

VOM MESSEN UND GEMESSEN WERDEN:

Potentiale & Grenzen bibliometrischer Methoden

WIEN
SEPTEMBER 2020

ÖSTERREICHISCHER
WISSENSCHAFTSRAT



**VOM MESSEN UND GEMESEN WERDEN:
Potentiale & Grenzen bibliometrischer Methoden**

Wien, im September 2020

Inhalt

1	Genese und Intention	5
2	Evaluation und Steuerung von Forschung	9
3	Grundlagen der Bibliometrie.....	21
3.1	Bibliometrische Datenquellen.....	21
3.2	Bibliometrische Impact-Indikatoren	23
3.3	Anforderungen und Grenzen bibliometrischer Methoden.....	36
4	Rolle bibliometrischer Methoden	41
4.1	Umfrage zur Bibliometrie an österreichischen Universitäten (2010)	41
4.2	Umfrage zur Bibliometrie an österreichischen Universitäten (2019)	42
4.3	Ausgewählte internationale Beispiele.....	51
5	Möglichkeiten bibliometrischer Analysen wissenschaftlicher Disziplinen	57
5.1	Klinische Forschung	60
5.2	Bildungsforschung.....	61
5.3	Informatik	68
6	Erkenntnisse und Empfehlungen	85
7	Literatur	91
8	Anhänge.....	97

1 Genese und Intention

„Die Wissenschaft ist es gewohnt zu messen. Sie ist es ebenso gewohnt, selbst gemessen zu werden, Rechenschaft über ihr Tun abzulegen. Diese ‚Vermessung‘ kann Wege der Forschung und ihre Erträge sichtbar machen und Anliegen einer wissenschafts- und forschungspolitischen Steuerung dienen; sie kann Basis zukünftiger Schwerpunktsetzungen, Anreiz zur Profilbildung einer forschenden Institution als Bildungsanbieter sein. Sie kann der Selbsteinschätzung von Fächern und Disziplinen und einer wissenschaftlichen Institution – wie gut sind wir? – dienen und Ansporn zur wissenschaftlichen Weiterentwicklung sein.“¹ Mit diesen Worten umriss der Wissenschaftsrat im Jahr 2014 seine ersten Überlegungen zu Mess- und Beurteilungsverfahren von Forschungsleistung.

Im Jahr darauf erinnerten uns die einleitenden Worte zu *The Metric Tide* an die stets mitschwingende Ambivalenz, wenn es um den Einsatz metrischer Methoden zur Leistungsevaluation geht. Der Überzeugung, mittels der Objektivität von Daten einen schnellen und unvoreingenommenen Blick auf Forschungsleistungen zu erhalten, steht die Skepsis gegenüber, ob eine solche Objektivität aufgrund historisch gewachsener Strukturen unterschiedlicher Disziplinen überhaupt möglich sei und noch mehr der unreflektierte Einsatz diverser Indices letztlich nicht mehr Schaden als Nutzen bringe.² Hierbei wirken tradierte Publikationsformen und -strategien, wissenschaftssprachliche Traditionen und rezeptive Bezugsräume, ob und in welcher Intensität metrische Methoden als taugliches Mittel der Evaluation erachtet werden.

Beide Aussagen gemahnen uns, dass Verfahren zur Steuerung und Evaluation von Forschung dem Wissenschaftssystem nicht nur immanent sind, sondern einer steten Neubewertung und Weiterentwicklung bedürfen. Dem Wissenschaftsrat ist es qua seiner Existenz Aufgabe, die Entwicklungen auf den Gebieten der Steuerung und Evaluation von Wissenschaft zu beobachten, einzuordnen und darauf beruhend Empfehlungen abzugeben.

Demgemäß setzte sich dieser, wie oben zitiert, unter anderem mit der Messung und Beurteilung von Qualität in der Forschung auseinander; im Rahmen der gemeinsam mit dem Deutschen

¹ Österreichischer Wissenschaftsrat, *Die Vermessung der Wissenschaft: Messung und Beurteilung von Qualität in der Forschung* (Wien, 2014), S. 3.

² Wilsdon, James u. a., *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management* (2015), S. iii.

Wissenschaftsrat ausgerichteten Tagung „Qualitätsstandard oder leeres Ritual? Begutachtungen in der Diskussion“³ wiederum rückten Gegenwart und Zukunft von Peer Review-Verfahren ins Zentrum des Interesses.

Im Laufe der vergangenen Jahre widmete sich der Wissenschaftsrat nicht nur metrik-basierten Methoden selbst, sondern brachte diese im Rahmen seiner Arbeiten in unterschiedlicher Intensität auch zur Anwendung. Im Zuge der Studie zur Klinischen Forschung in Österreich⁴ war der bibliometrischen Analyse noch eine ergänzende Rolle zugeschrieben; diese sollte seinerzeit lediglich einen stichprobenartigen Überblick über den Status quo des Forschungsgegenstandes verschaffen.⁵ In der gemeinsam mit dem Wissenschaftsfonds – FWF durchgeföhrten Studie „Standortbestimmung der Bildungsforschung in Österreich“⁶ nahm sie eine gewichtigere Rolle ein und bildete eine der Grundlagen für ein Informed Peer Review-Verfahren. Im Rahmen der Analyse der Informatik in Österreich kam der Bibliometrie zentrale Bedeutung bei.⁷ Hierbei wurde eine ausführliche Analyse in Auftrag gegeben, um auf Grundlage der Zwischenergebnisse im Rahmen einer Veranstaltung⁸ den Status quo der Informatik und Perspektiven der Weiterentwicklung zu diskutieren. Die Veranstaltung wiederum diente gewissermaßen als Feedbackschleife; in gemeinsamer kritischer Reflexion mit der wissenschaftlichen Community wurden Stärken und Schwächen der ersten Ergebnisse der Analyse eruiert und zur Weiterentwicklung und Verfeinerung der Analyse herangezogen. Ziel der Arbeiten war es aber nicht nur, Erkenntnisse über Stärken und Schwächen der Informatik zu gewinnen, sondern anhand der Analyse der Informatik die Potentiale und Grenzen bibliometrischer Methoden auszuloten. In diesem Sinne werden im Kapitel zur Informatik (Kapitel 5.3) Zwischenergebnisse präsentiert; der finale Bericht zum Status quo der Informatik in Österreich wird derzeit erarbeitet. Dieser Abschnitt

³ Österreichischer Wissenschaftsrat, *Tagungsbericht zur Veranstaltung „Qualitätsstandard oder leeres Ritual? Begutachtungen in der Diskussion“* (Wien, 2018).

⁴ Österreichischer Wissenschaftsrat, *Klinische Forschung in Österreich: Stellungnahme und Empfehlungen* (Wien, 2016a).

⁵ Die tatsächliche bibliometrische Analyse wurde durch CWTS Leiden durchgeführt. Dies war ebenso der Fall bei den beiden darauffolgenden bibliometrischen Analysen.

⁶ Hesse, Friedrich W. u. a., *Determination of the State of Educational Research in Austria* (Wien, 2019).

⁷ Noyons, Ed, *Bibliometric Analysis of the Austrian Computer Science Research* (unveröffentlichter Draft) (Leiden, 2019b); Noyons, Ed, „Bibliometric Analyses of the Austrian Computer Science Research“, Präsentation bei *Informatik in Österreich: Perspektiven und Strategien* (Wien, 2019c).

