

DOI: <https://doi.org/10.11588/ip.2019.1.49609>**Christian HAUSCHKE**

Problematische Aspekte bibliometrie-basierter Forschungsevaluierung

Zusammenfassung

Bibliometrie dient oftmals als Grundlage für die Beurteilung wissenschaftlicher Leistung. Dazu werden verschiedene Metriken und Indikatoren verwendet, die zur Vermessung von Publikationen oder des Outputs von Forschenden gedacht sind. Die Eignung der verwendeten Kennzahlen ist strittig. Oft ist es klar, dass sie völlig ungeeignet sind. Dennoch dienen sie nach wie vor als Grundlage, um über die Karrieren von Forschenden und somit über ihre individuelle Zukunft, aber auch die Zusammensetzung der Akteure in der Wissenschaftslandschaft zu entscheiden. Dies führt zu verschiedenen ethischen und auch ökonomischen Problemen. Initiativen wie die San Francisco Declaration of Research Assessment (DORA) versuchen, diesen Fehlentwicklungen entgegen zu wirken.

Schlüsselwörter

Bibliometrie, Forschungsevaluierung

Problematic aspects of bibliometric research assessment

Abstract

Bibliometrics often serves as a basis for assessing scientific performance. For this purpose, various metrics and indicators are used to measure publications or the output of researchers. The suitability of the key figures used is controversial; it is often clear that they are completely unsuitable. Nevertheless, they continue to serve as a basis for deciding on the careers of researchers and thus on their individual future, but also on the composition of the actors in the scientific landscape. This leads to a variety of ethical and economic problems. Initiatives such as the San Francisco Declaration of Research Assessment (DORA) try to counteract these undesirable developments.

Keywords

Bibliometrics, research assessment



Inhaltsverzeichnis

1 Was ist Bibliometrie?	2
2 Bibliometrische Methoden – eine Auswahl	3
2.1 Metriken auf Zeitschriftenebene	3
2.2 Metriken auf Artikelebene	3
2.3 Metriken auf Autorenebene	4
2.4 Alternative Metriken	4
3 Ethik und Assessment	4
4 Problematische Aspekte der Bibliometrie	6
5 Fazit	10
Literatur	13
Autor	17

1 Was ist Bibliometrie?

Der Begriff der Bibliometrie steht für eine Unterdisziplin der Szientometrie. Während sich die Szientometrie mit der quantitativen Erfassung der Wissenschaft und ihres Outputs beschäftigt, ist die Bibliometrie der Messung wissenschaftlicher Publikationen verpflichtet.

Gegenstand der Bibliometrie, wie sie sich heute darstellt, ist die Vermessung von Publikationen, ihrer Beziehung zueinander und zu anderen Entitäten wie zum Beispiel Personen, Organisationen, Zeitschriften oder abstrakten Konzepten wie Schlagwörtern und Fachgebieten. Ein wesentliches Motiv ist dabei die Quantifizierung der Reichweite und der Auswirkungen eines Artikels, was man in dem Begriff *impact* – oder auch Resonanz (Ball und Tunger 2005, S. 16) – zusammenfasst.

„Bibliometrische Analysen geben Auskunft über die Wahrnehmung von Veröffentlichungen einer Forschergruppe oder eines Instituts in der Fachöffentlichkeit. Sie geben Hinweise über die Wirkung dieser Veröffentlichungen und der dazu gehörenden wissenschaftlichen Ergebnisse. Sie sind ein Instrument, um Aufschlüsse über die wissenschaftliche Publikationsleistung, die Integration in die Wissenschaftslandschaft und die internationale Sichtbarkeit von Forschungsergebnissen zu erhalten.“ (Ball und Tunger 2005, S. 15)

Havemann (2009) fasst die zur Analyse des Gegenstands Wissenschaftskommunikation zur Verfügung stehenden Instrumentarien unter den Überschriften „Bibliometrische Verteilungen“ (z.B. nach Bradford 1934; Lotka 1926) und „Bibliometrische Netzwerke“ (vorwiegend Zitationsnetzwerke und Analysen von Koautorenschaften) zusammen. Gemessen werden damit – auf Basis verschiedener Modelle – unterschiedliche Indikatoren, die Produktivität (auch Outputanalyse, Ball und Tunger 2005, S. 21), Wirkung (Impact, Resonanz), Kooperation und Zitationsverhalten abbilden (Havemann 2009, S. 105–124).

2 Bibliometrische Methoden – eine Auswahl

Wie oben angedeutet existieren verschiedene Varianten bibliometrischer Analysen, aus denen zwei besonders hervorgehoben werden sollten. Die Outputanalyse fokussiert auf der statistischen Abbildung von Publikationszahlen und -typen.

Die Wahrnehmungsanalyse hingegen versucht, die Resonanz zu messen, die eine Publikation hervorruft. Dies geschieht in der Regel über Erwähnung von Publikationen in anderen Werken (Ball und Tunger 2005, S. 21). Diese Publikationen bilden dabei ein Zitationsnetzwerk, das die Beziehung der Publikationen zueinander visualisiert und analysierbar macht (Havemann 2009, S. 49–54). Diese Untersuchungen finden auf verschiedenen Ebenen statt.

2.1 Metriken auf Zeitschriftenebene

Journal-basierte Metriken beschäftigen sich mit der Zählung von Zitationen auf der Aggregationsebene einer Zeitschrift. Die bekannteste journal-basierte Metrik ist der von Eugene Garfield vorgeschlagene Journal Impact Factor (JIF, Garfield 1955, 2006) der Zitationsmengen auf Zeitschriftenebene messbar machen möchte. Der JIF ist ein kommerzielles Angebot der Firma Clarivate (ehemals Thomson Scientific).

Zur Berechnung des JIF wird gemessen, wie oft die in einer Zeitschrift pro Jahr erschienenen Artikel durchschnittlich zitiert wurden. Dazu wird die Anzahl der Zitate, die Artikel einer Zeitschrift in einem Zweijahreszeitraum erhalten haben, durch die Anzahl der Artikel in eben diesen beiden Jahren geteilt. Dabei werden nicht alle Artikel berechnet, sondern nur die von Clarivate als *citable object* (PloS Medicine 2006) deklarierten Artikel. Was genau *citable* ist, wird zwischen Verlag und Clarivate ausgehandelt. Dies macht den JIF intransparent und nicht nachvollziehbar. Der JIF leitet sich zwar ab aus den Zitationen, die in einer Zeitschrift erschienene Artikel erhalten. Es ist an dieser Stelle jedoch wichtig zu bemerken, dass dieser Indikator eine Aussage über die Zeitschrift in einem bestimmten Zeitraum trifft, nicht aber über die Artikel in der Zeitschrift oder gar deren AutorInnen.

