

Altmetrics zwischen Revolution und Dienstleistung: Eine methodische und konzeptionelle Kritik

Ulrich Herb

Zusammenfassung: Alternative Impact Metriken (Altmetrics) gelten oft als demokratisierende oder gar revolutionäre Verfahren zur Messung wissenschaftlicher Resonanz. Dieser Nimbus wird ihnen vor allem zuteil, da sie regelmäßig als den bekannten zitationsbasierten Verfahren der Impact-Messung (Journal Impact Factor, Hirsch-Index/ h-Index) überlegen angesehen werden. Dies wiederum vor allem, weil sie die Resonanz-Messung von wissenschaftlichen Texttypen erlauben, die von den Zitationsmetriken nicht erfasst werden (z.B. Bücher oder Sammelbände) oder sogar von Objekttypen (z.B. wissenschaftliche Daten und Software), die bisher per se von der Impact-Messung ausgeschlossen waren. Dieser Beitrag prüft, inwiefern Altmetrics aktuell dem Anspruch, eine demokratisierende Revolution, die immer auch eine rabiate Veränderung von Strukturen und Regeln bedeutet, auszulösen, gerecht werden kann oder ob sie nicht doch einfach eine der äußerst kritischen Reflexion bedürftige Dienstleistung sind, da sie neue Verfahren der Kontrolle, Evaluierung und Ökonomisierung der Wissenschaft erlauben.

Schlüsselwörter: Impact, Metriken, Zitationen, Altmetrics, Evaluierung, Ökonomisierung, Wissenschaft

Altmetrics between Revolution and Service: A Methodical and Conceptual Criticism

Abstract: Alternative impact metrics (Altmetrics) are often considered to be democratizing or even revolutionary methods for measuring scientific resonance. This nimbus mainly surrounds them because they are regarded as metrics that outclass the well-known citation-based metrics (e.g. the Journal Impact Factor or Hirsch-Index/ h-Index). This in turn happens mainly due to the fact that Altmetrics allow the resonance measurement of scientific document types (e.g. books or anthologies) or even object types (e.g. scientific data and software) that were previously excluded from the impact measurement. This contribution examines to what extent Altmetrics are sparking off a democratizing revolution, which necessarily always

implies a rigorous change in structures and rules, or whether they are simply not more than a service that has to be considered highly critical as they offer new tools to control, evaluate and economize science.

Keywords: impact, metrics, citations, altmetrics, evaluation, economization, science

Der Impact-Boom

Die Relevanz-Messung wissenschaftlichen Outputs nimmt stetig zu: Kannte man vormals mit dem in den 60er Jahren begründeten Journal Impact Factor (JIF) nur ein einziges Verfahren zur Bestimmung der Relevanz wissenschaftlicher Publikationen, traten zusehends und schon nahezu viral neue Verfahren auf den Plan. Nur ein Bruchteil der im Folgenden genannten Modelle wird in diesem Artikel näher vorgestellt und zwar der JIF, der Hirsch- oder h-Index sowie das Konzept der Altmetrics. Die restlichen Techniken werden an dieser Stelle nur erwähnt, um den Leserinnen und Lesern vor Augen zu führen, mit welchem Vermessungseifer Wissenschaft zu Leibe gerückt wird und unter wie vielen Mikroskopen versucht wird, die Bedeutung, Relevanz oder Wichtigkeit (vulgo Impact) von Wissenschaft zu betrachten.

War der JIF, wie erwähnt, lange Zeit die einzig wirklich gebräuchliche Impact-Metrik, schießen seit Mitte der 2000er Jahre immer neue Methoden der Relevanzmessung aus dem Boden: z.B. der Hirsch-Index (h-Index) im Jahr 2005, Eigenfactor 2007, der SCImago Journal Rank (SJR) 2008 oder die Metrik Source Normalised Impact per Paper (SNIP) 2010. Während die genannten Verfahren allesamt Informationen über Zitationshäufigkeiten von Journalen oder Artikeln auswerten, um Relevanz und Impact zu bestimmen, rückten ebenfalls Mitte der 2000er Jahre auch Modelle in den Vordergrund, die Nutzungshäufigkeiten (Brody, Harnad, und Carr 2005) oder (basierend auf Netzwerkanalysen) Nutzungsmuster (Bollen, Van De Sompel, und Rodriguez 2008; Bollen, Van de Sompel, et al. 2009; Bollen et al. 2005) elektronischer wissenschaftlicher Dokumente analysierten. 2010 betraten dann schließlich die Altmetrics den Ring, alternative Metriken, die eine Vielzahl an gänzlich unterschiedlichen Informationen über die Verwendung oder Erwähnung wissenschaftlicher Texte (teils auch Daten oder Software) heranzogen, um deren Impact oder den Impact der die Texte, Daten oder Software produzierenden Wissenschaftler zu erfassen.

Ein möglicher Grund für die epidemische Verbreitung dieser Verfahren der Impact-Messung mag der sich ebenfalls fortlaufend steigende Output an wissenschaftlichen Publikationen sein, so ermittelten Ware & Mabe (2009) ungefähr konstante jährliche Wachstumsraten von ca. 3,5 % für Journale bzw. 3 % für Artikel in den letzten drei Jahrhunderten. Dieser Trend dürfte durch

das Aufkommen der sogenannten Mega-Journale¹ sehr gesteigert werden. Mega-Journale unterziehen Artikel einer Peer Review, die weniger selektiv als gewöhnlich ist: Unter anderem sind die Neuartigkeit der Befunde sowie die fachliche Relevanz in aller Regel keine Kriterien, deren Fehlen gegen eine Annahme von Einreichungen sprechen, wichtig sind in erster Linie Faktoren wie die Wissenschaftlichkeit und die methodische Korrektheit des Artikels sowie die Übereinstimmung der berichteten Daten-/Forschungslage mit den gezogenen Schlussfolgerungen. Überdies sind Mega-Journale auch inhaltlich nicht sehr festgelegt und decken meist ein breites Spektrum an Disziplinen ab. Alleine das Mega-Journal PLOS One publizierte im Jahr 2015 sage und schreibe 29.815 Artikel.

Die Bewertung der Relevanz einer solchen Masse an Publikationen (und abgeleitet: die Bewertung der Relevanz der diese Veröffentlichungen schaffenden Wissenschaftler) scheint qualitativ oder intellektuell, sprich durch Lesen, nicht mehr möglich. Der Versuch, mittels quantitativer Informationen die Relevanz einer Publikation zu beschreiben, erfolgt daher in aller Regel in Form von Impact-Messung durch unterschiedliche Metriken. Diese Metriken zielen auf unterschiedliche Objekte (Journale, einzelne Texte, Personen, teils Daten oder Software) und versuchen anhand quantitativer Informationen eine Aussage über die Wirkung (den Impact) dieses Objekts zu geben. Die gängigen und verbreiteten Impact-Metriken nutzen dabei Zitationsraten oder –häufigkeiten als Impact-Indikator.

Zitationsbasierte Impact-Metriken

Auch wenn Wissenschaftler, Forschungsförderer und Wissenschaftsbürokraten vermutlich stets meinen, alle Diskutanten teilten dieselbe Vorstellung oder gar Definition dessen, was wissenschaftlicher Impact ist, herrschen darüber, was die Wirkung von Wissenschaft und Wissenschaftlern ausmacht, unterschiedliche Vorstellungen. Sie wird mit so unterschiedlichen Phänomenen wie *Produktivität*, *Resonanz* oder *Qualität* gleichgesetzt (Fröhlich 1999, S. 30). Während die Produktivität rein quantitativ zu verstehen ist und sich allein an der Menge hergestellter wissenschaftlicher Texte orientiert, wird versucht die Resonanz einer Publikation anhand der sie zitierenden Veröffentlichungen zu ermitteln. Diese ebenfalls rein quantitative Information der Zitationszahlen wird wiederum als Indikator für die Qualität eines wissenschaftlichen Werkes herangezogen. Der vereinfachende interpretative Dreisatz lautet:

1 Das erste Mega-Journal, PLOS One, ging 2006 online und fand rasch Nachahmer, unter anderem: Scientific Reports (der Nature Publishing Group) 2011, SAGE Open 2011, SpringerPlus 2012, IEEE Access 2013, PeerJ 2013, Heliyon (des Verlags Elsevier) 2015.

- Der Impact (oder die Wirkung) einer Publikation ist gleichzusetzen mit ihrer Qualität.
- Die Qualität einer Publikation ist gleichzusetzen mit ihrer Resonanz.
- Die Resonanz einer Publikation ist gleichzusetzen mit ihrer Zitationshäufigkeit.

Tautologische Definition

Dieses erkenntnistheoretische und methodische Manko der Impact-Messung beschreiben Bollen, Van De Sompel, et al. (2009) treffend als Tautologie der Impact-Messung: „In fact, we do not even have a workable definition of the notion of ‘scientific impact’ itself, unless we revert to the tautology of defining it as the number of citations received by a publication.“ Zusätzlich merken die Autoren zurecht an, dass der Versuch der Quantifizierung von Qualität scheitert, da Qualität ein mehrdimensionales Konstrukt ist, dessen Reduktion auf eine Ziffer fehlschlagen muss: “Our results indicate that the notion of scientific impact is a multi-dimensional construct that can not be adequately measured by any single indicator” (Bollen, Van De Sompel, et al. 2009). Die Bewertung von Wissenschaft und Wissenschaftlern mittels solcher Verfahren entspricht einer Komplexitätsreduktion unter Verwendung methodisch nicht fundierter Kennziffern.