⁸ Österreichischer Wissenschaftsrat, „Informatik in Österreich: Perspektiven und Strategien“ (Wien, 2019), <<https://www.wissenschaftsrat.ac.at/blog/informatik-in-österreich-perspektiven-und-strategien>> [zugegriffen 3. August 2020].

dient also vornehmlich der Darstellung von Prozess und Methodik bibliometrischer Analysen anhand der Disziplin.

Warum also befasst sich der Wissenschaftsrat erneut mit Fragen der Forschungsevaluation, und warum wird in dieser Publikation das Augenmerk auf metrische Methoden gelegt? Gibt es nicht Publikationen sonder Zahl über den Einsatz der Bibliometrie?

Die vorliegende Publikation hat nicht zum Ziel, das „bibliometrische Rad“ neu zu erfinden, vielmehr sollen die Grundlagen und neuesten Entwicklungen der Methodik dargelegt und bewertet werden. Demgemäß werden die wichtigsten Datenquellen für bibliometrische Analysen vorgestellt, sowie ein Überblick über verschiedene Arten bibliometrischer Indikatoren, deren Anwendungsmöglichkeiten und Grenzen gegeben. Zugleich soll damit die fortlaufende Weiterentwicklung der Methodik als datenbasiertes Instrument, Erkenntnisse über wissenschaftliche Aktivitäten und deren vielfältige Vernetzungen auf unterschiedlichen Aggregationsebenen zu gewinnen, veranschaulicht werden.

Es soll ein Überblick über die Einbettung der Methodik in den institutionellen Alltag österreichischer und internationaler Hochschulen geboten werden, um somit ein besseres Verständnis des gegenwärtigen Stellenwerts der Bibliometrie und ihres möglichen Nutzens zu erhalten.

Anhand der Fallbeispiele lassen sich zudem disziplinspezifische Herausforderungen an metrik-basierte Methoden aufzeigen; seien dies, wie bereits oben erwähnt, tradierte Publikationsformen und -strategien, wissenschaftssprachliche Traditionen oder rezeptive Bezugsräume. Gleichzeitig wird hierdurch deutlich, welche Analysepotentiale (aber auch -grenzen) moderne Verfahren der Bibliometrie über Disziplinengrenzen hinweg, und auch über deren engere Anwendung zu Evaluationszecken hinaus, für die unterschiedlichen Ebenen des Wissenschaftssystems aufweisen. Wichtig ist jedoch, dass disziplinspezifische Besonderheiten bereits im Rahmen der Datenerhebung und der Verwendung feldspezifisch normierter Indikatoren berücksichtigt werden.

Im Sinne der Selbstreflektion soll aber auch nachgezeichnet werden, wie eine Institution, deren Ziel die systemische Beobachtung und Beratung ist, eine bestimmte Methode lernend in ihre Arbeit integriert. Insofern stellt dieser Text in gewisser Weise einen Tätigkeitsbericht des Wissenschaftsrates zur Thematik „datenbasierte Analyse des österreichischen Wissenschaftsraums“ dar.

Vor diesem Hintergrund hat sich der Wissenschaftsrat zum Ziel gesetzt, zum einen eine kritische Reflexion der Bedeutung bibliometrischer Analysen für die Forschungsevaluation und zum anderen eine nicht als abschließend zu betrachtende Einschätzung über die Anwendungsbereiche, Potentiale und Grenzen bibliometrischer Methoden zur Analyse des wissenschaftlichen Outputs

auf unterschiedlichen Ebenen (Individualebene, Institutionelle Ebene und Systemebene) vorzulegen. Gleichermassen jedoch behalt sich der Wissenschaftsrat auch mit Blick auf bereits bestehende Dokumente mit empfehlendem Charakter vor, den Fokus in seinen Empfehlungen auf die Potentiale dieser Methoden zu legen.

2 Evaluation und Steuerung von Forschung

Seit es Forschung als soziale Praxis, mit entsprechenden Normen verantwortungsvollen Vorgehens, verlässlichen Aussagen und angemessenen Formen der Institutionalisierung gibt, geht Forschung immer auch mit Praktiken ihrer Evaluation einher. In der Vergangenheit diente Forschungsevaluierung vor allem der Wissenschaft selbst. Manuskripte wurden durch Experten, *Peers*, geprüft und kamen einst, wie mehrheitlich auch heute noch, erst nach eingehenden Begutachtungsverfahren zur Veröffentlichung. Im Laufe der Zeit entstand ein wachsendes Interesse an der Produktivität von Wissenschaftlerinnen und Wissenschaftlern und dem Impact ihrer Publikationen in der wissenschaftlichen Community. So hat der Mathematiker Alfred Lotka bereits 1926 einen quantitativen Vergleich der Produktivität von wissenschaftlichen Autoren veröffentlicht; ein Jahr später folgte die erste Zitationsanalyse, durchgeführt von P. Gross und E. Gross.⁹

Einhergehend mit einem stetigen Wachstum des Systems Wissenschaft insgesamt wird Forschung vermehrt aus einer Vielzahl unterschiedlicher Beweggründe heraus und von unterschiedlichen Interessensgruppen evaluiert.¹⁰ Einen ersten Überblick verschafft die vom Deutschen Wissenschaftsrat erstellte Tabelle über verschiedene Objekte von Begutachtungen und damit verbundenen Zielsetzungen (Tabelle 1).

Gegenstände	Ziele
Manuskripte	
Artikel in Fachzeitschriften (auch andere Publicationsformen)	Auswahl für Veröffentlichung, Fehler- und Plagiatskontrolle, Hinweise für Überarbeitung
Ressourcen	
Drittmittel bzw. Projektmittel zur Forschungsförderung	Wettbewerbliche Vergabe zusätzlicher Ressourcen
Grundmittel für Forschungseinrichtungen	Entscheidung über weitere Förderung von Instituten
Infrastrukturen	Entscheidung über Bau einer Forschungsinfrastruktur
Infrastrukturzugang	Bewilligung von Nutzungs-/Messzeiten
Personen	
Empfehlungsschreiben	Auswahl für Stellen auf allen Karrierestufen, Studiengänge oder summer schools
Qualifikationsarbeiten	Feststellung von Ausbildungserfolg oder Forschungsleistung
Berufungen	(Vergleichende) Auswahl für Rekrutierung
Stipendien	Auswahlentscheidung für Ressourenzuteilung
Preise	Auszeichnung und Zuteilung von Ressourcen

⁹ Havemann, Frank, *Einführung in die Bibliometrie* (Berlin, 2016).

¹⁰ Ebd.

Studiengänge	
Studiengänge	Programmakkreditierungen
Qualitätssicherungssystem einer Hochschule	Systemakkreditierung von Strukturen und Prozessen
Akkreditierungsagenturen	Akkreditierung der Agenturen
Institutionen	
Forschungsinstitut	Thematische Ausrichtung und Schwerpunktsetzung auf Forschungsfeldern
Nichtstaatliche Hochschulen	Einhaltung von Standards für Zulassung und Titelvergabe
Disziplin/Fach	Leistungsvergleich und Entwicklungsbedarf
Strategien	
Förderprogramme	Verbesserung von Förderprogrammen und -strategien
Institutionelles Gefüge	Ausrichtung einzelner Teile des Wissenschaftssystems

Tabelle 1: Gegenstand und Ziele von Begutachtungen im Wissenschaftssystem¹¹

Als Grundlage für wissenschaftspolitische und -strategische Entscheidungen gewinnt die Evaluierung von Forschungsleistung besonders im Hinblick auf die verstärkt wettbewerbsorientierte Vergabe von Fördermitteln, mit der Forschung qualitativ gesteuert werden soll, immer mehr an Bedeutung.¹² Das Konzept der Forschungssteuerung fand seit den 1970er Jahren zunächst vornehmlich im Rahmen der Organisation und Steuerung von Forschung außeruniversitärer Einrichtungen Verwendung und ist mittlerweile fester Bestandteil universitärer Governance-Strukturen. Im Rahmen der Forschungssteuerung müssen z.B. bei der Festlegung von Forschungsschwerpunkten und (somit) der Zuweisung von Ressourcen komplexe und weitreichende Entscheidungen getroffen werden. Dazu gehört auch die Bewertung der Qualität von Forschungsergebnissen auf unterschiedlichen Ebenen, von der Ebene einzelner Projekte bis zu ganzen Einrichtungen auf Basis geeigneter Leistungsindikatoren.¹³ Zur Gestaltung derartiger Evaluations bedienen sich die Wissenschaft, Politik und Förderorganisationen verschiedenster Methoden und Instrumente, die Aufschluss über die erbrachte Leistung und einen Vergleich – im besten Fall über Instituts-, Disziplinen- oder Systemgrenzen hinweg – ermöglichen sollen.