2.2 Metriken auf Artikelebene

Metriken auf Artikelebene (ALM, *article-level metrics*) versuchen, die Resonanz eines Artikels zu messen. Sie zielen somit auf den Impact eines einzelnen Artikels ab, nicht auf eine Zeitschrift insgesamt. Dadurch erhofft man sich, die Signifikanz einzelner Publikationen (und dadurch die Sichtbarkeit ihrer Autoren) gezielter zu erfassen und abbilden zu können. Zu diesem Zweck stehen verschiedene Instrumente zur Verfügung. Einfache, und von Verlagen oft verwendete Metriken sind zum Beispiel Downloadzahlen oder die Häufigkeit von Bookmarks in speziellen wissenschaftlichen Diensten (wie Mendeley) für einen spezifischen Artikel (Lin und Fenner 2013). Aber auch die Menge der Zitationen für einen einzelnen Artikel ist den ALM zuzurechnen.

2.3 Metriken auf Autorenebene

Neben den bisher genannten Metriken gibt es eine weitere Sparte, die sich der Quantifizierung der wissenschaftlichen Impacts einzelner Wissenschaftlerinnen und Wissenschaftler widmet. Der seit einigen Jahren am weitesten verbreitete ist der von Hirsch (2005) vorgeschlagene h-Index.

For the few scientists who earn a Nobel prize, the impact and relevance of their research is unquestionable. Among the rest of us, how does one quantify the cumulative impact and relevance of an individual's scientific research output? In a world of limited resources, such quantification (even if potentially distasteful) is often needed for evaluation and comparison purposes (e.g., for university faculty recruitment and advancement, award of grants, etc.). (Hirsch 2005, S. 16569)

Der von Hirsch entwickelte h-index lässt sich sehr einfach berechnen. Hat eine Person x Publikationen, die jeweils x Zitationen erhalten haben, liegt der h-Index bei x . Person A hat 30 Publikationen, von denen 7 mindestens 7 Mal zitiert wurden. Ungeachtet der Menge der Publikationen liegt der h-Index also bei 7.

Häufig kritisiert wird, dass der h-Index nicht zwischen den verschiedenen Autorenschaften unterscheidet. In vielen Disziplinen wird der erst- und der zuletzt genannten Person besonderer Wert beigemessen. Beim h-Index geht diese Differenzierung verloren. Als Reaktion sind leichte Abwandlungen entwickelt worden, z.B. der Ab-Index (Absolute Index, vgl. Biswal 2013), die solche Faktoren mit einbeziehen.

2.4 Alternative Metriken

Um über den wissenschaftlichen Bereich hinaus gehende Resonanz von Publikationen in der Gesellschaft zu erfassen, wurden verschiedene neue Metriken entwickelt, die man unter dem Sammelbegriff *alternative metrics* (Altmetrics) zusammenfasst. Diese Metriken basieren oft auf der Zählung von Zitationen in explizit nicht-wissenschaftlichem Kontext. Beliebt sind Erwähnungen in sozialen Medien (Facebook, Twitter), in Blogs oder in Wikipedia (Neylon und Wu 2009; Fenner 2013). Die Rezeption eines Artikels in der gesamten Gesellschaft bezeichnet man auch als *societal impact* (Bornmann 2013; Oberst 2017).

3 Ethik und Assessment

Assessment (engl. Bewertung, Beurteilung) bezeichnet in der Betriebswirtschaftslehre die Messung und Zählung von Werten, Vorgängen oder Dingen, die daraus abgeleitete Bewertung von Leistungen ermöglicht. Die daraus entstandenen Erkenntnisse dienen nach den Maßgaben des Controllings als Informationsgrundlage für die Steuerung von Organisationen. Dem Personalcontrolling fällt hierbei eine zunehmend wichtigere Rolle zu. Profitorientierten Unternehmungen wird nahegelegt, den Wert personeller Ressourcen als Teil des immateriellen Betriebsvermögens zu betrachten („Humankapital“, vgl. Müller 2011, S. 8).

Das Personalmanagement von Unternehmen ist den höchsten formalen Zielen dieser Organisationen untergeordnet. In der Wirtschaft ist dies die Erzielung von Gewinn, in Non-Profit-Organisationen kann dies auch ein inhaltlich geprägtes Ziel sein. In jedem Fall haben die Mitarbeiterinnen und Mitarbeiter mit ihrer Leistung dazu beizutragen, die Ziele ihres Arbeitgebers zu erreichen.

*„Seinen spezifischen Beitrag zur Gewinnerwirtschaftung leistet das Personalmanagement durch eine kontinuierliche Verbesserung bzw. Optimierung der Arbeitswirtschaftlichkeit, die sich mathematisch als Quotient aus Leistung und Kosten definiert und handlungspraktisch auf zwei grundlegende Bestrebungen verweist: eine Minimierung der personalen Kosten ('Mitarbeiter als Kostenfaktor') und eine Maximierung der personalen Leistung ('Mitarbeiter als Erfolgsfaktor').“
(Kuhn und Weibler 2012, S. 45)*

Unter Personalassessment versteht man also im betrieblichen Personalmanagement die leistungsbezogene Bewertung von Mitarbeitern, zu einem klar definierten Ziel:

„Menschliche Arbeit dient in den traditionellen Denkkonzepten primär der Erreichung der Unternehmensziele: Effizienz, Kostenwirtschaftlichkeit und Produktivität stehen im Vordergrund.“ (Heinen 1992, S. 74)¹

Dies stellt das Personalmanagement vor große Herausforderungen. Die Erfolgsausrichtung ist zwar essentiell für das Überleben von Unternehmen (auch im Non-Profit-Bereich), gleichzeitig müssen ethische Prinzipien bei der Messung und Bewertung von menschlicher Arbeit eingehalten werden. Bird et al. (2005) haben Empfehlungen zur Umsetzung abgegeben, die hier auszugsweise wiedergegeben werden sollen:

1. Die Art und Weise der Messung soll transparent und nachvollziehbar dokumentiert und offengelegt sein.
2. Die Ziele der Messung müssen klar definiert und mit strenger Methodik evaluiert werden. Die bewerteten Personen sollten Einfluss auf das Verfahren nehmen können.
3. Sollte die Erhebung Unsicherheitsfaktoren aufweisen, müssen diese klar genannt werden.
4. Sowohl die bewerteten Personen als auch die Empfänger des Berichtes über die gemessene Leistung müssen über Möglichkeiten und Grenzen der Leistungsmessung aufgeklärt sein (Bird et al. 2005, S. 2).

