monographien und anderer Werke berücksichtigt werden, sondern auch deren Qualität. Publikationsdaten sind auch deswegen von großem Interesse, weil man sich von ihnen internationale Vergleichbarkeit verspricht.

Gängigerweise wird der Journal Impact Factor (JIF) zur Bestimmung der Qualität einer Publikation herangezogen – was, euphemistisch formuliert, einer methodologischen Unschärfe gleichkommt, denn der JIF misst allenfalls Popularität und auch das mehr schlecht als recht [<http://www.bio-diglib.com/content/2/1/7>]. Der JIF wird aus dem Quotient der *Zahl der Zitate im Bezugsjahr auf die Artikel eines Journals der vergangenen zwei Jahre* und der *Zahl der Artikel eines Journals in den vergangenen zwei Jahren* berechnet Die Auswahl der Journals obliegt einzig Thomson Scientific [<http://scientific.thomson.com/isi/>], im Sample

dürften kaum *alle* relevanten Journals vertreten sein – de facto sind nichtenglischsprachige Journals unterrepräsentiert. Zudem ignoriert der JIF komplette Dokumentgattungen: z.B. Bücher, Buchbeiträge, graue Literatur. Schließlich bevorteilt er Zeitschriften, die zahlreiche Reviews beinhalten, denn diese werden im Zähler, aber nicht im Nenner berücksichtigt. Der JIF sagt auch nichts über die Zitationshäufigkeit eines *Artikels* aus, denn er wird auf Journalebene berechnet. Selbst wenn er sich auf Artikel bezöge wäre der ermittelte Wert schwerlich auf die Leistung einzelner Personen herunterzubrechen, denn gerade in den Naturwissenschaften finden sich teils Artikel mit Autorenzahlen im zweistelligen Bereich – wobei der Beitrag einzelner Autoren oft eher symbolischer Art ist: Man könnte fast einem Tauschhandel sprechen [<http://www.heise.de/tp/r4/artikel/22/22197/1.html>]. Ungeachtet der beachtlichen Mängelliste glaubt man mancherorts, aus dem Durchschnitt der JIF-Werte der Journale, in denen publiziert wurde, auf die Qualität der Forschung oder gar der Forscher schließen zu können.

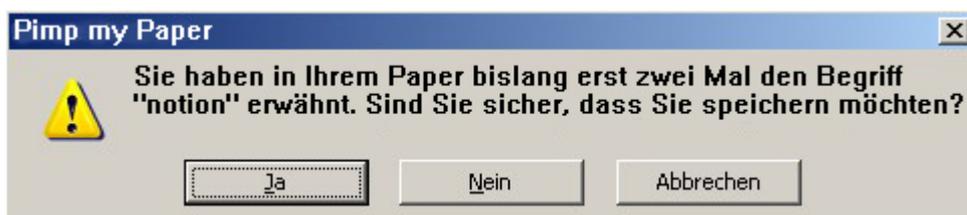
Ein anderes auf Zitationsdaten basierendes Verfahren, das in einigen Einrichtungen als Evaluationsindikator benutzt wird und womöglich dem JIF sogar den Rang ablufen könnte, ist der sogenannte Hirsch-Index oder h-Index [<http://de.wikipedia.org/wiki/H-Index>]. Der h-Index bezieht sich auf Autoren und nicht auf Journals, er wird folgendermaßen berechnet: Ein Autor hat einen Index h , wenn h von seinen insgesamt N Veröffentlichungen mindestens jeweils h Zitierungen haben und die anderen $(N-h)$ Publikationen weniger als h Zitierungen. Hat ein Forscher also 12 Artikel publiziert, die 12 mal zitiert wurden, hat er einen h-Index von 12; wurde der am wenigsten zitierte Artikel nur neun mal zitiert und die restlichen elf 12 mal, hat er einen h-Index von neun. Die herangezogenen Zitationsdaten werden aus der Datenbank *Web of Science* [http://de.wikipedia.org/wiki/Web_of_Science], ein Produkt des bereits erwähnten Konzerns Thomson Scientific bezogen. Der h-Index teilt einige Mängel mit dem JIF: Vernachlässigung ganzer Dokumentgattungen und nichtenglischer Publikationen, Ignoranz des Mehrautorenproblems. Außerdem ist die trennscharfe Autorenidentifikation im *Web of Science* nicht sichergestellt und auch beim h-Index stellt sich die Frage, ob er nicht eher Popularität als Qualität misst.