Gläser und Laudel (2001) kritisieren ebenfalls die Wahrnehmung, Zitationsinformationen seien ein Indikator für das nicht näher definierte Phänomen Impact, das wiederum als Äquivalent zu Qualität interpretiert wird. Diese Interpretation wird nach den Autoren durch die Wissenschaftspolitik und –bürokratie gefördert: “Firstly, it must be established what can be measured by citations, i.e., which sociologically relevant features of scientific work and interactions between scientists are reflected by citations. Scientometrics has narrowed down this discussion to the question of whether citations measure impact and whether impact can be related to quality or performance. This focus has been facilitated by an increasing orientation of scientometrics towards science policy.” (Gläser und Laudel 2001, S. 428).

Zitationsbasierte Impact-Metriken in der Kritik²

Es existierten umfassende Beschreibungen der Unmöglichkeit, wissenschaftliche Qualität mittels Zitationen zu erfassen (in Übersicht Herb 2015a, Kapitel B.4) sowie der dysfunktionalen Effekte, die Konstrukte wie der Journal Impact Factor (JIF) oder der Hirsch-Index (h-Index) auf das wissenschaftliche Publikationssystem haben (in Übersicht Brems, Button, und Munafò 2013). Beide Verfahren, JIF und h-Index, sind in der Bewertung wissenschaftlicher Leistungen

2 Von den eingangs erwähnten Spielarten zitationsbasierter Impact-Messung werden hier nur der Journal Impact Factor und der Hirsch-Index näher beschrieben, da sie in ihrer Wahrnehmung und ihrer Wirkung auf die Bewertung von Wissenschaft sowie den sich daraus ergebenden reaktiven Ausrichtungen der Publikationsentscheidungen von Wissenschaftlern die anderen zitationsbasierten Verfahren überragen.

besonders verbreitet und dennoch sind ihre Berechnung und Konstruktion erstaunlich oft unbekannt, daher werden beide Verfahren im Folgenden kritisch dargestellt.

Journal Impact Factor JIF

Besonders der JIF unterliegt umfangreicher Kritik, er wird anhand der Datenbank Journal Citation Reports (JCR) des Anbieters Thomson Scientific nach folgender Formel berechnet:

$$\frac{\text{Zahl der Zitate im Bezugsjahr auf Artikel (eines Journals) der vergangenen zwei Jahre}}{\text{Zahl der zitierfähigen Artikel des Journals der vergangenen zwei Jahre}}$$

Allerdings werden im Zähler Zitate auf *alle* Artikeltypen eines Journals mitgezählt, im Nenner aber nur die Anzahl der vom Datenbankproduzent als zitierfähig (citable) bezeichneten Typen, zu diesen zählen allein Research Article und Reviews (McKerahan und Carmichael 2012, S. 107; McVeigh und Mann 2009, S. 107). Die nicht-zitierfähigen (non-citable) Items sind umgekehrt alle publizierten Texte eines Journals, die weder Research Article noch Review sind, wie etwa Editorials, Letters, Nachrichten, Buchbesprechungen oder Tagungsberichte. Die Kategorisierung von Artikeln als citable oder non-citable wird jedoch teils als willkürlich dargestellt (The PLOS Medicine Editors 2006). Journale wie Nature, Science, The Lancet oder das New England Journal of Medicine, deren Inhalt zu bis zu einem oder zwei Dritteln aus Kommentaren, Nachrichten oder Berichten besteht (Brown 2004, S. 12), profitieren von dieser Unterscheidung der citable und non-citable Items bei der Ermittlung der JIF-Werte, denn ein hoher Anteil der nicht-zitierfähigen Dokumenttypen führt zu einem höheren JIF-Wert. Zudem ignoriert der JIF Verwertungszyklen in unterschiedlichen Disziplinen und benachteiligt daher Journale aus Fächern mit Zyklen, die länger als zwei Jahre betragen, wie Mathematik, Geistes- und Sozialwissenschaften.

Weiterhin ist sowohl bei JIF als auch h-Index der Scope der Datenbanken sehr beschränkt und die Auswahl der indizierten Materialien den Datenbankproduzenten überlassen (Dong, Loh, und Mondry 2005; The PLOS Medicine Editors 2006). Die JCR schließen komplette Dokumentarten aus (z.B. graue Literatur, Bücher, einen Großteil der Konferenzberichte, etc.). Die üblicherweise zur Berechnung der Indizes benutzen Datenbanken besitzen zudem einen Sprachbias zugunsten englischsprachiger Journale auf, Zeitschriften in anderen Sprachen haben einen niedrigeren JIF, da sie im Sample unterrepräsentiert sind (Archambault et al. 2006; Dong, Loh, und Mondry 2005; Seglen 1998; Van Leeuwen 2013). Auch der Wissenschaftssuchmaschine Google Scholar wird übrigens mitunter eine Überrepräsentation englischsprachiger Literatur nachgewiesen (Clermont und Dyckhoff 2012, S. 8 f.).

Hirsch-Index (h-Index)

Während der JIF den Impact von Journalen ermitteln soll, zielt der h-Index auf die Messung des Impacts von Autoren. Er versucht diesen mittels folgender Formel zu ermitteln (Hirsch 2005, S. 16572, Kursivstellung wie im Original): “A scientist has index h if h of his or her Np papers have at least h citations each and the other $(Np - h)$ papers have $\leq h$ citations each.” Ein Autor hat demnach einen h-Index von sieben, wenn er sieben Schriften veröffentlicht hat, die mindestens sieben Mal zitiert wurden. Auch wenn dieser Ansatz zunächst womöglich plausibel erscheinen mag, da er gewissermaßen die Höhe der Zitationen auf Artikel mit der Anzahl der hochzitierten Artikel eines Wissenschaftlers in Einklang zu bringen versucht, macht ein simples Exempel klar, dass der h-Index keinesfalls Impact-Gerechtigkeit schafft.

Tabelle 1: Rechenexempel zum Hirsch-Index

| Publikation/Wissenschaftler | W1 | W2 | W3 |
|-----------------------------|--------------|--------------|------------------|
| P1 | 3 Zitationen | 5 Zitationen | 1.000 Zitationen |
| P2 | 3 Zitationen | 5 Zitationen | 999 Zitationen |
| P3 | 3 Zitationen | 1 Zitationen | - |
| P4 | 0 Zitationen | - | - |
| P5 | 0 Zitationen | - | - |
| P6 | 0 Zitationen | - | - |
| P7 | 0 Zitationen | - | - |
| P8 | 0 Zitationen | - | - |
| P9 | 0 Zitationen | - | - |
| P10 | 0 Zitationen | - | - |
| h-Index | 3 | 2 | 2 |
| Gesamtzahl Zitationen | 9 | 11 | 1.999 |
| Zitationen pro Publikation | 0,9 | 3,67 | 999,5 |

Während Wissenschaftler W1 einen h-Index von 3 aufweisen kann, erhält W2 nur einen Index von 2, da nur zwei seiner Publikationen mindestens zwei Zitationen erhalten. Allerdings erhält er insgesamt 11 Zitationen und damit mehr als W1, zudem benötigt er nur drei Publikationen, um diese 11 Zitationen zu erhalten, wohingegen W1 insgesamt zehn Publikationen verfassen musste, um neun Zitationen zu ergattern. Anders als W2, der keine nicht-zitierte Publikation aufweist, wurden sieben Publikationen von W1 nicht zitiert. Noch schlechter schneidet W3 ab, der auch nur einen h-Index von zwei hat, aber insgesamt 1.999 mal zitiert wurde und damit um vielfaches mehr als W2 (elf Zitationen), der aber ebenfalls einen h-Index von zwei hat oder gar W1, der zwar nur neun Zitationen, aber sogar einen höheren h-Index als W3 vorweisen kann. Noch skurriler erscheint die Lage, wenn man sich vor Augen führt, dass W1, der den höchsten

h-Index hat, 0,9 Zitationen pro Publikationen aufweisen kann, wohingegen W2 und W3 einen geringeren h-Index haben, obwohl sie wesentlich mehr Zitationen pro Publikation erreichen (3,67 bzw. 999,5).

Jorge E. Hirsch, der den h-Index entwickelte, traf unterschiedliche Aussagen darüber, welche Dimension der Index messen soll. Teils sprach er vom “cumulative impact and relevance of an individual’s scientific research output” (Hirsch 2005, S. 16569), teils abweichend von “importance, significance, and broad impact of a scientist’s cumulative research contributions” (Hirsch 2005, S. 16572).

Anders als der JIF (der markenrechtlich an die JCR gekettet ist), ist der h-Index an keine feste Datenbank gebunden. Letztlich kann er mittels jeder Zitationsdatenbank berechnet werden, gängiger Weise wird er anhand der Datenbanken Web of Science bzw. Scopus oder der Suchmaschine Google Scholar ermittelt – wobei in aller Regel je Datenquelle ein anderer Wert ausgegeben wird.