Diese Kriterien dienen einerseits dem Schutz des Verfahrens, andererseits dem Schutz der Mitarbeiter und auch der Ziele der Organisation. Ist das Verfahren intransparent oder wird die Durchführung verschleiert, ist nicht mit Akzeptanz der Mitarbeiter zu rechnen. Ein Vertrauensbruch des Arbeitgebers entstände, was zur Ablehnung desselben und in Folge dessen zu verringerter Leistungsbereitschaft führt.

Auch ein methodisch unsauberes Verfahren hat verschiedene negative Auswirkungen. Es ist davon auszugehen, dass keine Akzeptanz durch die Mitarbeiter vorliegen wird. Dazu kommt, dass die erhobenen Daten faktisch nicht zu verwenden sind. Sind sie dennoch Entscheidungsgrundlage, wird auf Basis von Fehlwahrnehmungen entschieden.

„Generally, research on humans needs third-party approval of its ethics and methodology, and is respectful both of confidentiality and the properly informed consent of participants. Exceptionally and for good reason, individual consent may

1 Zitiert nach [Müller 2011, S. 20–21](#)

be relaxed in the public interest, or for specific benefit. But, it is necessary to ensure that a research participant's identity cannot be deduced from published data, or analyses“ (Bird et al. 2005, S. 24)

Werden personenbezogene Informationen zur Bewertung von Individuen herangezogen, so ist diesen das Recht zuzugestehen, die zu Grunde liegenden Daten zu verifizieren (Bird et al. 2005, S. 24), um Fehleinschätzungen der Leistung des Personals durch das Management zu verhindern. Wo großer Wert auf indikatorenbasierte Verfahren zur Leistungsmessung gelegt wird, liegt oftmals auch ein Anreiz zur Manipulation der zu Grunde liegenden Daten vor. Dies kann der Fall sein, wenn beispielsweise monetäre Anreize für das Erreichen bestimmten Kennzahlen gesetzt werden. Dies nennt man „game playing“ (Bruijn 2002, S. 581).

Darüber hinaus ist schon die für eine Leistungsmessung notwendige Rollenverteilung mit ethischen (und moralischen) Problemen behaftet, denn die am Evaluierungsprozess beteiligten Personen schlüpfen in eine Rolle, die eine bestimmte Rollenmoral mit sich bringt. So wird vom mittleren Management und von Angestellten erwartet, dass sie Daten liefern, die zur Messung ihrer eigenen Leistung dienen. Da Evaluatoren oft nicht in der Lage sind, die berichteten Daten zu verifizieren, müssen die Evaluierten zwischen den Polen Eigeninteresse (*self-interest*), Rollenmoral (*role morality*) und allgemeiner Moral (*common-sense morality*) abwägen. Stimmen sie den Kriterien oder Methoden der Evaluierung nicht zu, liegt eine Manipulation der Daten nahe. (Kerssens-van Drongelen und Fisscher 2003, S. 53)

Die Evaluatoren wiederum könnten vor den negativen Folgen eines von ihnen berichteten Fehlhandelns zurückschrecken, besonders, wenn sie eine beratende Tätigkeit für die Einrichtung einnehmen (ebd.).

4 Problematische Aspekte der Bibliometrie

Das ‚New Public Management‘ führt zunehmend dazu, dass auch akademische Einrichtungen ihre Mittel auf der Basis von Kennzahlen verteilen. Indikatoren, die auf wissenschaftlichem Output beruhen (z.B. akquirierte Drittmittel, Publikationen, Ergebnisse von Lehrevaluationen) spielen eine zunehmende wichtigere Rolle in Berufungs- und Einstellungsverfahren (Jungbauer-Gans und Gross 2012, S. 245). Bibliometrische Methoden werden zur Messung der Qualität von publizierter Forschung verwendet. Je öfter eine Publikation zitiert wird, desto höher wird ihre Qualität eingeschätzt (Adam 2002, S. 726). Wie oben festgestellt wurde, ist eine valide Datenbasis hierfür eine notwendige Voraussetzung. Einige in der Wissenschaft verbreitete Praktiken stehen diesem Anliegen entgegen.

Zuerst zu nennen sind dabei inkorrekte Angaben zur Autorenschaft.

„[S]cientific authorship is most likely the most central pillar of science publishing ethics.“ (Teixeira da Silva und Dobránszki 2016, S. 1459)

Der Deutsche Hochschulverband (2011) nennt verschiedene Rechtsgrundlagen, die für die Autorenschaft wissenschaftlicher Publikationen heranzuziehen sind. So ergibt sich aus dem Urheberrecht (§ 8 UrhG) die Bestimmung, dass sich Urheberschaft aus der konkreten Ausgestaltung des Textes ableitet, wissenschaftliche Beteiligung ohne Mitarbeit am Text ist hier nicht ausreichend. Das Hochschulrahmengesetz (HRG) legt fest, dass auch Nicht-Urheber als Beteiligte genannt werden sollten. Wer „einen eigenen wissenschaftlichen oder wesentlichen sonstigen Beitrag geleistet“ hat, sei als Mitautor zu nennen, „soweit möglich, [...] sei ihr Beitrag zu kennzeichnen“ (§ 24 HRG). Neben den rechtlichen bestehen auch wissenschaftsethische Probleme durch die Fehlattribuierung von Autorenschaften, denn diese kennzeichnet den Beitrag einer Person zum wissenschaftlichen Fortschritt, benennt ein persönliches Erfolgserlebnis und ist ein Beweis für intellektuelle Bemühungen eines Individuums. In der Folge fungiert die Autorenschaft als Beitrag zum beruflichen Ansehen einer Person und dient somit dem beruflichen Fortkommen, da Publikationen in Einstellungs- und Berufungsverfahren ausgewertet werden (Bennett und Taylor 2003, S. 264). Drei weit verbreitete irreguläre Praktiken in der Zuordnung von Autorenschaften sind hier zu nennen (vgl. Bennett und Taylor 2003, S. 264–266):