¹¹ Wissenschaftsrat, *Begutachtungen im Wissenschaftssystem: Positionspapier* (Berlin, 2017), S. 12.

¹² Schweizerischer Wissenschafts- und Technologierat, *Leistungsmessung und Qualitätssicherung in der Wissenschaft* (Bern, 2013), S. 7; Taubert, Niels Christian und Peter Weingart, „Wandel des wissenschaftlichen Publizierens - eine Heuristik zur Analyse rezenter Wandlungsprozesse“, in *Wissenschaftliches Publizieren - Zwischen Digitalisierung, Leistungsmessung, Ökonomisierung und medialer Beobachtung*, hg. von Peter Weingart und Niels Taubert (Berlin, 2016), S. 3–38 (S. 14).

¹³ Schweizerischer Wissenschafts- und Technologierat, 2013.

In der Diskussion über die Vor- und Nachteile unterschiedlicher Beurteilungsmethoden wird zwischen qualitativen und quantitativen Methoden unterschieden. Qualitative Evaluationen werden durch Wissenschaftlerinnen und Wissenschaftler aus verwandten Bereichen in einem strukturierten Prozess vorgenommen. Diese Form der qualitativen Begutachtung wird gemeinhin als *Peer Review* bezeichnet. Die quantitative Leistungsmessung bedient sich verschiedener Leistungsindikatoren und bibliometrischer Daten. Die Vor- und Nachteile beider Formen sowie ihre Chancen und Grenzen sind seit mehr als einem Jahrzehnt Gegenstand von Debatten, nicht zuletzt, weil durch die wachsende Bedeutung der Wissenschaft für die wirtschaftliche und gesellschaftliche Wettbewerbsfähigkeit die Notwendigkeit steuernder Eingriffe in die Forschungslandschaft immer häufiger betont wird.

Peer Review

Peer Review ist eines der weithin verbreitetsten Verfahren zur Evaluation von Forschung. Dieses Instrument wird auf unterschiedlichen Ebenen, Disziplinen und Institutionen herangezogen. Hier erfolgt die Bewertung von Forschungsleistung bzw. des Potentials eines Forschungsvorhabens durch Expertinnen und Experten, die möglichst selbst im gleichen Forschungsbereich tätig sind. Die Expertise der Gutachtenden über aktuelle Entwicklungen und Standards der eigenen Disziplin sowie deren Fachwissen sollen eine sachliche Bewertung von Forschung ermöglichen. Auch wenn Peer Review weite Anwendung findet, so ist das Instrument nicht völlig unumstritten. Seit den 1970er Jahren wird dem Peer Review immer wieder vorgeworfen, weder fair noch transparent zu sein.¹⁴ Wie einzelne Gutachterinnen bzw. Gutachter zu ihrem Ergebnis kommen, ist nicht in allen Fällen nachvollziehbar und unterschiedliche Expertinnen und Experten kommen oft zu gegensätzlichen Entscheidungen. Eine sich stetig verschärfende Problematik dieses Evaluationsverfahrens besteht nicht zuletzt darin, dass es zeit- und personalaufwendig ist; die steigende Zahl an Veröffentlichungen und Förderanträgen und der mit Peer Review-Prozessen verbundene Arbeitsaufwand setzt Forschungsförderer und die Editoren von Zeitschriften zunehmend unter Druck.¹⁵

¹⁴ Hirschauer, Stefan, „Peer Review Verfahren auf dem Prüfstand“, *Zeitschrift für Soziologie*, 33 (2002), S. 62–83.

¹⁵ de Vrieze, Jop, „Funders Groan under Growing Review Burden“, *Science*, 357 (2017), S. 343.

Bibliometrische Methoden der Leistungsevaluation und Sichtbarmachung von Forschungsaktivitäten

Bibliometrische Evaluierungsverfahren können manchen Limitierungen des Peer Review-Verfahrens entgegenwirken. Ist ein ausreichend großer und gut aufbereiteter Datenpool gegeben, so ermöglichen bibliometrische Analysen relativ rasch generierbare, statistisch gesicherte Informationen zum Forschungsoutput und Impact. Deren Aussagekraft ist auf der „Makroebene“, also der Ebene ganzer Länder, am höchsten, aber auch auf der „Mesoebene“ (Universitäten, Forschungsinstitute, Forschungsgruppen) ist die Aussagekraft der Bibliometrie hoch. Jedoch treten insbesondere auf der „Mikroebene“ (Personen, einzelne Projekte) häufig fachspezifische Aspekte in den Vordergrund, die rein metrik-basierte Bewertungen erheblich erschweren können.¹⁶ Da publikationsbezogene Daten letztlich implizit auf dem Peer Review-Begutachtungsverfahren der Zeitschriftenverlage beruhen, sind bibliometrische und Peer Review-Verfahren nie völlig zu trennen: „*Peer review and bibliometric data are not completely independent. Citation data are ultimately based on researchers who cite or do not cite particular publications. The same communities are the source of peer review data. [...] Peer review decisions may have been influenced by prior knowledge of bibliometric data.*“¹⁷

Zum anderen herrscht aber weitgehend Übereinstimmung, dass die Bibliometrie nicht als alleiniges Instrument der Forschungsevaluierung dienen kann. Um Forschung effektiv bewerten und steuern zu können, benötigt es das Wissen von Expertinnen und Experten. Quantitative Analysen zur Unterstützung von Peer Review-Verfahren werden im Allgemeinen dann befürwortet, wenn diese mit dem richtigen Augenmaß und Kontextwissen und damit verantwortungsbewusst durchgeführt werden.¹⁸ Die Bibliometrie bietet dann die Möglichkeit, mittels quantitativer Informationen die strategische Entscheidungsfindung im Rahmen der Forschungssteuerung zu unterstützen. Häufig unterstützt die Bibliometrie somit Peer Review-Prozesse bei der Bewertung einer Forschungseinheit (z.B. einer Forscherin oder eines Forschers, einer Forschungsgruppe, einer Abteilung oder einer Forschungseinrichtung) im Sinne eines sogenannten *Informed Peer Reviews*. Allerdings gibt es auch Verwendungen ohne Peer Review. So werden bibliometrische

¹⁶ Marx, Werner und Lutz Bornmann, „Bibliometrische Verfahren zur Bewertung von Forschungsleistung: Hintergründe, Aussagekraft und Grenzen“, *Soziale Welt*, 66 (2015), S. 161–179 (S. 167).

¹⁷ Wilsdon, James u. a., 2015, S. 64.