Zitationsverfahren in der Kritik

Die Kritik an den zitationsbasierten Verfahren der Impactmessung führte unter anderem zur Verabschiedung der San Francisco Declaration on Research Assessment DORA³. DORA wurde initiiert von der American Society for Cell Biology (ASCB), die Unterzeichner fordern die Abkehr vom Journal Impact Factor zur Bewertung wissenschaftlicher Leistung und plädieren für die Nutzung sogenannter Article Level Metrics: „Make available a range of article-level metrics to encourage a shift toward assessment based on the scientific content of an article rather than publication metrics of the journal in which it was published“ (“The San Francisco Declaration on Research Assessment DORA” 2012, S. 2). Überdies bemängeln die Verfasser und Unterzeichner von DORA die Beschränkung der Impact-Messung auf Journalartikel unter Exklusion anderer, durchaus sehr relevanter wissenschaftlicher Objekte textueller und nicht-textueller Art, wie z.B. Forschungssoftware oder Daten (“The San Francisco Declaration on Research Assessment DORA” 2012, S. 1). DORA wurde, Stand 03.02.2016, von 618 Organisationen und 12.789 Personen unterzeichnet. Unter den erstunterzeichnenden Einrichtungen finden sich so namhafte Akteure wie der Wellcome Trust, das Howard Hughes Medical Institute, die Proceedings of The National Academy Of Sciences (PNAS) oder der österreichische Fonds zur Förderung der wissenschaftlichen Forschung (FWF); unter den weiteren Unterzeichnern ist z.B. die Österreichische Akademie der Wissenschaften hervorzuheben.

3 <http://am.ascb.org/dora/> [Zugriff am 03.02.2016]

Closed Data & fehlende Reproduzierbarkeit

Doch nicht nur die methodischen Mängel der zitationsbasierten Verfahren der Impact-Messung oder deren Verwendung zur Bewertung von Wissenschaftlern wecken Kritik. Auch die Nutzung proprietärer und geschlossener Datenbanken zur Ermittlung von zitationsbasierten Impact-Metriken stößt auf Widerstände und provoziert Forderungen nach alternativen Verfahren der Impact-Messung: Die Herausgeber Rossner, Van Epps & Hill (2008; 2007) versuchten händisch die JIF-Werte eigener und fremder Journale zu verifizieren und scheiterten mehrmals dabei. Unter anderem wurden sie vom Datenbank-Produzenten Thomson Scientific, der mit den Journal Citation Reports JCR die Datenbank auflegt, mittels derer der JIF berechnet wird, mit der Aussage überrascht, es existierten mehrere Versionen des JCR – doch egal mit welcher Version des JCR: es tauchten immer neue JIF-Werte für die ausgewählten Journale auf und keiner passte zu den selbstbestimmten Werten. Diese ausbleibende Reproduzierbarkeit lässt Rossner, Van Epps und Hill (2007, S. 1092) die proprietären Datenbank in Frage stellen: „Just as scientists would not accept the findings in a scientific paper without seeing the primary data, so should they not rely on Thomson Scientific’s impact factor, which is based on hidden data.“ Diese Kritik trifft jedoch implizit nicht nur die Zitationsdatenbanken von Thomson Scientific als Anbieter des JCR und des JIF, sondern auch das Angebot Scopus des Thomson-Konkurrenten Elsevier oder Googles Wissenschaftssuchmaschine Google Scholar: Alle der genannten Dienste machen die Daten zu den ermittelten Zitationswerten nicht offen zugänglich, sondern maximal (im Falle Google Scholars) entgeltfrei verfügbar. Eine Re-Analyse der Daten ist so aber nicht möglich. Übereinstimmend fordert auch DORA (“The San Francisco Declaration on Research Assessment DORA” 2012, S. 3) von den Produzenten metrischer Daten: „. Be open and transparent by providing data and methods used to calculate all metrics. (...) Provide the data under a licence that allows unrestricted reuse, and provide computational access to data, where possible.“

Altmetrics?

Sowohl eine Ausweitung der Reichweite an Objekttypen, für die Impact erfasst werden soll, als auch eine Erfassung des Impacts einzelner Objekte versprechen die sogenannten alternativen Metriken oder Altmetrics⁴. Altmetrics wollen Nutzungsinformationen über eine Vielzahl sehr unterschiedlicher wissenschaftlicher Objekte (Textpublikationen, Daten, Software, Poster und andere wissenschaftliche Produkte) aus einer Vielzahl sehr unterschiedlicher Datenquellen auswerten. Unter anderem werden für Altmetrics Daten über die Verbreitung wissenschaftlicher Objekte in sozialen Medien (teils wissenschaftsspezifischer und teils laienhafter Art, zur ers-

4 <http://altmetrics.org/manifesto/> [Zugriff am 03.02.2016]

ten Gattung zählt etwa das Vorkommen von Dokumenten in Literaturverwaltungssystemen, zur zweiten die Erwähnung der Objekte z.B. in Facebook oder Twitter) ausgewertet.

In nicht-wissenschaftlichen Kontexten existieren einige Untersuchungen, die sozialen Medien eine gewisse Vorhersagekraft für Ereignisse und gewissermaßen auch den Impact/ oder Wirkung von Produkten attestieren: Mestyán, Yasseri und Kertész (2013) konnten eine positive Korrelation zwischen der Häufigkeit, mit der noch nicht angelaufene Kinofilme in Twitter erwähnt wurden bzw. die Wikipedia-Einträge dieser gelesen oder bearbeitet wurden, und dem späteren Kassenerfolg der Streifen nachweisen. Bollen, Mao und Zeng (2011) zufolge lassen sich Aktienkurse aus Twitter-Meldungen vorhersagen und Apic, Betts und Russell (2011) entwickelten auf Basis der Muster, mit der Wikipedia-Einträge eines Landes geändert werden, den Disput-Index, der die geopolitische Instabilität dieses Landes vorhersagt. Und schon längst hat die Marktforschung die sozialen Medien für sich entdeckt, so kooperieren Twitter und das Marktforschungsunternehmen Nielsen: Zusammen entwickelte man ein Verfahren mit dessen Hilfe die Resonanz (oder der Impact) von TV-Sendungen im Kurznachrichtendienst gemessen werden kann.⁵ Warum also nicht den Impact wissenschaftlicher Objekte in ähnlicher Manier erfassen?

Anders als die genannten Studien, die jeweils nur wenige Social-Media-Angebote als Datenquellen berücksichtigen, nutzen existierende Altmetrics-Dienste eine Vielzahl an Input-Systemen und -informationen, darunter auch Download-Häufigkeiten oder mitunter die bekannten Zitationen. Wie erwähnt werden teils eher wissenschaftliche Systeme als Quellen herangezogen, teils eher Systeme, die nicht überwiegend von Wissenschaftlern genutzt werden. Zu den letzteren zählen etwa der bereits mehrfach erwähnte Dienst Twitter oder die sozialen Netzwerke Google+ und Facebook. Zu den wissenschaftsnahen Datenquellen gehören Literaturverwaltungen wie Mendeley oder CiteULike (bei denen ermittelt werden kann, wie viele Nutzer ein bestimmtes Werk in ihrer privaten Bibliothek vorhalten), Downloadzahlen von Forschungsdaten (z.B. aus dem Forschungsdatenserver Dryad) oder von Forschungssoftware (z.B. aus dem Software-Host github). Für Software kann allerdings nicht nur die absolute Zahl der Downloads, sondern auch die Häufigkeit der Weiterverwendung/ Anpassung zu neuen Zwecken durch so genannte Forks ermittelt werden. Bei einigen der Quellsysteme können z.B. auch qualitative Bewertungen der Objekte ausgelesen werden, so kann ermittelt werden wie das durchschnittliche Nutzerrating für Forschungssoftware in github ausfällt und als Relevanzindikator verwendet werden. Neben den eher wissenschaftlichen und eher nicht-wissenschaftlichen Quellsystemen gibt es auch Datenquellen, die fallweise der einen oder anderen Kategorie zugeordnet werden können wie etwa Blogs.

5 <http://www.nielsensocial.com/nielsentwittertratings/weekly/> [Zugriff am 03.02.2016]

Altmetrics-Dienste

Es existieren bereits einige Dienste, die Altmetrics-Werten erfassen und als Impact-Informationen aufbereiten; sie ähneln dahingehend gewissermaßen den Anbietern von Zitationsindizes. Die Dienste kennen sehr unterschiedliche Reichweiten, Datenquellen und Geschäftsmodelle, worin sie sich wiederum in einigen Fällen von den Erstellern der Zitationsdatenbanken unterscheiden: Scopus und JCR werten mehr oder weniger zur Journalartikel aus, Google Scholar zumindest auch andere Textobjekte – allerdings auch keine wissenschaftlichen Objekte, die nicht als Text vorliegen.