1. *Ghost authorship* bezeichnet die Beteiligung nicht genannter Personen am Artikel. Dies ist beispielsweise der Fall, wenn Autoren, die bei einer Lobbyorganisation oder in der Industrie beschäftigt sind, sich an der Erstellung eines Textes beteiligen, ihre Mitarbeit jedoch verschleiern möchten.
2. *Guest authorship*: Im Gegensatz zur *ghost authorship* wird hier eine nicht existierende Autorenschaft behauptet. Dies kann u.U. eine Ehrenbezeugung gegenüber einem respektierten Wissenschaftler sein, der zwar am Thema des Artikels, nicht aber am Text gearbeitet hat. Oder aber eine fiktive Autorenschaft, bei der sich zum Beispiel eine Institutsleiterin oder ein Doktorvater durch implizite oder explizite Druckausübung in die Autorenliste zwingt, um den eigenen Publikationsoutput zu erhöhen.
3. *Manipulation der Autorenreihenfolge* kann dazu führen, dass die angesehene Erstautorenposition nicht von der maßgeblich verantwortlichen Person eingenommen wird.

Ein weiteres Problem in der Datenbasis kann durch Selbstzitate entstehen. Diese treten immer dann auf, wenn Autoren ihre eigene Publikation zitieren. Im positiven Fall kann dies ein Nachweis sein, dass Forschende sich über längere Zeit einem Thema widmen und dieses weiterentwickeln. Es kann aber auch ein Ausdruck einer auf quantitativen Output optimierten Publikationsstrategie sein. Man spricht hier von *least publishable units* (Broad 1981, S. 1137) oder von der Salamiaktik, also der Aufteilung wissenschaftlicher Erkenntnisse auf möglichst viele Publikationen: Der Inhalt wird dünn geschnitten wie Salamischeiben (Binswanger 2011, S. 9; Taubert und Weingart 2016, S. 23):

„Dies kann für die Allgemeinheit sogar fatale Folgen haben. Wenn beispielsweise mehrere Publikationen ein neues Medikament als medizinisch unbedenklich beschreiben, sich aber alle diese Publikationen auf das gleiche Experiment stützen, wiegt man die Öffentlichkeit in einer (häufig gewollten) falschen Sicherheit.“ (Binswanger 2011, S. 9)

Da einfache Selbstzitate vergleichsweise einfach zu entdecken sind, greifen einige Wissenschaftler zu Manipulationen, um die Zahl der von ihnen gesammelten Zitationen zu erhöhen. Eine beliebte Methode zur Verschleierung ist das Zitationskartell, bei der eine Gruppe von Wissenschaftlern sich darauf verständigt, sich gegenseitig zu zitieren (vgl. Stock 2010, S. 268).

All dem zu Grunde liegt oft das in den letzten Jahrzehnten stark zunehmende Phänomen der Mehrautorenschaft (Adams 2012).

„Der Einsatz von bibliometrischen bzw. auf dem Publikationsoutput basierenden Indikatoren in Steuerungsverfahren könnte daher zu einer Erhöhung von in Mehrautorenschaft verfassten Publikationen führen“ (Tafertshofer 2017, S. 94)

Diese Befürchtung scheint zuzutreffen, ist in einigen Disziplinen inzwischen sogar von Hyperautorenschaft die Rede, bei der eine Publikation von wenigen Seiten von mehreren Tausend Personen gemeinsam verfasst wurde (Cronin 2001). Dies ist reizvoll für Wissenschaftler, da durch die Mitarbeit von mehreren (oder vielen) Personen schneller eine Beteiligung an einer höheren Zahl an Publikationen zu Stande kommen kann, was sich wiederum positiv auf den messbaren Output jedes einzelnen auswirkt. Teixeira da Silva und Dobránszki (2016) führen dieses Dilemma zum Teil auf einen Mangel an einer universellen Definition wissenschaftlicher Autorenschaft zurück:

„This is because authors within such complex groups come from different cultures, and from different research institutes in which distinct codes of conduct (CoCs) may be in place. Moreover, it is not uncommon for manuscripts that are rejected by one journal to be accepted in another, and should the authorship definitions as defined by the publishers of both journals have differing or contrasting definitions, then a moral or ethical dilemma would arise.“ (Teixeira da Silva und Dobránszki 2016, S. 1458)

Wo kein gemeinsames Verständnis des Autorenbegriffs existiert, sind Missverständnisse vorprogrammiert. Es scheint einfacher, gegen nicht-universelle Bestimmungen zu verstoßen.

Das große Gewicht, das Zitationen beigemessen wird, birgt weitere Probleme. In der Regel wird das Vorhandensein vieler Zitationen als Qualitätsmerkmal verstanden. Die Motivationen zur Zitation sind jedoch überaus vielfältig. Case und Higgins (2000, S. 641) untersuchten die verbreitetsten Motivationen. Nach dem Zitieren von Klassikern wurden am zweithäufigsten soziale Gründe für eine Zitation genannt. Die Zitation solle unter Beweis stellen, dass die Autoren relevante (und oftmals in einem prestigeträchtigen Journal erschienene) Literatur in ihrem Feld kennen. Weitere Zitationsmotivationen seien unter anderem negative Zitationen, bei dem auf Schwächen des zitierten Werkes hingewiesen wird. Auch wird auf „oberflächliche“ (*perfunctory*) Zitationen hingewiesen, bei denen nur oberflächlicher Bezug zum Artikel besteht, beziehungsweise der Bezug im Text nicht deutlich ist (Lin 2018). Der auf bibliometrischer Analyse begründete Wert eines Zitats wird jedoch nicht in dem Kontext verstanden, aus dem er stammt. Vielmehr wird

einfach gezählt – je mehr, desto besser. Es zählt die reine Aufmerksamkeit, die eine Publikation erzielt, gleich, ob es sich um positive oder negative handelt. Dies verleitet zu sensationalisierter Forschung.