¹⁸ Hicks, Diana u. a., „The Leiden Manifesto for Research Metrics“, *Nature*, 520 (2015), S. 9–11; Wilsdon, James u. a., 2015; Marx, Werner und Lutz Bornmann, 2015.

Methoden z.B. gelegentlich auch verwendet, um Indikatoren, die für die Vergabe von Forschungsmitteln verwendet werden können, zu definieren.¹⁹

Das Wissensgebiet der Bibliometrie, das in den letzten Jahren in den Vordergrund der Evaluationstätigkeit gerückt ist, wird gemeinhin als „*the application of mathematics and statistical methods to books and other media of communication*“²⁰ verstanden. Hier nimmt die Zitationsanalyse, deren Ursprung auf die Entwicklung des ersten Zitationsindex (SCI) in den 1960er Jahren zurückzuführen ist, eine dominante Stellung ein. Dieser und ähnliche Indizes wurden ursprünglich als Rechercheinstrument insbesondere für Bibliothekarinnen und Bibliothekare konzipiert und herangezogen. Wenig später folgte eine erste Umdeutung der Zitationsanalyse als Instrument zum Zweck der wissenschaftlichen Erforschung von Wissenschaft selbst, bis schließlich das Potenzial für die auf quantitative Indikatoren gestützte Evaluation akademischer Leistung entdeckt wurde, wo auch der heutige primäre Anwendungsbereich von Bibliometrie angesiedelt ist.²¹

Die Anwendbarkeit bibliometrischer Methoden für die Evaluation von Forschungsleistung und Publikationstätigkeit ist seither Gegenstand reger Diskussionen. Fast zeitgleich mit der Entwicklung des ersten Zitationsindexes wurden Warnungen vor einer missbräuchlichen Verwendung – unter anderem als Indikator für Personalentscheidungen – ausgesprochen.²² Die Möglichkeiten und Grenzen sowie die positiven und negativen Effekte bibliometrischer Analysen im Bereich der Evaluation haben bereits eine eingehende wissenschaftliche Auseinandersetzung erfahren. Die Potentiale der Bibliometrie als Mittel für unabhängiger, objektivere, billigere und weniger zeitaufwändige Bewertungen werden den Limitierungen hinsichtlich der Aussagekraft über die Qualität wissenschaftlicher Leistungen gegenübergestellt.²³

¹⁹ CWTS, *Bibliometrics for Research Management and Research Evaluation: A Brief Introduction* (Leiden, 2018), S. 2-3.

²⁰ Pritchard, Allen, „Statistical Bibliography or Bibliometrics?“, *Journal of Documentation*, 25 (1969), S. 348–349 (S. 348).

²¹ Taubert, Niels Christian, „Bibliometrie in der Forschungsevaluation. Zur Konstitution und Funktionslogik wechselseitiger Beobachtung zwischen Wissenschaft und Politik“, in *Quoten, Kurven und Profile. Zur Vermessung der sozialen Welt.*, hg. von Jan Passoth und Joseph Wehner (Wiesbaden, 2013), S. 179–204.

²² Garfield, Eugene, „Citation Indexes in Sociological and Historical Research“, *Essays of an Information Scientist*, 1 (1963), S. 43-46.

²³ Für eine ausführliche Version der Peer Review vs. Bibliometrie-charakterisierten Debatte siehe Österreichischer Wissenschaftsrat, 2014.

Je nach Kontext, in dem die Bibliometrie genutzt wird, werden verschiedene Arten bibliometrischer Analysen verwendet. Die strategische Entscheidungsfindung auf Ebene einer ganzen Forschungseinrichtung erfordert häufig einen Überblick, der durch eine begrenzte Anzahl numerischer Indikatoren rasch zu erzielen ist. Entscheidungsfindungen auf niedrigeren Ebenen innerhalb einer Organisation können von umfangreicher bibliometrischen Analysemethoden z.B. Cluster-, *Strength, Potential and Risk* (SPR)-Analysen oder bibliometrischen Visualisierungen, profitieren, bei denen numerische Indikatoren durch weitere bibliometrische Informationen ergänzt werden.

Bibliometrische Analysen bieten eine gewisse Flexibilität in Bezug auf den Detaillierungsgrad, mit dem die Informationen dargestellt werden, sodass eine bestmögliche Anpassung an den jeweiligen Zweck möglich ist. In einfachen Fällen besteht eine bibliometrische Analyse nur aus einer sehr begrenzten Zahl an numerischen Indikatoren, wie z.B. Zitationsindikatoren oder h-Faktoren. Aufgrund der länderübergreifenden Verfügbarkeit bibliometrischer Daten können jedoch detailliertere Analysen durchgeführt und graphisch aufbereitet werden, z.B. Aufschlüsselungen nach Wissenschaftsbereichen, Zeittrends, geographische Verteilungen bibliometrischer Indikatoren auf der Ebene von Regionen, Ländern oder International oder Visualisierungen wissenschaftlicher Netzwerke (z.B. Autorennetzwerke oder Zitationsnetzwerke). Eine grafische Darstellung der wichtigsten Schlagworte z.B. aus Titeln und Abstracts von Veröffentlichungen in sogenannten *term maps* ermöglicht einen raschen Überblick über den wissenschaftlichen Inhalt einer großen Anzahl von Publikationen. Mit Hilfe von *Overlay*-Darstellungen ist es möglich, zusätzliche Informationen einzubinden, um z.B. einen Vergleich der Forschungsschwerpunkte verschiedener Universitäten zu ermöglichen.²⁴ Ein Beispiel hierfür ist die Visualisierung des Forschungsprofils der *Leiden University* (Abbildung 1). Die Daten hierfür basieren auf Publikationen der Universität in der WoS-Datenbank zwischen 2007 und 2010, sowie Zitationen bis 2011.

²⁴ CWTS, 2018, S. 9.

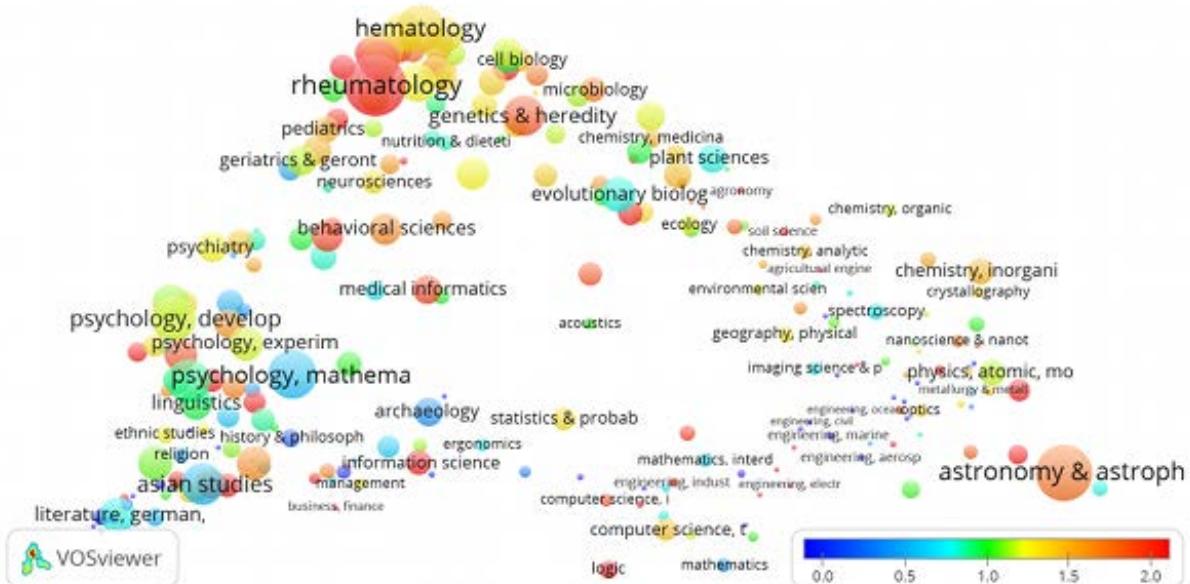


Abbildung 1: Overlay-Darstellung der Forschungsschwerpunkte der Leiden University²⁵