Die Verwendung der aus dem JIF destillierten Kennziffern oder des h-Index birgt auch die große Gefahr, fachliche Normierungen zu übersehen: Techniker und Ingenieure publizieren kaum in internationalen Journals, sondern in deutschsprachigen, ihre Publikationen sind folglich im *Web of Science* unterrepräsentiert. Wissenschaftler anderer Disziplinen trifft das

gleiche Schicksal, denn sie publizieren eher Bücher (etwa Geisteswissenschaftler) oder Konferenzbeiträge (etwa Informatiker) als Journalartikel. Die Verwendung von h-Index und JIF leidet auch unter einer starken Asynchronität, schließlich werden Forschung und Publikationen vergangener Jahren belohnt.

Auch Techniken zur automatisierten Ermittlung evaluationsrelevanter Werte gewinnen an Bedeutung: Das EU-geförderte Projekt EERQUI will computerlinguistische Verfahren der Qualitätsbestimmung elektronischer Dokumente erproben und versucht, die Qualität von Dokumenten über statistische Analysen der Semantik und Syntax zu bestimmen. Dazu soll unter anderem das Vorkommen als relevant eingeschätzter, abstrakter Keywords wie etwa *central concept*, *argument*, *research question*, *principle*, *notion* überprüft werden. Das Verfahren soll auch eine Konsistenzprüfung leisten und feststellen, ob die im Abstract eines Dokuments gemachten Versprechungen hinsichtlich der zu beantwortenden Forschungsfragen eingehalten werden. Perspektivisch soll die Qualitätsbeurteilung Faktoren wie Nutzungs- und Zitationsdaten der Dokumente berücksichtigen. Dabei will man auch die Valenz eines Zitats berücksichtigen: Dokumente, die im zitierenden Dokument negativ kommentiert und bewertet werden, erfahren eine Abwertung, andere eine Aufwertung - ein Feature, das man bei klassischen Datenbanken wie Web of Science oder Scopus [<http://www.scopus.com>] bislang vergeblich sucht. In der Projektphase beschränkt man sich auf Inhalte aus der Erziehungswissenschaft.

Ob die Technik resistent gegen Keyword-Spamming ist, wird sich zeigen: Die Verwendung gewisser Begriffe, das Befolgen eines bestimmten Aufbaus und Nutzung einer vorteilhaften Struktur dürften das Ranking positiv beeinflussen. Anbieter von Software zur Literaturverwaltung etwa könnten auf die Idee kommen neue Funktionalitäten anzubieten, die nicht nur das Einhalten eines fachspezifischen Zitations- und Bibliographierstils erleichtern, sondern die Dokumente gleichzeitig für eine positive Bewertung aufzubereiten helfen.



Dialog-Fenster der noch fiktiven Software „Tune my paper“

Defätisten könnten weitere nachteilige Effekte für das wissenschaftliche Arbeiten wittern: Einförmigkeit der Wissenschaftssprache, vollkommen vereinheitlichter Aufbau von Artikeln, sprachliche und inhaltliche Erwünschtheitseffekte. Die Technik ist außerdem nur für traditionelle Publikationsformen, in erster Linie Artikel, geeignet - Blogs und Diskussionslisten bleiben außen vor, innovative Konzepte wie living documents [http://en.wikipedia.org/wiki/Living_document] oder Publizieren im Wiki-Stil sind aufgrund der ihrer Kollaborität geschuldeten Uneinheitlichkeit und ihrer Diskursivität wohl schwer integrierbar. Allerdings soll die Technik des Pilotprojekts den menschlichen Gutachter nicht ersetzen, sondern unterstützen: Das letzte Wort hat der Reviewer - falls er sich von der Auswertung nicht steuern lässt.