Eine andere von Case/Higgins aufgeführte Motivation verweist auf einen weiteren kritischen Aspekt. Der Wert, der Zitationen beigemessen wird, verleitet zu ethisch fragwürdigem Verhalten von Begutachtern. Gutachter im wissenschaftlichen Begutachtungsprozess (Peer Review) können ihre Machtposition ausnutzen, um sich selbst (oder Mitgliedern ihres Zitationskartells, s.o.) einen Vorteil zu verschaffen.

„Peer Review bedeutet, dass eine Begutachtung von wissenschaftlichen Leistungen und Akteuren in geregelten Verfahren von Fachkollegen vorgenommen wird. Betroffen sind Forschungsanträge, Manuskripte und Kandidaturen für Stellen und Preise, auch (wenngleich seltener) die Universitätslehre. Gesichert werden soll, dass wissenschaftliche Qualität wissenschaftlich kompetent eingeschätzt wird.“ (Neidhardt 2010, S. 281)

Da diese Verfahren bisher in der Regel anonymisiert (*blind*, bzw. *double-blind*) durchgeführt werden, können die Gutachter den Autoren des von ihnen begutachteten Werks nahelegen, eine bestimmte Publikation zu zitieren. In der Hoffnung auf eine erhöhte Chance auf Veröffentlichung besteht ein gewisser Druck, dieser Empfehlung nachzukommen. Case und Higgins bezeichnen die Auswahl eines Artikels aus diesem Grund als „was authored by someone who might have been influential in the review process“. Hier liegt also zum Teil eine Art vorauseilender Gehorsam vor, teils jedoch auch Reaktion auf eine direkte Aufforderung im Begutachtungsprozess. In manchen Fällen wird gar von Herausgebern dazu aufgerufen, die von ihnen herausgegebene Zeitschrift zu zitieren, um die über Indikatoren wie den JIF gemessene Relevanz der Zeitschrift künstlich zu erhöhen (Thombs und Razykov 2012).

Nicht außer Acht gelassen dürfen zuletzt die Auswirkungen, die bibliometrische Methoden auf die Wissenschaftler haben, deren Publikationen vermessen werden. Für den Fall der Evaluation von Forschungsinstitutionen gibt es Indizien, dass diese einen Stressfaktor für die Angehörigen der Organisationen darstellen. So beschreibt beispielsweise Howard Newby, ehemaliger Direktor des Higher Education Funding Council for England (HEFCE), die das Research Assessment Exercise (RAE) durchführen, negative Auswirkungen sowohl auf Forschungspraktiken, aber auch als stressverursachend für die Forschende (Lucas 2010, S. 39).

Eine weitere, sehr konkrete Auswirkung auf Wissenschaftler hat die Vorhersage des wissenschaftlichen Erfolgs Einzelner auf Basis bibliometrischer Indikatoren (Bartheld et al. 2015; Acuna et al. 2012). Besonders Metriken auf Autorenebene werden hierzu verwendet. Das Ziel ist, schon bei jungen Wissenschaftlern späteren Erfolg oder Misserfolg zu prognostizieren, um auf dieser Basis Entscheidungen über (Weiter-)Beschäftigung zu treffen. Die Probleme, die sich hieraus ergeben, liegen auf der Hand. Der Anreiz der Wissenschaftler kann gar nicht mehr darauf liegen, fundierte Wissenschaft zu betreiben. Stattdessen muss – mit den oben skizzierten Mitteln

wie z.B. der Salami-Taktik – möglichst hohe Kennziffern erreicht werden. Auch neuere Metriken können diesen Wettbewerb um hohe Punktzahlen befeuern:

“Some fear that altmetrics will introduce a new form of competition not based on scientific quality.” (Wilsdon et al. 2017, S. 12)

In diesem Fall verschiebt sich der Wettbewerb, neue Anreize zu unethischem Verhalten werden gesetzt (Cheung 2013).

5 Fazit

Bibliometrische Analysen und Indikatoren können Aussagen über die Aufmerksamkeit treffen, die eine Publikation, eine Zeitschrift und auch eine im wissenschaftlichen Publizieren engagierte Person erzielt. Die Methoden, um diese Aufmerksamkeit zu berechnen, sind je nach Einsatzszenario genau abzuwägen und immer wieder zu evaluieren. Insbesondere dürfen Metriken nicht auf ein Gebiet angewendet werden, für die sie nicht gedacht sind. Ein besonders prominentes negatives Beispiel ist der JIF, der trotz massiver Evidenz der Untauglichkeit bei der Beurteilung von wissenschaftlicher Exzellenz von Artikeln oder Personen genau hierfür immer noch verwendet wird.

Mit unterschiedlichen Methoden wird versucht, vieler der in dieser Arbeit skizzierten Fehlanreize entgegen zu wirken. Um Anreize zur Veröffentlichung von möglichst vielen Werken zu entschärfen, reduzieren Forschungsförderer die Anzahl der maximal erlaubten erwähnten Publikationen in einem Antrag. Die Deutsche Forschungsgemeinschaft (DFG) zum Beispiel lässt pro antragstellender Person höchstens zehn Publikationen zu.² Ein anderer Ansatz ist die genaue Zuordnung der Art einer Beteiligung an einem Werk. Das Consortia Advancing Standards in Research Administration Information (CASRAI) hat dazu CRediT (Contributor Role Taxonomy) entwickelt. Diese Taxonomie definiert 14 verschiedene Rollen, die an einem wissenschaftlichen Werk beteiligte Personen einnehmen können und mit denen auch Aktivitäten wie Drittmittelakquise, das Entwickeln wissenschaftlicher Software oder die Kuratierung von Forschungsdaten abgedeckt sind (Brand et al. 2015).

Dazu mühen sich derzeit mehrere Initiativen, ein Bewusstsein für die oftmals ethisch problematische Verwendung der Bibliometrie zu schaffen. So ließ zum Beispiel die Europäische Kommission die Auswirkungen von Altmetrics untersuchen (Wilsdon et al. 2017), man spricht dort gar von Responsible Metrics. Im Report „The Metric Tide“ (Wilsdon et al. 2015) ließ die HEFCE untersuchen, wie man verantwortlich mit Metriken umzugehen habe.