Die Kreise repräsentieren die Fachkategorien (*subject categories*) der WoS-Datenbank. Die Größe der einzelnen Kreise stellt den relativen Grad an Aktivität (Zahl der Publikationen) in diesem Fachbereich dar, gemessen am Weltdurchschnitt. Die Farbe der Kreise wiederum gibt den Impact im Vergleich zum Weltdurchschnitt an. Grün repräsentiert den Weltdurchschnitt, Blau indiziert einen sehr unterdurchschnittlichen Impact und Rot einen sehr überdurchschnittlichen.²⁶

Bibliometrische Methoden können, unter anderem, folgende Informationen zugänglich machen:²⁷

- Wissenschaftlicher Output – Informationen über die Anzahl der Veröffentlichungen einer Forschungseinheit.
- Wissenschaftlicher Impact – Informationen über die Anzahl der Zitate, die Publikationen einer Forschungseinheit erhalten haben.

²⁵ CWTS, „University Profile Maps“, (Leiden, 2019b) <<https://www.vosviewer.com/university-profile-maps>> [zugegriffen 31. Juli 2019].

²⁶ Ebd.

²⁷ CWTS, 2018, S. 10.

- Wissenschaftliche Zusammenarbeit – Informationen über kooperativ verfasste Publikationen, die sich z.B. auf nationale und internationale Zusammenarbeit oder auf die Zusammenarbeit zwischen Universitäten und Industrie konzentrieren.
- Mobilität – Informationen über Forscherinnen und Forscher, die ihre Institutzugehörigkeit ändern.
- Interdisziplinarität – Informationen über die Zugehörigkeit von Publikationen zu einzelnen Fachdisziplinen, deren Verbindungen und gegenseitigen thematischen Bezüge.
- Emergente Themen – Informationen über neue Themenbereiche und Themenkonjunkturen.
- Geschlecht – Informationen über das Geschlecht der Forschenden.
- Open Access-Publishing – Informationen über den Open Access-Status von Publikationen.

Bibliometrische Methoden erfahren in der Universitätsmedizin, den Natur- und teilweise auch in den Ingenieurwissenschaften auf internationaler Ebene bereits breite Anwendung; in den Geistes-, Sozial- und Kulturwissenschaften werden sie jedoch nur sehr zurückhaltend angenommen. Begründet wird dies mit den fächerspezifischen Gegebenheiten, wie der Publikationskultur, dem Erfassungsgrad der jeweiligen Datenbanken, oder landesbezogener Spezifika, wie der Sprache der wissenschaftlichen Veröffentlichungen, die nicht angemessen berücksichtigt werden können.²⁸

Auf die Anwendungsgrenzen der Bibliometrie wurde mit der Weiterentwicklung bestehender Indikatoren und der Entwicklung alternativer Instrumente reagiert, die einerseits die allgemeinen Schwachstellen und andererseits die disziplinspezifischen Gegebenheiten berücksichtigen sollen. Vor allem in Hinblick auf die Geisteswissenschaften wird versucht, vorhandene Indikatoren anzupassen oder zu erweitern. So wurde der von Jorge E. Hirsch entwickelte *h-Index* bereits mehrfach angepasst; mittlerweile existieren mehr als 40 Variationen, wie z.B. der m-

²⁸ ÖAW, *Qualitätskriterien für Publikationen in den Geistes-, Sozial- und Kulturwissenschaften: Bericht der AD-Hoc-Kommission der Phil.-Hist. Klasse der ÖAW* (Wien, 2012), S. 2–3.

Quotient von Hirsch selber oder der hla-Index.²⁹ Weiters wurde der *Book Citation Index* eingeführt, der allerdings noch nicht für eine breitere Anwendung geeignet ist.³⁰ Mit Referenzwerten wird versucht, Faktoren, die einen Einfluss auf Zitationen haben, aber nichts über die Qualität der Leistung oder den Impact aussagen (z.B. Publikationsjahr oder Fachgebiet), zu neutralisieren.³¹ Hinzu kommt, dass der Grad der Abdeckung von Veröffentlichungen in internationalen Datenbanken wie WoS oder Scopus je nach Forschungsdisziplin stark variiert, weshalb auch alternative Datenquellen, wie beispielsweise Google Scholar, Google Book Search oder Bibliothekskataloge, genutzt werden.³²

Neuere, mit dem Begriff *Altmetrics* assoziierte Trends, versuchen mit „alternativen Metriken“ das Maß der Wahrnehmung einer Publikation zu messen. Der Fokus schwenkt dabei von einer klassischen Impact-Analyse auf den Ansatz der *Societal Impact*-Messung um, da gesellschaftliche Wirkung von Forschung in wissenschaftsstrategischen und -politischen Entscheidungen mehr Gewicht zugemessen wird. *Altmetrics*-basierte Indikatoren sollen die gesamte Aufmerksamkeit, die eine Veröffentlichung bekommt, erfassen, sei es in sozialen Netzwerken, Zeitungen, Blogs oder Policy-Dokumenten.³³ Der Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) hat alle Publikationen, welche aus FWF-geförderten Projekten hervorgingen, ausgewertet, da der Druck „den Erfolg der geförderten Forschung zu belegen“³⁴ stetig wächst und

²⁹ Harzing, Anne-Wil, Satu Alakangas und David Adams, „hla: An Individual Annual h-Index to Accommodate Disciplinary and Career Length Differences“, *Scientometrics*, 99 (2014), S. 811–821; Niederklapfer, Thomas, *Der Hirsch-Index und seine Varianten: Eine Abhandlung zur Theorie und Praxis* (Innsbruck, 2014); der h-Index wird ab Seite 24 detaillierter behandelt.

³⁰ Einschränkungen bei der Anwendung ergeben sich unter anderem durch den hohen Anteil an Publikationen ohne Adressinformationen, die zum Teil fehlende Möglichkeit der Unterscheidung zwischen verschiedenen Buchtypen und hinsichtlich der Sprache, da hauptsächlich englischsprachige Veröffentlichungen inkludiert sind; vgl. Gorraiz, Juan, Philip Purnell und Wolfgang Gläzel, „Opportunities for and Limitations of the Book Citation Index“, *Journal of the American Society for Information Science and Technology*, 64 (2013), S.1388-1398.

³¹ Bornmann, Lutz, *Messung und Bewertung von Forschungsleistungen: Ein externes Gutachten für die Arbeitsgruppe des Österreichischen Wissenschaftsrates* (München, 2013), S.12-13.

³² Hammarfelt, Björn, „Beyond Coverage: Toward a Bibliometrics for the Humanities“, in *Research Assessment in the Humanities: Towards Criteria and Procedures*, hg. von Michael Ochsner, Sven E. Hug und Hans-Dieter Daniel (Zürich, 2016), S. 115–131.

³³ Tunger, Dirk, Marcel Clermont und Andreas Meier, „Altmetrics: State of the Art and a Look into the Future“, in *Scientometrics*, hg. von Mari Jibu und Yoshiyuki Osabe (London, 2018), S. 123–134.