PLOS ALM

Der Open-Access-Verlag Public Library of Science (PLOS)⁶ bietet für jeden seiner Artikel Altmetrics-Scores an und betreibt zusätzlich mit den Article-Level Metrics ALM-Reports⁷ ein eigenes Portal zur Recherche von Werten zu einzelnen Artikeln. Die Metriken sind umfangreich und umfassen Nutzungsinformationen (z.B. PDF-Downloads und HTML-Downloads), Zitationen (z.B. aus Thomson Scientifics Web of Science, Scopus, CrossRef, Google Scholar oder der Datenbank Pubmed), Impact aus Social Media (z.B. Mendeley, CiteULike, Twitter, Facebook, Blogs, Wikipedia) oder anderen wissenschaftlichen Datenquellen (Erwähnung in Datenservern wie figshare). Als Non-Profit-Verlag macht PLOS ALM-Scores zu allen Artikel soweit möglich kostenfrei und offen zur Weiterwendung zugänglich. Zusätzlich stellt PLOS eine Schnittstelle⁸ bereit, über die die Werte von Artikeln in einzelnen Altmetrics-Quellen maschinell heruntergeladen werden können. Auch ein kompletter Download der ALM-Informationen zu allen PLOS-Artikeln ist möglich.⁹ Diese Daten stehen zum kostenfreien Download bereit, eine weitere Verwendung der Daten z.B. analog der Bedingungen der Open-Source- oder Open-Knowledge-Prinzipien (Herb 2011) ist nicht uneingeschränkt möglich, da Daten proprietärer Anbieter mit ausgegeben werden (z.B. im Falle von Scopus oder des Web of Science). Als Zielgruppen für die Nutzung der Altmetrics-Informationen nennt PLOS neben den Autoren auch Forschungsförderer und Wissenschaftseinrichtungen, womit das Moment der Evaluierung und Fremdbewertung von Wissenschaft auch den PLOS-ALM inhärent ist.

6 <https://www.plos.org/> [Zugriff am 03.02.2016]

7 <http://almreports.plos.org/> [Zugriff am 03.02.2016]

8 <http://api.plos.org/alm/examples/> [Zugriff am 03.02.2016]

9 <http://article-level-metrics.plos.org/plos-alm-data/> [Zugriff am 03.02.2016]

Impactstory

Der Dienst Impactstory¹⁰ entstand aus einer wissenschaftsinternen Initiative der Elektroingenieurin und Bioinformatikerin Heather Piwowar und des Informationswissenschaftlers Jason Priem und stellt wohl am ehesten eine Art wissenschaftlicher Selbstorganisation dar. Die Startphase des Dienstes wurde 2012 mit £17.000 durch die Open Society Foundation unterstützt. Von 2012 bis heute wird der Service von der Alfred P. Sloan Foundation mit einem Zuschuss von \$625.000 gefördert. Zusätzliche Mittel in Höhe von \$300.000 erhielt man von der National Science Foundation. Anders als bei PLOS ALM steht bei Impactstory der Wissenschaftler im Mittelpunkt, der sich entweder für \$60 pro Jahr ein Altmetrics-Profil anlegen kann oder ohne Registrierung einfach eine ephemere Übersicht seiner Resonanz generieren lassen kann. Anders als PLOS ALM berücksichtigt Impactstory nicht nur Artikel, sondern eine Vielzahl an wissenschaftlichen Objekten wie Textpublikationen, Forschungsdaten, Forschungssoftware, Präsentationen.¹¹ Als Input-Informationen werden neben anderen genutzt: Zitationen (aus den Datenbanken PubMed und Scopus), Vorkommen in Literaturverwaltungen (z.B. Mendely, CiteULike, Delicious), Tweets (Erwähnungen in Twitter), Nutzungshäufigkeiten (Anzahl der Ansichten einer Präsentationen auf der Plattform Slideshare, Download von Programmcode bzw. Publikationen von Code auf github oder figshare), Empfehlungen (Raten oder Favorisieren von Daten, Präsentationen oder anderen Items z.B. im Software-Repository github oder auf Slideshare). Genau wie PLOS-Altmetrics erlaubt Impactstory den Zugriff auf einen Großteil der erhobenen Impact-Daten, nicht aber aller, da diese teils aus Fremdsystemen (z.B. Scopus) stammen, so dass ihre Weitergabe an Nutzer rechtlich nicht möglich ist.

Altmetric

Anders als PLOS ALM oder Impactstory verfolgt der Dienst Altmetric¹² eindeutig ambitionierte kommerzielle Interessen und bietet kostenpflichtige Altmetrics-Information für einzelne Artikel unterschiedlicher Verlage. Nutzer können diese Altmetrics-Scores der Artikel entgeltfrei einsehen, die Kosten trägt der Verlag, der die Scores zu den Artikeln einblendet. Auch der so genannte Altmetric-Score unterscheidet Altmetric von PLOS-ALM und Impactstory, denn er reduziert die Resonanz eines Journalartikels in ausgewählten Quellsystemen auf einen quantitativen Wert – dieses Prinzip findet sich bei den beiden bereits vorgestellten Altmetrics-Diensten nicht. In die Berechnung des Scores fließen drei Faktoren mit ein: Die Häufigkeit, mit der ein Text

10 <https://impactstory.org/> [Zugriff am 03.02.2016]

11 Der immer wieder aufkeimenden Kritik an der Berücksichtigung dieses Objekttyps sei entgegengehalten, dass in nicht wenigen Fächern auch das Präsentieren von Postern oder Vorträgen (Talks) als Indikator wissenschaftlicher Bedeutung gilt.

12 <https://www.altmetric.com/> [Zugriff am 03.02.2016]

in einem der Quellsysteme nachgewiesen resp. erwähnt wird, die Art der Quelle (denn diese Quellen sind unterschiedlich gewichtet), Provenienz/ Typus des nachweisenden/ erwähnenden Nutzers (Nachweise/ Erwähnungen durch Wissenschaftler werden höher bewertet als Laiennutzung). Zudem unterscheidet der Dienst zwischen automatisch generierten Erwähnungen eines Artikel (z.B. in Form eines Tweets eines Journal-Accounts in Twitter) und Erwähnungen durch Personen. Mehr noch: Altmetric verspricht tendenziöse Bevorzugung der Publikationen eines Journals durch einzelne Nutzer erkennen zu können und bei der Berechnung des Altmetric-Scores berücksichtigen zu können (Altmetric 2015). Das Verfahren, nach dem diese Unterscheidungen getroffen werden, ist nicht bekannt und kann daher nicht bewertet werden. Als Quellen berücksichtigt Altmetric zusätzlich zu den erwähnten oder diesen ähnlichen Diensten auch Online-Nachrichten. Als Zielgruppen definiert die Website Verlage, Institutionen, Wissenschaftler und Forschungsförderer. Kunden können Daten des Dienstes herunterladen und zu internen Zwecken nutzen, auch ein Vergleich eigener Scores mit denen der Konkurrenz, z.B. anderer Journale, ist möglich.

Plum Analytics

Ähnlich wie Altmetric ist auch der Anbieter Plum Analytics eindeutig kommerzieller Natur. Die Perspektive scheint eher stark für Organisationen gestaltet zu sein, so heißt es auf der Website zu den PlumX Dashboards (Plum Analytics 2016a) „You can organize your research into user-defined groups and hierarchies to analyze and understand what is happening with research in your organization. You can group your metrics by researcher or groups within your organization such as by lab, by subject, by journal or any other group where you need to tell the story of research.“ Auch ein anderes Angebot von Plum richtet sich explizit an Wissenschaftsverwaltungen: PlumX +Grants erlaubt es ihnen, die in Altmetrics gemessene Performanz projektbezogener Publikationen zu protokollieren und für zukünftige Mitteleinwerbungen zu nutzen, Ziel ist es unter Wissenschaftlern der eigenen Einrichtung diejenigen ausfindig zu machen, deren Einreichungen auf aktuelle Projektausschreibungen die vermutlich größten Erfolgsaussichten haben. Das Konzept wirkt etwas dirigistisch, der Werbeslogan sogar sehr martialisch (Plum Analytics 2016b): „Arm your researchers to compete for funding“. Auch der Vergleich mit anderen Einrichtungen (*competitors*) wird durch das Tool PlumX +Benchmarks ermöglicht. PlumX Funding Opportunities dient hingegen dazu, automatisch Projektlinien von Drittmittelgebern mit den Wissenschaftlern der eigenen Einrichtung abzugleichen.

Altmetrics & Impact

Auch wenn Social Media und anderen Datenquellen der Altmetrics mitunter mangelnde Seriosität und ephemerer Charakter attestiert wird, existieren doch einige Studien, die Korrelationen

zwischen den Erwähnungen wissenschaftlicher Texte in diesen Datenquellen und Zitationshäufigkeiten auf Ebene einzelner Artikel belegen.

Bar-Ilan (2012) wies für Publikation aus der Informationswissenschaft positive Korrelation zwischen Mendeley User Counts (der Häufigkeit, mit der Nutzer der Literaturverwaltung Mendeley eine Publikation in ihre Mendeley-Bibliothek aufnehmen) und Zitationen im Web of Science, Scopus und Google Scholar nach. Li und Thelwall (2012) ermittelten für Publikationen aus Genomik und Genetik positive Korrelation zwischen Mendeley User Counts und Zitationswerten aus dem Web of Science, Scopus und Google Scholar. Li, Thelwall und Giustini (2012) wiesen positive statistische Zusammenhänge zwischen Mendeley User Counts und Zitationen nach dem Web of Science auf Basis von 1.613 in Nature und Science im Jahr 2007 publizierter Artikel nach. Shuai, Pepe und Bollen (2012) stießen für wissenschaftliche Dokumente, die als Preprints¹³ zwischen Oktober 2010 und Mai 2011 auf dem Open Access Repository arXiv publiziert wurden, auf positive Korrelationen zwischen Erwähnung in Twitter (Twitter Mentions) und Zitationen nach Google Scholar. Auch Eysenbach (2011) ermittelte zwischen Twitter Mentions und Zitationen nach Google Scholar sowie Scopus für Artikel des Journal of Medical Internet Research JMIR positive Korrelationen.