“There is legitimate concern that some quantitative indicators can be gamed, or can lead to unintended consequences; journal impact factors and citation counts are two prominent examples.” (Wilsdon et al. 2015, VIII)

2 http://www.dfg.de/formulare/54_01/54_01_de.pdf, abgerufen am 11. Dezember 2018

Damit verknüpft sei die oft intransparente Datenbasis und nicht durchschaubare Verfahren in denen bibliometrische Indikatoren verwendet würden. In den beiden zuletzt erwähnten Reports wird auch von *Responsible Metrics* gesprochen, also verantwortungsvollen Metriken.

“Robustness: basing metrics on the best possible data in terms of accuracy and scope;

Humility: recognising that quantitative evaluation should support – but not supplant – qualitative, expert assessment;

Transparency: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results;

Diversity: accounting for variation by field, and using a range of indicators to reflect and support a plurality of research and researcher career paths across the system;

Reflexivity: recognising and anticipating the systemic and potential effects of indicators, and updating them in response.” (Wilsdon et al. 2015, X)

Doch auch von Seite der Forschenden wird dem Phänomen der missbräuchlich genutzten Metriken vermehrte Aufmerksamkeit gewidmet. Die San Francisco Declaration of Research Assessment (DORA, Cagan 2013) wurde schon von über 700 Institutionen und mehr als 13.000 Wissenschaftlerinnen und Wissenschaftlern unterzeichnet.³ In dieser Deklaration werden verschiedene Empfehlungen ausgesprochen, die sich an verschiedene Akteure in der Wissenschaft richten: Forschungsförderer, Verlage, wissenschaftliche Institutionen, Anbieter von bibliometrischen Informationen und Wissenschaftler selbst. In der Deklaration wird klar, dass alle Akteure gemeinsame Verantwortung tragen, damit bibliometrische Indikatoren verantwortungsvoll genutzt werden können. Unter anderem sollen Forschungsförderer und wissenschaftliche Institutionen auch nicht-konventionelle Publikationen wie Software oder Forschungsdaten zur Evaluierung hinzuziehen, vor allem aber sollen sie deutlich machen und sicherstellen, dass die Qualität einer wissenschaftlichen Publikation wichtiger ist als der gemessene Impact. Dies heißt mit anderen Worten: Bei einer Evaluierung muss man sich mit den Inhalten der Publikationen auseinandersetzen und deren Qualität beurteilen, nicht nur Zitierungen der Publikation zählen. Verlage sollen die Überbetonung des JIF beenden und Anbieter von bibliometrischen Informationen Missbrauch öffentlich ächten, Wissenschaftler sollen sich ihrer eigenen Verantwortung bewusst werden und als Gutachter oder in Berufungskomitees nicht auf Basis von bibliometrischen Indikatoren urteilen. Über allem steht die gemeinsame Forderung an alle Akteure in der Wissenschaft:

³ <https://sfдора.org/signers/>, Stand: 16. Dezember 2018

“Do not use journal-based metrics, such as journal impact factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion or funding decisions.” (Cagan 2013, S. 869)

Das *Leiden Manifesto for Research Metrics* (Hicks et al. 2015) schlägt in die gleiche Kerbe und stellt 10 Prinzipien für den verantwortungsvollen Umgang mit Leistungsindikatoren in der Forschungsevaluierung auf. Die von Konkiel (2018) beschriebene HuMetricsHSS-Initiative geht noch weiter und spricht nicht nur von der Notwendigkeit von verantwortungsvollen Metriken (*responsible metrics*), hier werden *humane metrics* gefordert. Die Entwicklung von Kriterien für den verantwortungsvollen Umgang mit (bibliometrischen) Indikatoren ist freilich nur der Anfang. Ob all diesen Forderungen in der Praxis nachgekommen wird, bleibt abzuwarten.

Darüber hinaus bleibt viel Forschungsbedarf, ob und wie quantitative wissenschaftlerfreundliche und wissenschaftsadäquate Leistungserfassung und -bewertung stattfinden darf und kann. Als Ausgangspunkt können hier die Veröffentlichungen von Cronin und Sugimoto (2014), Haustein und Larivière (2015) sowie Bornmann (2017) dienen. Sowohl die Metriken, ihre möglichst offengelegte und allen zugängliche Datenbasis als auch die Anwendungspraxis kann und sollte erforscht und verbessert werden.

Literatur

Acuna, Daniel E.; Allesina, Stefano; Kording, Konrad P. (2012): Future impact. Predicting scientific success. In: *Nature* 489 (7415), S. 201–202. DOI: [10.1038/489201a](https://doi.org/10.1038/489201a).

Adam, David (2002): The counting house. In: *Nature* 415 (6873), S. 726–729. DOI: [10.1038/415726a](https://doi.org/10.1038/415726a).

Adams, Jonathan (2012): Collaborations: The rise of research networks. In: *Nature* 490 (7420), S. 335–336. DOI: [10.1038/490335a](https://doi.org/10.1038/490335a).

Ball, Rafael; Tunger, Dirk (2005): Bibliometrische Analysen - Daten, Fakten und Methoden. Grundwissen Bibliometrie für Wissenschaftler, Wissenschaftsmanager, Forschungseinrichtungen und Hochschulen. Jülich: Forschungszentrum Jülich; Zentralbibliothek (Schriften des Forschungszentrums Jülich Reihe Bibliothek = Library, 12). Online verfügbar unter <http://hdl.handle.net/2128/381> zuletzt geprüft am 06.05.2018.

Bartheld, Christopher S. von; Houmanfar, Ramona; Candido, Amber (2015): Prediction of junior faculty success in biomedical research: comparison of metrics and effects of mentoring programs. In: *PeerJ* 3, e1262. DOI: [10.7717/peerj.1262](https://doi.org/10.7717/peerj.1262).

Bennett, Dianne M.; Taylor, David McD (2003): Unethical practices in authorship of scientific papers. In: *Emerg Med Australas* 15 (3), S. 263–270. DOI: [10.1046/j.1442-2026.2003.00432.x](https://doi.org/10.1046/j.1442-2026.2003.00432.x).