³⁴ FWF, „Förderstatistiken: Altmetrics“ <<https://www.fwf.ac.at/de/ueber-den-fwf/foerderungsstatistiken/altmetrics/>> [zugegriffen 4. September 2019].

hierbei auch der gesellschaftliche Impact von zunehmender Bedeutung ist. Gemäß FWF bietet *Altmetrics* hierfür eine „Annäherung, die der FWF testen möchte“.³⁵

Abbildung 2 stellt den sogenannten *Altmetric Donut* einer der rund 19.500 vom FWF geförderten Publikationen dar, die mittels *Altmetric.com* ausgewertet wurde. Der Donut zeigt unter anderem, dass dieser Artikel, welcher im Dezember 2015 in *Biotechnology Advances* veröffentlicht wurde, einen *Altmetric Attention Score* von 4409 hat und er fast 10.000 mal auf Twitter erwähnt wurde. Auch andere Informationen sind zugänglich, z.B. dass es sich um einen Artikel handelt, welcher sich unter den Top fünf Prozent aller von Altmetric.com ausgewerteten Publikationen befindet.

Die Anwendung von *Altmetrics* zur Forschungsevaluierung im Rahmen von Forschungssteuerung und -management wird derzeit noch weithin kritisch bewertet. Im Zentrum der Kritik steht, dass es keine Standards gibt, was mit *Altmetrics* tatsächlich gemessen wird und wie. Nichtsdestotrotz wird das Potential einer erweiterten Messung gesellschaftlichen Impacts von Forschungsleistungen positiv bewertet.³⁶ In diesem Zusammenhang sind auch Bemühungen zu nennen, weitere Indikatoren zu entwickeln, um über Publikationsleistung hinaus weitere wissenschaftlich relevante Aspekte zu erfassen, wie etwa Transferleistungen und wissenschaftliche Reputation, die sich z.B. durch Funktionen in Kommissionen, Gutachtergremien oder Zeitschriften-Boards, Einladungen zu Vorträgen etc., manifestieren. Befürworterinnen und Befürworter solcher Indikatoren argumentieren, dass solche Leistungen eher über disziplinäre Grenzen hinweg vergleichbar sind, und somit eine sinnvolle Ergänzung zu den klassischen bibliometrischen Indikatoren bieten.³⁷

³⁵ Ebd.

³⁶ Tunger, Dirk, Marcel Clermont und Andreas Meier, 2018.

³⁷ Leitner, Karl-Heinz, *Messung und Bewertung von Forschungsleistung in Österreich: Externes Gutachten* (Wien, 2013), S. 21.

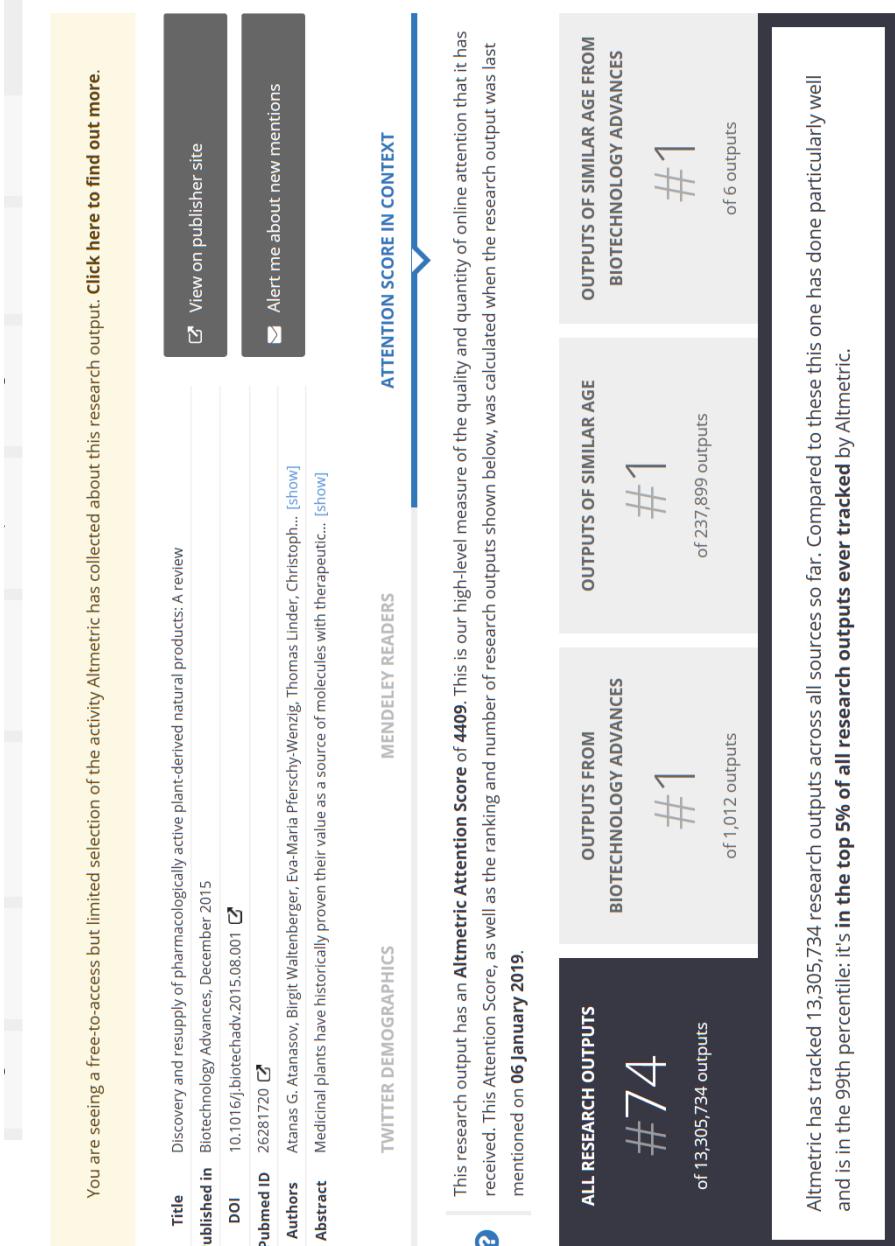


Abbildung 2: Beispiel eines Altmetric Donuts³⁸

38 Altmetric, „Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review“ (2019) <<https://www.altmetric.com/details/4411153>> [zugegriffen 31. Juli 2019].

Es herrscht weitgehende Übereinstimmung, dass bibliometrische Indikatoren zwar keine exakten Angaben über wissenschaftliche Bedeutsamkeit liefern, aber sie bieten Informationen über die wissenschaftliche Aufmerksamkeit, die Publikationen, Forscherinnen und Forscher oder Forschungseinrichtungen erfahren. Für wichtige Entscheidungen der Forschungssteuerung sind diese Informationen zu begrenzt, um als einzige Informationsquelle genutzt werden zu können. Bibliometrische Informationen müssen daher als lediglich ein Element innerhalb einer breiteren Palette von Evaluationsmethoden (z.B. Peer Review), erweiterter Datengrundlagen (Preise oder Third Mission Engagements), und eventuell alternativer Indikatoren (z.B. *Altmetrics*) betrachtet werden, die in der Entscheidungsfindung im Rahmen der Forschungssteuerung unterstützend zur Verfügung stehen. Weiters ist es wichtig, sich ins Bewusstsein zu rufen, dass es sich bei bibliometrischen Indikatoren um Daten handelt, die der menschlichen Auswahl und Interpretation bedürfen. Die Interpretation dieser Daten kommt nicht ohne das Fachwissen von Expertinnen und Experten, d.h. ohne *Peers*, aus, oder ohne die differenzierte Einbettung in oftmals komplexe Fachzusammenhänge.