Impact & Manipulierbarkeit

Neben den Zweifeln an der Seriosität und Tauglichkeit der Altmetrics, wissenschaftlichen Impact zu messen, werden sie auch regelmäßig als manipulierbar dargestellt (Davis 2012; eine Bilanz dieser Diskussion findet sich in Herb 2015a, S. 218 ff.). Diese Kritik trifft auf die Altmetrics zu – in gleichem Maße aber trifft sie allerdings auch auf die Verfahren der zitationsbasierten Impact-Messung zu (Brembs, Button, und Munafò 2013; Smeyers und Burbules 2011).¹⁴ Ein ausführliche Darstellung dieser Manipulationsmöglichkeiten für Altmetrics soll an dieser Stelle aus Platzgründen unterbleiben, einen Einblick gewährt wie erwähnt Davis (2012) – festzuhalten ist vorrangig, dass sich Altmetrics und Zitationsmetriken diesbezüglich prinzipiell nicht unterscheiden.

Ebenso lassen sich ihnen die auch den Zitationszählungen bekannten Matthäus-Effekte nachsagen: Tweets auf Artikel eines prominenten Wissenschafts-Stars fliegen diesem vermutlich in gleichem Maße wie Zitationen auf seine Publikationen zu und sind wohl nicht selten eher Ausdruck einer Self-fulfilling Prophecy als einer objektiven inhaltlichen Prüfung des Inhaltes. Diese Limitierungen scheinen jedoch jedem Versuch der Impact-Messung inhärent, gleich ob diese via Zitationen oder Altmetrics erfolgt.

13 Sprich in einer zur formalen Verlagspublikation präfinalen Form.

14 Auch die bereits erwähnte und zu Trickserien einladende Berücksichtigung der citebale und non-citable Items in der JIF-Berechnung ist hier erwähnenswert.

Aussagekraft der Korrelationen & methodische Fundierung der Altmetrics

Interessanter scheint Frage, was uns denn nun die berichteten positiven Korrelationen zwischen Zitationen und Vorkommen von Publikationen in Altmetrics-Datenquellen nutzen und lehren können.

Unklar bleibt für Altmetrics, ob diese ein Indikator für wissenschaftliche Relevanz, Bedeutung, Wirkung (Impact), Qualität oder die Stärke einer anderen Eigenschaft sind. Da diese Frage auch für Zitationen unbeantwortet ist, lässt sich aus den positiven Korrelationen zwischen Werten in Altmetrics-Quellen und Zitationen nichts weiter ermitteln, als dass zwei Phänomene korrelieren, deren Bedeutung ungeklärt ist. Es ist also sowohl für Zitationen als auch die Altmetrics-Quellen nicht belegt, welche Eigenschaften ihre Häufigkeitswerte eigentlich beschreiben. Überdies ist offen, ob die Scores in den Altmetrics-Diensten sich überhaupt auf die *gleiche* Eigenschaftsdimension wie Zitationen beziehen. Das Auffinden von Korrelationen zwischen beiden Werten garantiert dies nicht: Genauso gut könnte es sich um Scheinkorrelationen handeln oder es könnte eine noch nicht isolierte intervenierende Variable wirken. Aber selbst wenn beide Häufigkeitswerte sich auf dieselbe (wenn auch noch nicht bestimmte) Dimension beziehen, stellt sich die Frage, was man mit Altmetrics gewinnt. Pragmatisch ließen sich zumindest drei Funktionen vorstellen:

- Sie könnten der Validierung der Zitationswerte dienen. Diese ist aber so lange nutzlos, solange man die Eigenschaften nicht benennen kann, deren Ausprägungen beide Arten an Werten messen.
- Sie könnten als Prädiktor der Zitationswerte genutzt werden. Da die Ermittlung von Zitationswerten mit einem gewissen Zeitverzug (Verstreichen einer Generation an Publikationen plus Ermittlung der Zitationen) verbunden ist, könnten Altmetrics dazu herangezogen werden, den (immer noch nicht definierten) Impact gemessen in Zitationen vorherzusagen.
- Sie könnten als offener Datenpool herangezogen werden, der uns kostenlos gleiche Informationen wie kostenpflichtige Zitationsdatenbanken liefert. Dazu müssten die Altmetrics Daten jedoch kostenfrei einsehbar und überprüfbar sein. Dies ist jedoch nur bei PLOS ALM der Fall.

Eine verlässliche Nutzung als Prädiktor setzte allerdings voraus, dass die Korrelation nahezu vollkommen wäre. Diese dürfte aber nicht anzunehmen sein. Möglicherweise werden Korrelationen auch durch Interdependenzen wie Selektionsprozesse produziert und stellen empirische Artefakte dar: So könnte man vermuten, dass unter Umständen Wissenschaftler eher neue Artikel bekannter Autoren rezipieren, da sie annehmen auch diese neuen Artikel seien wichtig, und diese auch eher via Social Media weiterverteilen oder in ihre Mendeley-Bibliothek aufnehmen. Aus dem gleichen Grund (dem Status des Autors) zitieren sie diese Artikel später auch und

produzieren die Korrelation en passant – ohne, dass Qualität oder Relevanz die unabhängige Variable hinter den abhängigen Effekten wäre.

Überdies stieß Eysenbach (2011) in der bereits erwähnten Studie auf eine interessante Selektivität: Die besagte Korrelation zwischen Twitterwerten und Zitationen fand sich durchgehend – ausgenommen waren nur Artikel, die sich mit der gesundheitlichen Situation älterer, einkommensschwacher Menschen befassten. Eysenbachs Erklärung: Die Untersuchungsgruppe war offensichtlich selbst nicht Social-Media-affin genug, um hohe Twitter-Werte generieren zu können. Als ebenfalls nachteilig muss die meist geringe und fachlich ungleiche Abdeckung wissenschaftlicher Literatur in den einzelnen Quellsystemen der Altmetrics-Dienste erwähnt werden (Herb 2015a, S. 214 ff.). Zudem fehlt es an Standardisierung zwischen den Diensten: Sowohl hinsichtlich der berücksichtigten Quellsysteme, als auch hinsichtlich der Auswertung deren Daten.¹⁵ Ebenso ist unklar, welche Maßnahmen die Dienste ergreifen um Manipulationen der Altmetric-Scores zu verhindern (eine Ausnahme bildet der Verlag PLOS, s. Lin 2012). Die fehlende Standardisierung kann aber zumindest teils dem Umstand geschuldet sein, dass bei kommerziellen Anbietern (Altmetric, PlumAnalytics) eine Vereinheitlichung der Methoden Alleinstellungsmerkmale und Geschäftsmodelle in Frage stellen würde.

Die fehlende methodische Fundierung der Altmetrics ist evident, für sie gilt, was Gerhard Fröhlich (1999, S. 28) für die Szientometrie als Ganzem konstatiert: Es „handelt sich bei der Szientometrie um eine eigentümlich theorieleose, offensichtlich von der Illusion des Induktivismus befallene Disziplin (d.h. es dominiert die wissenschaftstheoretische Grundüberzeugung, man könne oder solle auf Theorien verzichten und durch reines Sammeln von Daten zu wissenschaftlichen Erkenntnissen gelangen).“

Potentiale der Altmetrics

Trotz dieser, nicht zu leugnenden methodischen und theoretischen Mängel haben Altmetrics auf einer pragmatischen Ebene, auf der allein zumeist ja auch die Zitationsmetriken betrachtet werden, Vorteile im Vergleich zu den zitationsbasierten Verfahren der Impact-Messung:

Hinsichtlich der berücksichtigten Items sind die Altmetrics offener als die Zitationsdatenbanken. Während Google Scholar noch ein breites Spektrum an Textpublikationen berücksichtigt, wertet Scopus nahezu ausschließlich Journalartikel aus und auch der JIF berücksichtigt nur Zeitschriftenartikel. Jedoch manifestieren sich wissenschaftliche Arbeit und Beiträge zum wissenschaftlichen Fortschritt auch außerhalb kanonisierter Textpublikationen: z.B. bei der Bereitstellung von Forschungsdaten oder der Veröffentlichung von Forschungssoftware. Die Reso-

¹⁵ Als Beispiel sei nur die Unterscheidung der Tweets von Wissenschaftlern und Laien bei Altmetric genannt, die sich sonst nicht findet.

nanz oder Wirkung dieser Objekte erfassen Zitationen aber in aller Regel nicht¹⁶, wohl aber z.B. der beschriebene Altmetrics-Dienst Impactstory. Folglich können Altmetrics einen wesentlichen facettenreicheren und vermutlich realitätsadäquateren Eindruck vom Wert wissenschaftlicher Arbeit vermitteln als es bei der Journalfixierung der meisten Zitationsdatenbanken möglich ist.¹⁷ Zudem gelingt es den Altmetrics zumindest ansatzweise Einfluss, Wirkung oder Rezeption, die sich nicht als Zitationen manifestieren, zu erfassen, z.B. bei Nutzung durch sogenannte Read-Onlys, nicht-publizierende Wissenschaftler. Von dieser Eigenschaft profitieren können auch die Verfasser von Methodenartikeln, die vielfach genutzt aber kaum zitiert werden.