Binswanger, Mathias (2011): Der Publikationswettbewerb in der Forschung: Arroganz, Ignoranz, Redundanz (LIFIS Online - Internet-Zeitschrift des Leibniz-Instituts für interdisziplinäre Studien e.V.). Online verfügbar unter http://leibniz-institut.de/archiv/binswanger_05_04_11.pdf, zuletzt geprüft am 21.05.2018.

Bird, Sheila M.; Cox, David; Farewell, Vern T.; Goldstein, Harvey; Holt, Tim; Smith, Peter C. (2005): Performance Indicators: Good, Bad, and Ugly. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 168 (1), S. 1–27, Online verfügbar unter: <https://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/performance-indicators-report-jrssa.pdf>, zuletzt geprüft am 18.05.2018.

Biswal, Akshaya Kumar (2013): An absolute index (Ab-index) to measure a researcher's useful contributions and productivity. In: *PloS one* 8 (12), e84334. DOI: [10.1371/journal.pone.0084334](https://doi.org/10.1371/journal.pone.0084334).

Bornmann, Lutz (2013): What is societal impact of research and how can it be assessed? a literature survey. In: *J. Am. Soc. Inf. Sci.* 64 (2), S. 217–233. DOI: [10.1002/asi.22803](https://doi.org/10.1002/asi.22803).

Bornmann, Lutz (2017): Measuring impact in research evaluations: a thorough discussion of methods for, effects of and problems with impact measurements. In: *Higher Education* 73 (5), S. 775–787. DOI: [10.1007/s10734-016-9995-x](https://doi.org/10.1007/s10734-016-9995-x).

Bradford, Samuel C. (1934): Sources of information on specific subjects. In: *Engineering* 137, S. 85–86.

Brand, Amy; Allen, Liz; Altman, Micah; Hlava, Marjorie; Scott, Jo (2015): Beyond authorship: attribution, contribution, collaboration, and credit. In: *Learn. Pub.* 28 (2), S. 151–155. DOI: [10.1087/20150211](https://doi.org/10.1087/20150211).

Broad, William J. (1981): The Publishing Game: Getting More for Less. In: *Science* 211 (4487), S. 1137–1139.

Bruijn, Hans de (2002): Performance measurement in the public sector: strategies to cope with the risks of performance measurement. In: *Intl Jnl Public Sec Management* 15 (7), S. 578–594. DOI: [10.1108/09513550210448607](https://doi.org/10.1108/09513550210448607).

Cagan, Ross (2013): The San Francisco Declaration on Research Assessment. In: *Disease models & mechanisms* 6 (4), S. 869–870. DOI: [10.1242/dmm.012955](https://doi.org/10.1242/dmm.012955).

Case, Donald O.; Higgins, Georgeann M. (2000): How can we investigate citation behavior? A study of reasons for citing literature in communication. In: *Journal of the Society for Information Science and Technology* 51 (7), S. 635–645. DOI: [10.1002/\(SICI\)1097-4571\(2000\)51:7<635::AID-ASI6>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1097-4571(2000)51:7<635::AID-ASI6>3.0.CO;2-H).

Cheung, Man Kit (2013): Altmetrics: Too soon for use in assessment. In: *Nature* 494 (7436), S. 176. DOI: [10.1038/494176d](https://doi.org/10.1038/494176d).

Cronin, Blaise (2001): Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? In: *Journal of the Association for Information Science and Technology*. 52 (7), S. 558–569. DOI: [10.1002/asi.1097](https://doi.org/10.1002/asi.1097).

Deutscher Hochschulverband (2011): Wissenschaftsadäquates Publikationsverhalten. Potsdam. Online verfügbar unter <https://www.hochschulverband.de/877.html>, zuletzt aktualisiert am 12.04.2011, zuletzt geprüft am 21.05.2018.

Fenner, Martin (2013): What can article-level metrics do for you? In: *PLoS biology* 11 (10), e1001687. DOI: [10.1371/journal.pbio.1001687](https://doi.org/10.1371/journal.pbio.1001687).

Garfield, Eugene (1955): Citation indexes for science; a new dimension in documentation through association of ideas. In: *Science* 122 (3159), S. 108–111.

Garfield, Eugene (2006): The history and meaning of the journal impact factor. In: *JAMA: The Journal of the American Medical Association* 295 (1), S. 90–93. DOI: [10.1001/jama.295.1.90](https://doi.org/10.1001/jama.295.1.90).

Haustein, Stefanie; Larivière, Vincent (2015): The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects. In: Isabell M. Welp, Jutta Wollersheim, Stefanie

Ringelhan und Margit Osterloh (Hg.): *Incentives and Performance*. Cham: Springer International Publishing, S. 121–139.

Havemann, Frank (2009): Einführung in die Bibliometrie. Berlin, Berlin: Gesellschaft für Wissenschaftsforschung e.V; Inst. für Bibliotheks- und Informationswiss. der Humboldt-Univ.

Heinen, Edmund (1992): Menschliche Arbeit aus betriebswirtschaftlicher Sicht. In: Horst Albach (Hg.): *Globalisierung und Wettbewerb*. In memoriam Alfred Herrhausen. Wiesbaden: Gabler, S. 69–81.

Hicks, Diana; Wouters, Paul; Waltman, Ludo; Rijcke, Sarah de; Rafols, Ismael (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature* 520(7548), 429–431. DOI: [10.1038/520429a](https://doi.org/10.1038/520429a).

Hirsch, Jorge E. (2005): An index to quantify an individual's scientific research output. In: *Proceedings of the National Academy of Sciences of the United States of America* 102 (46), S. 16569–16572. DOI: [10.1073/pnas.0507655102](https://doi.org/10.1073/pnas.0507655102).

Jungbauer-Gans, Monika; Gross, Christiane (2012): Veränderte Bedeutung meritokratischer Anforderungen in wissenschaftlichen Karrieren. In: *Die Hochschule* (2), S. 245–324.

Kerssens-van Drongelen, Inge C.; Fisscher, Olaf A. M. (2003): Ethical dilemmas in performance measurement. In: *Journal of Business Ethics* 45 (1/2), S. 51–63. DOI: [10.1023/A:1024120428490](https://doi.org/10.1023/A:1024120428490).