3 Grundlagen der Bibliometrie³⁹

Das folgende Kapitel bietet eine kurze Einführung in die Verwendung der Bibliometrie für die Forschungssteuerung und -evaluierung. Es werden die wichtigsten Datenquellen für bibliometrische Analysen vorgestellt, sowie ein Überblick über verschiedene Arten bibliometrischer Indikatoren, deren Anwendungsmöglichkeiten und Grenzen gegeben.

3.1 Bibliometrische Datenquellen

Die meisten Universitäten und Forschungseinrichtungen verfügen über ein eigenes Forschungsinformationssystem, das einerseits aus internen Datenquellen gespeist wird und andererseits über externe Datenquellen abgeglichen und ergänzt wird. Für die Generierung bibliometrischer Indikatoren sind international zugängliche Datenbanken unumgänglich. Eugene Garfields Arbeit am *Institute of Scientific Information* zur Erstellung von Literaturdatenbanken, welche erstmals auch Zitierungen der katalogisierten Publikationen erfassten, war der Ausgangspunkt für den 1980 entstandenen *Science Citation Index*.⁴⁰

Heute sind *Web of Science* (WoS), bereitgestellt von *Clarivate Analytics*, und *Scopus*, bereitgestellt von *Elsevier*, die beiden am häufigsten verwendeten bibliometrischen Datenquellen. Auch *Google Scholar* wird zunehmend eingesetzt. Darüber hinaus stützen sich bibliometrische Analysen, die auf bestimmte Wissenschaftsbereiche beschränkt sind, zuweilen auf fachspezifische Datenquellen wie *INSPIRE*, *MathSciNet*, *PsycINFO* und *PubMed*. Interessante neuere Datenquellen und Analysewerkzeuge, die für bibliometrische Analysen potenziell von Nutzen sein können, sind *Microsoft Academic* und *Crossref*. Deren Nutzung befindet sich jedoch noch in einem frühen Stadium; somit sind die Erfahrungen mit diesen Datenquellen noch begrenzt.⁴¹

Web of Science enthält Informationen über eine Reihe von Zitationsindizes. Die am häufigsten verwendeten sind Teil der sogenannten *Web of Science Core Collection*. Dazu gehören der *Sci-*

³⁹ Dieses Kapitel basiert weitgehend auf dem Handbuch *Bibliometrics for Research Management and Research Evaluation: A Brief Introduction*, welches von CWTS (2018) im Auftrag des Österreichischen Wissenschaftsrates erstellt wurde, sowie eigenen Recherchen und Gesprächen, unter anderem mit CWTS, zu den neuesten Entwicklungen auf dem Gebiet der Szentometrie. Das Handbuch befindet sich in englischer Sprache und voller Länge im Anhang.

⁴⁰ Marx, Werner und Lutz Bornmann, 2015.

⁴¹ CWTS, 2018, S. 6.

ence Citation Index Expanded, der *Social Sciences Citation Index*, der *Arts & Humanities Citation Index* und der *Emerging Sources Citation Index*. Diese Indizes umfassen wissenschaftliche Zeitschriften in den Bereichen Natur- und Ingenieurwissenschaften, Sozialwissenschaften, Kunst und Geisteswissenschaften. Darüber hinaus sind der *Conference Proceedings Citation Index* und der *Book Citation Index*, die Tagungsbeiträge und Bücher umfassen, ebenfalls in der *Web of Science Core Collection* enthalten. *Scopus* umfasst, ähnlich wie WoS, Zeitschriften, Konferenzberichte und Bücher in den Natur-, Ingenieur-, Sozial-, Kunst- und Geisteswissenschaften, allerdings wird hier nicht nach den verschiedenen Zitationsindizes unterschieden.

Scopus und WoS benötigen beide ein Abonnement, um über eine Web-Oberfläche wissenschaftliche Literaturrecherchen bzw. einfache bibliometrische Analysen durchführen zu können. Für großangelegte bibliometrische Analysen sind sie jedoch nicht geeignet. Dies erfordert eine spezielle Lizenz, die einen direkten Zugriff auf die Daten ermöglicht und einen flexibleren Einsatz der Daten sowie deren manuelle Bereinigung erlaubt. Solche Analysen können mit kommerziellen bibliometrischen Analysewerkzeugen wie *InCites* und *SciVal* durchgeführt werden, die auf Daten von WoS und *Scopus* zurückgreifen.

CWTS erstellt ein eigenes *Citation Index System (CI-System)*, basierend auf der *Web of Science Core Collection*. Dieses verfügt über spezielle Algorithmen für den Zitationsabgleich sowie die Filterung von Selbstzitaten, und baut darüber hinaus ein eigenes Vereinheitlichungssystem für Namens- und Adressvarianten von Institutionen auf, um aussagekräftigere und genauere Analysen zu ermöglichen. Das CI-System verwendet auch ein selbstgenerierendes Klassifikationssystem (siehe Seite 29), welches auf den Zitationsbeziehungen der einzelnen Artikel basiert.⁴²

Google Scholar ist eine Online-Suchmaschine für wissenschaftliche Literatur, kann aber auch als Datenquelle für bibliometrische Analysen dienen. Im Vergleich zu WoS und *Scopus* hat *Google Scholar* zwei wesentliche Vorteile: es ist frei verfügbar und bietet eine umfassendere Abdeckung der wissenschaftlichen Literatur. Dies gilt insbesondere für bibliometrische Analysen in der Informatik und den Sozial- und Geisteswissenschaften, aber auch für Analysen, die sich nicht ausschließlich auf internationale wissenschaftliche Literatur, sondern auch auf nationale und regionale Publikationen konzentrieren. Für derartige Vorhaben bieten WoS und *Scopus* nur eine begrenzte Abdeckung. *Google Scholar* weist jedoch auch bedeutende Einschränkungen auf. Es fehlt weitgehend an Transparenz, da nur sehr wenige Informationen darüber

⁴² Noyons, Ed, *Advanced Bibliometric Analysis in the Context of Research Evaluation - Part 1: Strength, Promising & Risk Areas (SPRA)*, Vortrag für den Österreichischen Wissenschaftsrat, (Wien, 2019a), S. 8.

vorhanden sind, was abgedeckt wird und was nicht. *Google Scholar*-Daten können leichter manipuliert werden. Auch sind die Daten von *Google Scholar* von geringerer Qualität als jene von WoS und *Scopus*. Letztlich besteht ein signifikanter Nachteil darin, dass es fast unmöglich ist, mit *Google Scholar* großangelegte bibliometrische Analysen durchzuführen, da kein umfangreicher Zugriff auf die Rohdaten unterstützt wird.