Altmetrics zwischen Dienstleistung, Evaluationswerkzeug, wissenschaftlicher Selbstrechtfertigung und akademischem Quantified Self

Altmetrics als Konzept und als manifeste Dienste oszillieren zwischen den Polen der Evaluierung (wie unter anderem an der Ausrichtung des Angebotes Plum Analytics deutlich wird) und der akademischen Selbstrechtfertigung, so gelingt es Wissenschaftlern mittels der Funktionen von Impactstory (geradezu konträr zur Anlage von Plum Analytics) zu beweisen, dass ihre Arbeit mehr Objekte als Journalartikel hervorbringt und mehr Resonanz entfacht als sich in Zitationen fassen ließe, sprich, dass sie mehrdimensional ist und sich eine Quantifizierung ihrer verbieten sollte. Konsequenterweise nimmt Impactstory, anders als etwa Altmetric, auch keine Reduktion der Werte auf einen Score vor. Diese Selbstrechtfertigung paart sich zuweilen mit den etwas narzisstischen Motiven der auch als Quantified Self bekannt gewordenen Selbstvermessung und steht zugleich im Gegensatz zur Evaluierung und Quantifizierung von Wissenschaft. Der Soziologe Marc Carrigan (2013) beschreibt diese Attitude prägnant: "In my own case, it often confuses me that my hostility towards audit culture can co-exist with a strange affection for google scholar. I still find it exciting to discover that my work has been cited. It confirms people have read it and that it made a sufficient impact for them to choose to reference it. Google Scholar automatically indexes these citations for me, thereby avoiding what would feel like a depressingly narcissistic exercise of searching for them myself." Noch stärker gegen die Reduzierung von Wissenschaft auf Metriken positioniert sich der Biologe Carl Boettiger (2013) mit

16 Abgesehen von Thomson Scientific's Data Citation Index (DCI), http://wokinfo.com/products_tools/multidisciplinary/dci/ [Zugriff am 03.02.2016], der an deutschsprachigen Universitäten aber eher selten lizenziert ist. Der DCI ignoriert jedoch auch Software und andere wissenschaftliche Objekte.

17 Aus einer methodischen Perspektive könnte es vielleicht auch sinnvoll sein, die Bedeutung von Altmetrics-Werten zu entschlüsseln, in dem man sich nicht an Korrelationen zu Zitationsdaten für Textobjekte abmüht, sondern in dem man Designs entwickelt, um zu erfahren, was genau den Unterschied zwischen hohen und niedrigen Altmetrics-Werten für andere Objekte (Daten, Software, etc.) ausmacht. Man könnte gar vermuten, dass Werte für unterschiedliche Objekttypen sich auf unterschiedliche Eigenschaftsdimensionen beziehen.

einem Posting in der Open Science Mailinglist: „Personally, I would prefer the self-assessment over assessment by committee for both practical and ideological reasons. (...) Also, I don't like the concept of a „score“ on several levels: (a) the number would be meaningless. You cannot add ranks across different dimensions (papers vs lab notes). A highly multidimensional vector space does not deserve to be reduced to a single scalar quantity. Yes this happens all the time and we should not encourage it. (...) (b) I don't like the competitive element. I share my own lab notebook in order to do better science, not in order to be 'more open' than the next guy.“

Boettigers Statement stellt gewissermaßen die Gretchenfrage für Altmetrics: Entwickeln sie sich zu Werkzeugen der umfassenden sowie durch Vergleiche Konkurrenzdruck erzeugenden Wissenschaftsevaluierung und der Quantifizierung des bislang und auch durch Altmetrics völlig undefinierten Konstrukts Impact? Oder taugen sie als Hilfsmittel für Wissenschaftler, die zu Zwecken wissenschaftlicher Selbstreflexion mehr über die Verbreitung ihrer Forschung und den durch sie produzierten Fortschritt erfahren wollen und sich zugleich von der eindimensionalen Wahrnehmung ihrer Arbeit durch das stupide Zählen von Zitaten auf Textpublikationen emanzipieren wollen?

Die Marktgängigkeit der Altmetrics wurde zumindest von prominenten und in der Vermarktung wissenschaftlicher Arbeit erfahrenen Akteuren erkannt: Der eindeutig auf Evaluierung mittels Altmetrics ausgerichtete Anbieter Plum wurde 2014 vom Informationsdienstleister EBSCO aufgekauft (Enis 2014) und auch der ebenfalls kommerzielle Dienst Altmetric ist im Besitz eines Branchenriesen: Altmetric ist ein Produkt des Anbieters Digital Science (dazu auch Piwo-war 2013), der wiederum zu MacMillan Publishers Ltd. gehört, in dessen Portfolio sich auch die Nature Publishing Group findet und dessen Besitzer die Holtzbrinck Gruppe ist.

Altmetrics auf dem Prüfstand der Informationswissenschaft

Hermann Hesses Bonmot „jedem Anfang wohnt ein Zauber inne“ trifft auch auf die Altmetrics zu: Ihr Konzept versprach eine Resonanzmessung wissenschaftlicher Leistungen, die über das Zählen von Zitationen auf Textpublikationen hinausgeht und so andere Formen der Resonanz zu erfassen vermag – ein Umstand, der gerade für Sozial- und Geisteswissenschaften, in denen Zitationen weniger stark als in Naturwissenschaften und Medizin Ausdruck wissenschaftlicher Wertschätzung sind, bedeutsam sein dürfte. Doch bei allem Idealismus, der Altmetrics als Teil einer digitalen Revolution in der Wissenschaft (Dickel und Franzen 2015) sehen oder ihnen demokratisierendes Potential (Franzen 2015) attestieren will, – welche ihrer Versprechungen können die alternativen Metriken wirklich halten?

Methodik und Technik¹⁸

Die Verheißung Impact umfassend zu messen, erfüllen Altmetrics schon rein technisch-methodisch nicht. Alle Dienste versuchen in der Regel Zitationen oder Referenzen auf wissenschaftliche Objekte durch maschinelles Aufspüren (Parsen) des Digital Object Identifiers (DOI)¹⁹ der Objekte zu erfassen. Verfügt ein Objekt über keinen DOI wird die Referenz auf es nicht erfasst, das gleiche gilt für den Fall, dass ein Objekt zwar einen DOI besitzt, diese aber bei der Referenzierung nicht erwähnt wird. Die Problematik lässt sich beispielhaft an der Altmetrics-Datenquelle Mendeley (einem Literaturverwaltungssystem) darstellen. Altmetrics-Dienste wie Plum, Altmetric und Impactstory prüfen Mendeley dahingehend,²⁰ wie viele Mendeley-Nutzer ein Objekt in ihrer Bibliothek vorhalten und nutzen dazu das Datenbankfeld des DOI. Verfügt das Objekt über keinen DOI zeigen Plum, Altmetric und Impactstory an, dass kein einziger Nutzer dieses in seiner Bibliothek verwaltet – obwohl hunderte sich das Objekt in der Bibliothek abgespeichert haben können. Analog zeigen die Dienste an kein Nutzer habe das Objekt in seiner Bibliothek, selbst wenn hunderte es darin vorhalten, allerdings keinen oder eine falschen DOI eingetragen haben. Folglich dürften Wissenschaften, in denen die Vergabe von DOI für Texte und/oder deren Verwendung in Referenzen ungebräuchlich sind, bei Impact-Messung mit Altmetrics benachteiligt sein. Beides dürfte vor allem Publikationen aus kleineren Verlagen betreffen, wie man sie häufig in Sozial- und Geisteswissenschaften findet, sowie Autoren aus diesen Wissenschaften. Eine der unbestritten wichtigsten Zeitschriften der deutschsprachigen Soziologie, die Zeitschrift für Soziologie ZfS, nutzt z.B. keinen DOI für Artikel. Für solche Artikel können Altmetrics-Dienste, die Resonanz über DOI-Parsing erfassen wollen, niemals Impact-Werte ermitteln, egal wie wertvoll ein Werk ist. Die DOI-Fokussierung hat zudem fatale Folgen für Objekte die keinen DOI kennen, weil sie nicht elektronisch vorliegen.²¹ Diesbezüglich sind Altmetrics noch angreifbarer als z.B. Google Scholar oder die erwähnten Zitationsdatenbanken, die zumindest prinzipiell Referenzen auf Objekte ohne DOI oder Referenzen ohne Erwähnungen der Objekt-DOI erfassen.

Ebenfalls beachtlich ist der Mangel an Standardisierung, denn die Altmetrics-Dienste nutzen unterschiedliche Datenquellen, teils, z.B. bei Altmetric und Plum, sind diese nicht einmal

18 Ein Altmetrics-Selbstversuch (Herb 2015b) basierend auf eigenen Publikationen förderte noch weitere Merkwürdigkeiten zu Tage, die hier aus Platzgründen nicht berichtet werden.

19 Der DOI ist ein persistenter, also dauerhafter digitaler Verweis auf ein analoges oder digitales Objekt. Für Online-Dokumente funktioniert er mehr oder minder als Weiterleitung auf eine URL. Während eine URL sich ändern kann, bleibt der DOI gleich und verweist bei Online-Dokumente immer auf die aktuelle URL. Anders als die URL ist der DOI damit prinzipiell auch zitierfähig.