Seit langen Jahren in der Bewertung von Hochschulen aktiv ist das 1994 von der Bertelmannsstiftung und der Hochschulrektorenkonferenz gegründete *Centrum für Hochschulentwicklung CHE* [<http://www.che.de/>]. Vorrangig berät das CHE bei der Modernisierung und Liberalisierung des Hochschulwesens und prägt sehr aktiv den Bildungsdiskurs. Dem CHE wird, auch wegen der Nähe zur Bertelmannstiftung, teils mangelnde Objektivität und eine starke Präferenz für die Verwirtschaftlichung der Hochschullandschaft unterstellt [<http://www.studis-online.de/HoPo/art-689-che-politik.php>]. Wegen Zweifeln an der Methodik des CHE-Rankings verzichtet seit 2007 die Schweizer Rektorenkonferenz auf die Teilnahme, 2008 schloss sich die österreichische Vertretung der Hochschulen an [<http://www.heise.de/tp/r4/artikel/26/26469/1.html>].

Zu den prominentesten Produkten des Hauses gehören das CHE-Hochschulranking und das CHE-Forschungsranking. Während das Hochschulranking Studierenden oder Studienanfängern bei der Wahl der Universität Orientierung geben soll, versteht sich das Forschungsranking als Benchmarkinginstrument für Hochschulleitungen und Ministerien. Unter den relevanten Forschungsindikatoren finden sich teils klassische Kriterien, die auch bei den bereits erwähnten LOM-Verfahren herangezogen werden: Über Fachbereichsbefragungen wird neben dem Aufkommen an Drittmitteln auch die Anzahl der Promotionen erfasst, zur Relativierung dieser Werte wird die Anzahl der Personalstellen abgefragt. Teils versucht man Fachspezifika über die Erfindungsmeldungen an der lokalen Hochschule zu berücksichtigen; dieser Indikator ist in erster Linie für die Ingenieurwissenschaften und Medizin relevant.

Das Renommee der Fachbereiche wird über eine Hochschullehrerbefragung ermittelt. Welche Fachbereiche von den Befragten für besonders wichtig oder exzellent gehalten werden, kann aber von verzerrenden Faktoren, zum Beispiel der Berichterstattung über Exzellenzinitiativen, beeinflusst werden. Fraglich auch, was die Antworten eigentlich wiedergeben:

Wahrgenommene Qualität, Prominenz oder Bekanntheit. Nicht zuletzt könnte auch eine Kovariation mit dem Publikationsaufkommen der Fachbereiche unterstellt werden.

Auch das CHE verzichtet bei der Erstellung des Forschungsrankings nicht auf bibliometrische Analysen und nutzt dazu Informationen zum Publikationsaufkommen und Zitationsindizes. Zur Ermittlung der Publikationszahlen zieht man Datenbanken heran – allen voran das bereits erwähnte multidisziplinäre *Web of Science*, seltener Datenbanken, die schwerpunktmäßig Publikationen eines Faches nachweisen. Allerdings weisen diese Datenbanken auch nicht alle Publikationen einer Einrichtung nach. Die bibliometrischen Analysen berücksichtigen innerhalb eines Zeitfensters von drei Jahren unter anderem die Publikationen pro Kopf, Zitationen je Papier und die Publikationen pro Jahr. Zitationen können aber nicht in jedem Fach berücksichtigt werden: Teils weisen die Fachdatenbanken keine Zitationen nach, teils weisen die Datenbanken zwar Zitationen nach, erfassen aber nicht die innerhalb des Fachs üblichen Dokumenttypen. Für die Analysen in der Soziologie wird z.B. das Web of Science herangezogen, was eine deutliche Verzerrung wegen dessen Bevorzugung von Journal-Artikeln und der Nichtberücksichtigung der präferierten Soziologensprache zur Folge hat. Soziologen forschen meist regional und publizieren in deutsch, Auswertungen auf Basis des Web of Science spiegeln nicht die Forschung wider. Fast schon konsequenterweise ignoriert auch das CHE innovative Publikationsmodelle und -praxen und beschränkt sich auf Papieranalogue wie Bücher, Sammelbandsbeiträge, Zeitschriftenbeiträge.

Anscheinend genügte das CHE-Modell auch nicht den Ansprüchen des Wissenschaftsrates [<http://www.wissenschaftsrat.de/>], der Bund und Länder in Fragen der Weiterentwicklung des Hochschulsystems berät. Daher modellierte er ein eigenes Bewertungsinstrument und erstellte in einer Pilotstudie Forschungsratings - wohlgerne keine Rankings - [http://www.wissenschaftsrat.de/pilot_start.htm] für die Fächer Chemie (publiziert 2007) und Soziologie (publiziert 2008).