Konkiel, Stacy (2018). Approaches to creating 'humane' research evaluation metrics for the humanities. In: *Insights the UKSG journal* 31(1), 39. DOI: [10.1629/uksg.445](https://doi.org/10.1629/uksg.445).

Kuhn, Thomas; Weibler, Jürgen (2012): Ethikbewusstes Personalmanagement. erfolgsstrategische Selbstverständlichkeit oder moralische Herausforderung? In: Stephan Kaiser und Arjan Kozica (Hg.): *Ethik im Personalmanagement. Zentrale Konzepte, Ansätze und Fragestellungen*. 1. Auflage. Mering: Rainer Hampp Verlag (Dnwe schriftenreihe), S. 45–63.

Lin, Chi-Shiou (2018): An analysis of citation functions in the humanities and social sciences research from the perspective of problematic citation analysis assumptions. *Scientometrics* 116 (2), S. 797–813. DOI: [10.1007/s11192-018-2770-2](https://doi.org/10.1007/s11192-018-2770-2).

Lin, Jennifer; Fenner, Martin (2013): The many faces of article-level metrics. In: *Bulletin of the American Society for Information Science and Technology* 39 (4), S. 27–30. DOI: [10.1002/bult.2013.1720390409](https://doi.org/10.1002/bult.2013.1720390409).

Lotka, Alfred J. (1926): The frequency distribution of scientific productivity. In: *Journal of the Washington Academy of Sciences* 16 (12), S. 317–323.

Lucas, Lisa (2010): *The research game in academic life*. Maidenhead, England, New York, London: Open University Press.

- Müller, Stefanie (2011): Humankapitalethik. Ein handlungsleitendes Modell zum verantwortungsvollen Umgang mit Humanvermögen. 1. Auflage. Hg. v. Christian Scholz. Mering: Hampp-Verlag (Strategie- und Informationsmanagement).
- Neidhardt, Friedhelm (2010): Selbststeuerung der Wissenschaft: Peer Review. In: Dagmar Simon, Andreas Knie und Stefan Hornbostel (Hg.): Handbuch Wissenschaftspolitik. Wiesbaden: VS Verl. für Sozialwiss, S. 280–292.
- Neylon, Cameron; Wu, Shirley (2009): Article-level metrics and the evolution of scientific impact. In: *PLoS biology* 7 (11), e1000242. DOI: [10.1371/journal.pbio.1000242](https://doi.org/10.1371/journal.pbio.1000242)
- Oberst, Ursula (2017): Measuring the Societal Impact of Research with Altmetrics: An Experiment. In: *027.7 Zeitschrift für Bibliothekskultur* 5 (1), S. 16–21. DOI: [10.12685/027.7-5-1-167](https://doi.org/10.12685/027.7-5-1-167)
- PLoS Medicine Editors (2006): The impact factor game. It is time to find a better way to assess the scientific literature. *PLoS medicine* 3(6), e291. DOI: [10.1371/journal.pmed.0030291](https://doi.org/10.1371/journal.pmed.0030291)
- Stock, Wolfgang G. (2010): Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung. In: Klaus Fuchs-Kittowski, Hubert Laitko, Heinrich Parthey und Walther Umstätter (Hg.): *Wissenschaft und Digitale Bibliothek*. 2. Aufl. Berlin: Gesellschaft für Wissenschaftsforschung e.V. c/o Inst. f. Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin (*Wissenschaftsforschung Jahrbuch*, 1998), S. 239–282. Online verfügbar unter www.wissenschaftsforschung.de/JB98_239-282.pdf
- Tafertshofer, Lorenz (2017): Effekte hochschulinterner indikatorenbasierter Steuerungs- und Anreizverfahren auf das Forschungsverhalten von Professorinnen und Professoren an sozialwissenschaftlichen Fachbereichen. Dissertation. Eberhard-Karls-Universität, Tübingen.
- Taubert, Niels; Weingart, Peter (2016): Wandel des wissenschaftlichen Publizierens – eine Heuristik zur Analyse rezenter Wandlungsprozesse. In: Peter Weingart und Niels C. Taubert (Hg.): *Wissenschaftliches Publizieren. Zwischen Digitalisierung, Leistungsmessung, Ökonomisierung und medialer Beobachtung*. Berlin, Boston: Walter de Gruyter (Forschungsberichte (Berlin-Brandenburgische Akademie der Wissenschaften), Band 38), S. 3–38. Online verfügbar unter: <https://edoc.bbaw.de/frontdoor/index/index/docId/2662>, zuletzt geprüft am 21.05.2018.
- Teixeira da Silva, Jaime A.; Dobránszki, Judit (2016): Multiple Authorship in Scientific Manuscripts: Ethical Challenges, Ghost and Guest/Gift Authorship, and the Cultural/Disciplinary Perspective. In: *Science and engineering ethics* 22 (5), S. 1457–1472. DOI: [10.1007/s11948-015-9716-3](https://doi.org/10.1007/s11948-015-9716-3).
- Thombs, Brett D.; Razykov, Ilya (2012): A solution to inappropriate self-citation via peer review. In: *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 184 (16), S. 1864. DOI: [10.1503/cmaj.120597](https://doi.org/10.1503/cmaj.120597).
- Wilsdon; James; Bar-Ilan; Judit; Frodeman; Robert et al. (2017): Next-generation metrics. responsible metrics and evaluation for Open Science. Report of the European Commission Expert

Group on Altmetrics. European Commission. Brüssel. Online verfügbar unter <https://ec.europa.eu/research/openscience/pdf/report.pdf>, zuletzt geprüft am 21.05.2018.

Wilsdon, James; Allen, Liz; Belfiore, Eleonora; Campbell, Philip; Curry, Stephen; Hill, Steven et al. (2015): The Metric Tide. Report of the Independent Review of the Role of Metrics in Research Assessment and Management: HEFCE. Online verfügbar unter: http://blogs.lse.ac.uk/impactofsocialsciences/files/2015/07/2015_metrictide.pdf, zuletzt geprüft am 24.03.2018.

Autor

Christian HAUSCHKE
Technische Informationsbibliothek (TIB)
Welfengarten 1B
30167 Hannover
<https://www.tib.eu>
christian.hauschke@tib.eu