20 PLOS ALM fehlt an dieser Stelle, das PLOS nur Werte für eigene Texte, die auch alle über eine DOI verfügen, ermittelt.

21 Wie in Fußnote 18 erwähnt ist der DOI ein Verweis auf elektronische oder analoge Objekte, bei wissenschaftlichen Publikationen wird er aber nahezu ausschließlich für Online-Dokumente vergeben.

einsehbar, und kommen mitunter zu abweichenden Werten (z.B. das Vorkommen von Items in Mendeley-Bibliotheken), auch wenn sie die selbe Quelle (Mendeley) nutzen (s. wiederum Herb 2015b). Weiterhin werden zuweilen verschiedene Altmetrics-Datenquellen für denselben Parameter genutzt: In der Regel wird beim Dienst Impactstory die Anzahl der Erwähnungen eines Textes in Twitter (die Tweets) offensichtlich über den Dienst Topsy bezogen, teils aber auch vom Konkurrenten Altmetric. Altmetric versucht Tweets von Autoren auf eigene Texte in der Zählung zu eliminieren, da Topsy dies unterlässt, sind die Twitter-Daten, die Impactstory für Texte anbietet, nicht vergleichbar.

Transparenz & Open Metrics

Wie erwähnt, dokumentierten Rossner, Van Epps & Hill (2008; 2007) die fehlende Reproduzierbarkeit von Zitationsdaten der Journal Citation Reports. Ursache der Reproduzierbarkeit ist die fehlende Verfügbarkeit der Rohdaten, die in Händen proprietärer Anbieter sind. Zwei der Altmetrics-Anbieter (PLOS und Impactstory), erlauben zumindest den Download des Teils der Daten, die sie nicht von Drittanbietern beziehen.²² Da PLOS allerdings nur Altmetrics-Informationen für verlagseigene Publikationen ausgibt und daher keine echter Altmetrics-Dienstleister ist, bleibt nur Impactstory als Dienst, der *teilweise* transparente Daten anbietet. Die kommerziellen Anbieter Plum und Altmetric erlauben dies nicht. Spätestens hier wird klar, dass alternative Metriken nicht per se offene Metriken (Herb 2015a, S. 221 ff.) sind, denn den Kriterien-Katalog für offene Metriken erfüllt keiner der Dienste (Herb 2015a, S. 221): „Offene Metriken (...) müssten sich (...) nicht nur durch die Ausweitung der Impact-Erfassung über Zitationen hinaus von den bislang etablierten Verfahren unterscheiden. Open Metrics müssten auch die Daten, die der Berechnung der Impact-Scores zugrunde liegen, offen bereitstellen. Zudem sollten die ermittelten Werte im Idealfall auch automatisiert abfragbar sein, z. B. über ein Application Programming Interface API. Unter Gesichtspunkten der Transparenz sollte nachvollziehbar dokumentiert sein, wie die ausgewählten Datenquellen bestimmt wurden und mittels welcher Formel oder Parameter die Werte berechnet wurden, z. B. bei Entscheidungen hinsichtlich des Einschlusses von Datenquellen oder methodischen Fragen wie der Gewichtung unterschiedlicher Datentypen bzw. der Informationen aus unterschiedlichen Datenquellen. Offenheit im Sinne der Open Metrics bezieht sich folglich auf möglichst offene Zugänglichkeit und Transparenz bei der Berechnung von Metriken.“

Zudem sind wie bereits erwähnt die ausgewerteten Datenquellen bei zumindest zwei der Dienste (Plum, Altmetric) nicht einsehbar. Impactstory macht diese Liste öffentlich, es ist aber, wie ebenfalls schon dargestellt, etwas unklar, woher man die Daten bezieht. Ebenso intransparent bleibt z.B. nach welchen Parametern Altmetric aus der Fülle der Daten den Score jeder

²² Diese sind mitunter sogar unter offenen Daten-Lizenzen, die mit Open-Source-Lizenzen kompatibel sind, verfügbar.

Publikation berechnet – zwar wurden grobe Informationen dazu publiziert, jedoch fehlen die detaillierte Angaben zur Berechnung des Scores (s. Altmetric 2015, wo jede Information zu den so bezeichnenden Score Modifiers fehlt). Nicht nachvollziehbar ist weiterhin wie der Dienst z.B. Wissenschaftler in Twitter identifiziert, um die versprochene Höhergewichtung deren Tweets herzustellen oder mit welchem Faktor man diese Gewichtung vornimmt.

Wissenschaftliche Fundierung, Steuerungscharakter und ein Fazit

Genau wie zitationsbasierten Impact-Maße fehlt auch den Altmetrics offensichtlich jede theoretische und wissenschaftliche Fundierung, es ist völlig unklar, welche Eigenschaft eines Objekts sie mit welchen Operationalisierungen messen. Die Desiderate für Altmetrics und Zitationsmessung sind gleich, leiten sie sich doch aus der wissenschaftlichen Leichtgewichtigkeit der Impact-Messung an sich ab, es fehlt ihnen schlichtweg an wissenschaftstheoretischer Aufklärung und sozialwissenschaftlicher Läuterung (Fröhlich 1999, S 36).

Neben den methodischen und wissenschaftlichen Mängeln werden Altmetrics praktisch auch sehr stark als Werkzeuge zur Bewertung und Fremdsteuerung von Wissenschaft beworben (man bedenke den dirigistischen Charakter der Plum-Angebote oder den kompetitiven Geist der Plum- sowie Altmetric-Beschreibungen), auch diesbezüglich unterscheiden sie sich nicht von der Zitationsmessung.

Trotz zweier verbleibender Vorteile, der Ausweitung²³ der relevanten Objekte von den in der Zitationsmessung meist alleine berücksichtigten Journalartikeln hin zu Texten, Daten, Software und Präsentationen sowie der Erfassung der Nutzung und Erwähnung von Objekten (und nicht nur ihrer Zitationen), bleibt die Bilanz also mager: Sowohl hinsichtlich Wissenschaftlichkeit und Methodik als auch Transparenz heben sie sich weder in der Theorie noch in der Praxis von den bekannten zitationsbasierten Verfahren ab.

Vor allem aber können sie die ihnen zugesprochenen revolutionären und demokratisierenden Versprechungen nicht einlösen: Es ist nicht zu erkennen, wie Altmetrics die Strukturen und Kräfteverhältnisse im wissenschaftlichen Feld positiv verändern werden. Im Gegenteil: Faktisch (und auch wenn dies nicht in der Absicht der Altmetrics-Pioniere Piwowar und Priem lag) taugen sie als neuartige Evaluierungsinstrumente und werden auch nachdrücklich mit dieser Funktion beworben. Insofern sind sie geeignet die Tendenzen der Fremdsteuerung von Wissenschaft zu verstärken und die Autonomie des wissenschaftlichen Feldes zu schwächen.

23 Leider müsste man noch weiter einschränken: Nur ein Dienst, Impactstory, berücksichtigt, Texte, Daten Software und Präsentationen – und dies gelingt ihm nur sehr fehlerhaft (Herb 2015b).

Literatur

- Altmetric. 2015. "How Is the Altmetric Score Calculated?" *Altmetric Website*. <https://help.altmetric.com/support/solutions/articles/6000060969-how-is-the-altmetric-score-calculated-> [Zugriff am 03.02.2016].
- Apic, Gordana, Matthew J Betts, and Robert B Russell. 2011. "Content Disputes in Wikipedia Reflect Geopolitical Instability." *PLoS One* 6 (6): e20902. doi:10.1371/journal.pone.0020902.
- Archambault, Éric, Étienne Vignola-Gagné, Grégoire Côté, Vincent Larivière, and Yves Gingras. 2006. "Benchmarking Scientific Output in the Social Sciences and Humanities: The Limits of Existing Databases." *Scientometrics* 68 (3): 329–42. doi:10.1007/s11192-006-0115-z.
- Bar-Ilan, Judit. 2012. "JASIST@mendeley." In *ACM Web Science Conference 2012 Workshop*. Evanston, Illinois. <http://altmetrics.org/altmetrics12/bar-ilan/> [Zugriff am 03.02.2016].
- Boettiger, Carl. 2013. "Open Scholar Foundation." *Open Science Mailing Liste*. <https://lists.okfn.org/pi-permail/open-science/20131206/002950.html> [Zugriff am 03.02.2016].
- Bollen, Johan, Huina Mao, and Xiaojun Zeng. 2011. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2 (1): 1–8. doi:10.1016/j.jocs.2010.12.007.
- Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis Bettencourt, Ryan Chute, Marko A. Rodriguez, and Lyudmila Balakireva. 2009. "Clickstream Data Yields High-Resolution Maps of Science." *PLoS One* 4 (3): e4803. doi:10.1371/journal.pone.0004803.
- Bollen, Johan, Herbert Van De Sompel, Aric Hagberg, and Ryan Chute. 2009. "A Principal Component Analysis of 39 Scientific Impact Measures." *PLoS One* 4 (6): e6022. doi:10.1371/journal.pone.0006022.
- Bollen, Johan, Herbert Van De Sompel, and Marko A. Rodriguez. 2008. "Towards Usage-Based Impact Metrics." *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries - JCDL '08*. New York, New York, USA: ACM Press, 231–40. doi:10.1145/1378889.1378928.
- Bollen, Johan, Herbert Van De Sompel, Joan A. Smith, and Rick Luce. 2005. "Toward Alternative Metrics of Journal Impact: A Comparison of Download and Citation Data." *Information Processing & Management* 41 (6): 1419–40. doi:10.1016/j.ipm.2005.03.024.
- Brembs, Björn, Katherine Button, and Marcus Munafò. 2013. "Deep Impact: Unintended Consequences of Journal Rank." *Digital Libraries; Physics and Society. Frontiers in Human Neuroscience* 7 (January). doi:10.3389/fnhum.2013.00291.
- Brody, Tim, Stevan Harnad, and Leslie Carr. 2005. "Earlier Web Usage Statistics as Predictors of Later Citation Impact." *Journal of the American Association for Information Science and Technology* 57 (8): 1060–72. doi:10.1002/asi.20373.
- Brown, Tracey. 2004. *Peer Review and the Acceptance of New Scientific Ideas*. London. <http://www.senseaboutscience.org/data/files/resources/17/peerReview.pdf> [Zugriff am 03.02.2016].
- Carrigan, Marc. 2013. "The Academic Quantified Self." *The Sociological Imagination*. <http://sociological-imagination.org/archives/14454> [Zugriff am 03.02.2016].