Herangezogene Kriterien waren vor allem Qualität und Quantität von Publikationen und der Drittmittelprojekte. Publikationsbezogene Kriterien, die erhoben wurden, waren

Publikationstyp, Sprache und Herkunftsland sowie die stattgefundenen oder nicht stattgefundenen Begutachtung der Dokumente durch eine Peer Review. Die Qualität der Publikationen versuchte man wie üblich vorrangig über Zitationswerte zu bestimmen, musste allerdings für die Soziologie aus Mangel an verlässlichen Zitationsdaten auf ein stichprobenartiges Prüfverfahren ausweichen: Publikationen wurden von Experten hinsichtlich der Qualität bewertet, gesichtet wurden vor allem Journalbeiträge und Auszüge aus Monographien. Es wurde gar überlegt, den Verlag, bei dem ein Werk erschien, als Qualitätskriterium heranzuziehen. Die Qualität der Drittmittelprojekte versuchte man im Fall der Soziologie ebenfalls durch Expertenurteil nach Sichtung der Projektbeschreibungen zu bestimmen. Ratings für andere Fächer aus den Bereichen Technik und Geisteswissenschaften sollen folgen.

Überraschenderweise widmen die geschilderten Verfahren durch die Bank Open-Access-Publikationen keine besondere Aufmerksamkeit. Der Wissenschaftsrat etwa hat die Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen [http://de.wikipedia.org/wiki/Berliner_Erkl%C3%A4rung_%C3%BCber_offenen_Zugang_zu_wissenschaftlichem_Wissen] unterzeichnet und damit die Bedeutung des entgeltfreien Zugangs zu wissenschaftlicher Information anerkannt – allerdings mag man Open Access anscheinend nicht als evaluierungsrelevant akzeptieren. Fast schon ärgerlich auch, dass beim in allen Verfahren herangezogenen Indikator Publikationsverhalten ausschließlich traditionelle, sehr formalisierte und streng kanonisierte Publikationsformen, vorrangig Journalartikel, berücksichtigt werden. Alternative und innovative Publikationsformen werden nicht erfasst und folglich nicht belohnt. Nicht nur die erwähnten Blogs bleiben außen vor, auch die Notebookscience wird ignoriert: Wer am Ende des Arbeitstages seine Forschungsergebnisse live von seinem Arbeitsplatz ins Netz stellt und mit anderen Forschern online diskutiert [<http://www.nature.com/news/2008/080915/full/455273a.html> , <http://drexel-coas-elearning.blogspot.com/2006/09/open-notebook-science.html>], kommt dem ursprünglichen Zweck des wissenschaftlichen Publizierens nahe. Für eine positive Evaluierung hätte er besser die Strategie des tacit knowledge verfolgt, bis er einen Artikel in einem JIF-starken Journal publiziert – kommunikative Logik ist unter Evaluierungsaspekten zweitrangig.

Man mag daran zweifeln, ob angesichts sich rasch wandelnder Publikationsgepflogenheiten die Fixierung auf Zitationsindizes und deren kommerzieller Anbieter zeitgemäß ist. Längst fordern Wissenschaftler angesichts der intransparenten Auswertungen der

Zitationsdatenbanken „Open Access to Citation Data“

[<http://www.jcb.org/cgi/content/full/179/6/1091>]. Generell stellen sich die gebräuchlichen Evaluierungsverfahren erstaunlich unelektronisch und hausbacken dar, auch das Projekt EERQUI widmet sich ausschließlich den elektrifizierten Pendanten der klassischen Papierpublikationen. Über ausbleibende Berücksichtigung und Belohnung für die Nutzung aktueller und innovativer Publikationspraktiken, wird selbstredend die Nutzung alter Publikationsformen fortgeschrieben. Auch neuartige bibliometrische Daten wie Nutzungsdaten elektronischer Dokumente finden aktuell keine Berücksichtigung. Dabei existieren vielversprechende Ansätze, die nicht nur Nutzungshäufigkeiten, sondern Verwandtschaftsbeziehungen (etwa über häufigen gemeinsamen Download) zwischen Dokumenten abbilden: Die Berücksichtigung der Kombination von Nutzungshäufigkeit und Struktur, bildet die Wertschätzung der Wissenschaft für Publikationsorgane oft besser ab als Zitationsindizes und –werte [<http://library.lanl.gov/cgi-bin/getfile?LA-UR-05-9439.pdf>].