- Clermont, Marcel, and Harald Dyckhoff. 2012. "Coverage of Business Administration Literature in Google Scholar: Analysis and Comparison with EconBiz, Scopus and Web of Science." *Bibliometrie – Praxis Und Forschung* 1 (5). <http://www.bibliometrie-pf.de/article/view/165> [Zugriff am 03.02.2016].
- Davis, Philip M. 2012. "Tweets , and Our Obsession with Alt Metrics." *The Scholarly Kitchen*. <http://scholarlykitchen.sspnet.org/2012/01/04/tweets-and-our-obsession-with-alt-metrics/> [Zugriff am 03.02.2016].
- Dickel, Sascha, and Martina Franzen. 2015. "Digitale Inklusion : Zur Sozialen Öffnung Des Wissenschaftssystems." *Zeitschrift Für Soziologie* 44 (5): 330–47.
- Dong, Peng, Marie Loh, and Adrian Mondry. 2005. "The 'Impact Factor' Revisited." *Biomedical Digital Libraries* 2 (7). doi:10.1186/1742-5581-2-7.
- Enis, Matt. 2014. "EBSCO Acquires Altmetrics Provider Plum Analytics." *The Digital Shift*. <http://www.thedigitalshift.com/2014/01/publishing/ebSCO-acquires-altmetrics-provider-plum-analytics/> [Zugriff am 03.02.2016].
- Eysenbach, Gunther. 2011. "Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact." *Journal of Medical Internet Research* 13 (4). doi:10.2196/jmir.2012.
- Franzen, Martina. 2015. "Der Impact Faktor War Gestern." *Soziale Welt* 66 (2): 225–42.
- Fröhlich, Gerhard. 1999. "Das Messen Des Leicht Meßbaren: Output-Indikatoren, Impact-Maße: Artefakte Der Szientometrie?" In *Kommunikation Statt Markt : Zu Einer Alternativen Theorie Der Informationsgesellschaft*, edited by Jörg Becker and Wolf Göhring, 27–38. GMD- Forschungszentrum Informationstechnik GmbH. <http://eprints.rclis.org/9115/> [Zugriff am 03.02.2016].
- Gläser, Jochen, and Grit Laudel. 2001. "Integrating Scientometric Indicators into Sociological Studies: Methodical and Methodological Problems." *Scientometrics* 52 (3): 411–34. doi:10.1023/A:1014243832084.
- Herb, Ulrich. 2011. "Wissensorganisation à La Open Source. Wie Die Open Knowledge Foundation Freies Wissen Fördert." *C'ï Magazin Für Computer Und Technik*. http://www.heise.de/artikel-archiv/ct/2011/19/142_kiosk [Zugriff am 03.02.2016].
- Herb, Ulrich. 2015a. *Open Science in Der Soziologie: Eine Interdisziplinäre Bestandsaufnahme Zur Offenen Wissenschaft Und Eine Untersuchung Ihrer Verbreitung in Der Soziologie*. Glückstadt, Germany: Verlag Werner Hülsbusch. doi:10.5281/zenodo.31234.
- Herb, Ulrich. 2015b. "Impact-Lotterie Mit Impactstory." *Scinoptica Blog*. <http://www.scinoptica.com/pages/topics/impactstory-eine-magere-bilanz.php> [Zugriff am 03.02.2016].
- Hirsch, Jorge E. 2005. "An Index to Quantify an Individual's Scientific Research Output." *Proceedings of the National Academy of Sciences of the United States of America* 102 (46): 16569–72. doi:10.1073/pnas.0507655102.
- Li, Xuemei, and Mike Thelwall. 2012. "F1000, Mendeley and Traditional Bibliometric Indicators." In *17th International Conference on Science and Technology Indicators*, 3:1–11. http://2012.sticonference.org/Proceedings/vol2/Li_F1000_541.pdf [Zugriff am 03.02.2016].

- Li, Xuemei, Mike Thelwall, and Dean Giustini. 2012. "Validating Online Reference Managers for Scholarly Impact Measurement." *Scientometrics* 91 (2): 461–71. doi:10.1007/s11192-011-0580-x.
- Lin, Jennifer. 2012. "A Case Study in Anti-Gaming Mechanisms for Altmetrics: PLoS ALMs and DataTrust." In *altmetrics12 - ACM Web Science Conference 2012, 21. Juni 2012*. Evanston, Illinois. <http://altmetrics.org/altmetrics12/lin/> [Zugriff am 03.02.2016].
- McKerahan, Tiffany L, and Stephen W Carmichael. 2012. "What Is the Impact Factor, Anyway?" *Clinical Anatomy* 25 (3): 283. doi:10.1002/ca.21291.
- McVeigh, Marie E, and Stephen J Mann. 2009. "The Journal Impact Factor Denominator: Defining Citable (counted) Items." *JAMA : The Journal of the American Medical Association* 302 (10): 1107–9. doi:10.1001/jama.2009.1301.
- Mestyán, Márton, Taha Yasseri, and János Kertész. 2013. "Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data." *PLoS One* 8 (8): e71226. doi:10.1371/journal.pone.0071226.
- Piwovar, Heather A. 2013. "Altmetrics: Value All Research Products." *Nature* 493 (7431): 159. doi:10.1038/493159a.
- Plum Analytics. 2016a. "PlumX Dashboards." *Plum Website*. <http://plumanalytics.com/products/plumx-dashboards/> [Zugriff am 03.02.2016].
- . 2016b. "PlumX +Grants." *Plum Website*. <http://plumanalytics.com/products/plumx-grants/> [Zugriff am 03.02.2016].
- Rosner, Mike, Heather Van Epps, and Emma Hill. 2007. "Show Me the Data." *The Journal of Cell Biology* 179 (6): 1091–92. doi:10.1083/jcb.200711140.
- . 2008. "Irreproducible Results: A Response to Thomson Scientific." *The Journal of Experimental Medicine* 205 (2): 260–61. doi:10.1084/jem.20080053.
- Seglen, Per O. 1998. "Citation Rates and Journal Impact Factors Are Not Suitable for Evaluation of Research." *Acta Orthopaedica* 69 (3): 224–29. doi:10.3109/17453679809000920.
- Shuai, Xin, Alberto Pepe, and Johan Bollen. 2012. "How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations." *Digital Libraries; Physics and Society*. Edited by Christos A. Ouzounis. *PLoS ONE* 7 (11): e47523. doi:10.1371/journal.pone.0047523.
- Smeyers, Paul, and Nicholas C. Burbules. 2011. "How to Improve Your Impact Factor: Questioning the Quantification of Academic Quality." *Journal of Philosophy of Education* 45 (1): 1–17. doi:10.1111/j.1467-9752.2011.00787.x.
- The 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 [Zugriff am 03.02.2016].
- "The San Francisco Declaration on Research Assessment DORA." 2012. <http://www.ascb.org/files/SFDeclarationFINAL.pdf> [Zugriff am 03.02.2016].

- Van Leeuwen, Thed. 2013. "Bibliometric Research Evaluations , Web of Science and the Social Sciences and Humanities : A Problematic Relationship?" *Bibliometrie - Praxis Und Forschung 2*. <http://www.bibliometrie-pf.de/article/view/173> [Zugriff am 03.02.2016].
- Ware, Mark, and Michael Mabe. 2009. *The STM Report. An Overview of Scientific and Scholarly Journal Publishing*. Oxford. http://www.stm-assoc.org/2009_10_13_MWC_STM_Report.pdf [Zugriff am 03.02.2016].

Ulrich Herb, Soziologie-Studium an der Universität des Saarlandes (Abschluss Diplom-Soziologe), Promotion (ebenfalls an der Universität des Saarlandes) in Informationswissenschaft. Seit 2001 an der Saarländischen Universitäts- und Landesbibliothek tätig und für die Betreuung von Drittmittelprojekten und elektronischen Publikationsangeboten zuständig, freiberuflich als Wissenschaftsberater und Journalist tätig. Lehrbeauftragter an verschiedenen Hochschulen, etwa an der Hochschule der Medien in Stuttgart oder der HTW Chur.

E-Mail: u.herb@sulb.uni-saarland.de