3.2 Bibliometrische Impact-Indikatoren

Der wissenschaftliche Impact einer Publikation wird weitgehend durch die Zählung der erhaltenen Zitationen analysiert (*Citations per Paper*). „Zitierungen sind“, wie Marx und Bornmann treffend formulieren, „die Fäden im Netzwerk der wissenschaftlichen Arbeiten, welche diese inhaltlich verbinden“.⁴³ Sie werden aus unterschiedlichen Gründen gesetzt: einige deuten darauf hin, dass die zitierende Veröffentlichung auf der angeführten Publikation aufbaut. Dies kann als Anerkennung der Bedeutung der zitierten Veröffentlichung interpretiert werden. Negative Zitationen sind von entgegengesetzter Natur und spiegeln eine kritische Perspektive auf die zitierte Veröffentlichung wider. Viele Zitationen sind jedoch weder positiv noch negativ und repräsentieren einen lediglich oberflächlicheren Zusammenhang zwischen den beiden Publikationen. Geht der Veröffentlichung einer Publikation ein Peer Review-Verfahren voraus, so werden auch die Zitierungen in der Regel evaluiert.⁴⁴ Dennoch, angesichts der Vielfalt der Zitationen, erlauben Zitationszahlen nur eine vage Annäherung an den tatsächlichen wissenschaftlichen Impact. Vielfach werden Zitationszählungen auch als Indikatoren für wissenschaftliche Qualität interpretiert, diese Interpretation ist jedoch verzerrend und unzulässig.

Unter den Impact-Indikatoren sind der Journal-Impact Faktor und der h-Index die bekanntesten Beispiele. Impact-Indikatoren können auf vielfältige Weise klassifiziert werden. Größenabhängige Impact-Indikatoren spiegeln den gesamten wissenschaftlichen Impact der Veröffentlichungen einer Forschungseinheit wider, während größtenteils unabhängige den durchschnittlichen wissenschaftlichen Impact widerspiegeln.

Tabelle 2 illustriert die Unterschiede, welche sich ergeben, je nachdem ob ein großenabhängiger (*Total Citation Score*, TCS) oder ein größtenteils unabhängiger Indikator (*Mean Citation Score*, MCS) zur Analyse des *Leiden Rankings* österreichischer Universitäten für die Jahre 2014-2017 herangezogen wird. Betrachtet man den TCS, so belegen die Medizinische Universität Wien,

⁴³ Marx, Werner und Lutz Bornmann, 2015, S. 165.

⁴⁴ Ebd.

die Universität Wien und die Technische Universität Wien die Ränge 1-3 (Rot, Orange, Gelb) innerhalb Österreichs. Rang 1 des größtenunabhängigen MCS wird weiterhin durch die Medizinische Universität Wien belegt, Rang 2 und 3 hingegen von der Medizinischen Universität Innsbruck und der Universität Graz. Die Universität Wien befindet sich in diesem Kontext auf Rang 5, die Technische Universität auf Rang 7.

	TCS	MCS
Medizinische Universität Wien	27.943	7,24
Universität Wien	24.713	6,34
Technische Universität Wien	15.245	5,71
Universität Innsbruck	11.129	5,67
Medizinische Universität Innsbruck	10.497	7,01
Medizinische Universität Graz	9.941	6,53
Universität für Bodenkultur Wien	9.413	6,33
Universität Graz	9.335	6,63
Technische Universität Graz	7.588	5,36
Universität Linz	5.260	4,75

Tabelle 2: Total Citation Score (TCS) und Mean Citation Score (MCS) österreichischer Universitäten im Zeitraum 2014-2017; Bedeutung der Schattierung: 1. Reihung Rot, 2. Reihung Orange, 3. Reihung Gelb⁴⁵

Weiterhin unterscheidet man zwischen Impact-Indikatoren, die Zitationen direkt erfassen, und jenen, die zunächst die am häufigsten zitierten Publikationen identifizieren und dann nur diese Publikationen analysieren. Hier muss ein Grenzwert n festgelegt werden, ab wie vielen Zitationen eine Veröffentlichung als hoch zitiert gilt. Im Vergleich zu den einfachsten Impact-Indikatoren, die direkt auf der Zählung von Zitationen basieren, sind jene, die auf der Zählung von viel zitierten Publikationen basieren, weniger empfindlich gegenüber Publikationen mit einer sehr großen Anzahl von Zitationen. Impact-Indikatoren, die auf der Zählung viel zitierter Publikationen basieren, sind daher robuster. Dies wird oft als Vorteil dieser Indikatoren angesehen.

Einer der bekanntesten Impact-Indikatoren ist der h-Index. Er wird derart bestimmt, dass die Analyseeinheit h Publikationen hat, die jeweils mindestens h Zitationen erhalten haben. Der h-Index ist ein größtenabhängiger Impact-Indikator. Er ist relativ unempfindlich gegenüber einzelnen Publikationen mit einer sehr hohen Anzahl von Zitationen. Ein Kritikpunkt des h-Index liegt

⁴⁵ Adaptiert von CWTS, „Leiden Ranking Listview 2019“ (Leiden, 2019a) <<https://www.leidenranking.com/ranking/2019/list>> [zugegriffen 1. August 2019].

in seiner möglichen Inkonsistenz.⁴⁶ So können zwei Forschungseinheiten bei gleicher Verbesserung in Bezug auf ihre Publikationen den h-Index unterschiedlich verbessern. Abbildung 3 illustriert diesen Sachverhalt. Die linke Grafik zeigt den Ausgangs-h-Index von jeweils fünf bzw. vier für zwei Einheiten. Beide Einheiten publizieren zwei zusätzliche Veröffentlichungen mit jeweils acht Zitationen. Dies führt zur rechten Grafik, welche den h-Index für die eine Analyseinheit unverändert bei fünf verbleiben lässt, für die andere jedoch auf sechs erhöht. Beide Einheiten konnten ihren Output und ihre Wirkung gleichermaßen verbessern, im Ranking des h-Index hat sich ihre Situation jedoch umgekehrt.

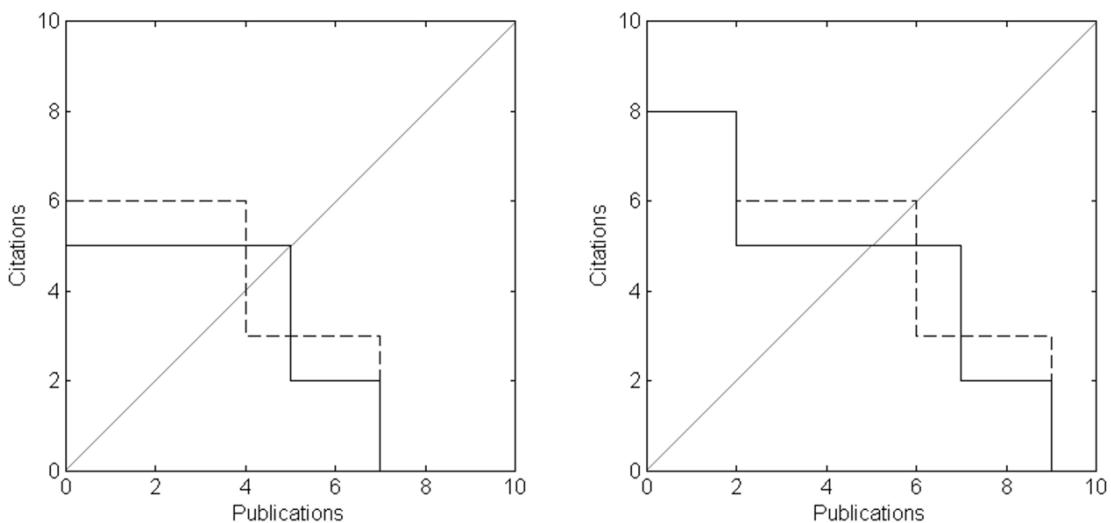


Abbildung 3: Illustration der möglichen Inkonsistenz des h-Index am Beispiel von 2 Forschungseinheiten. Links: Ausgangszustand, rechts: Veränderung durch eine zusätzliche Publikation mit 8 Zitierungen⁴⁷

Der h-Index kombiniert somit den Publikationsoutput und den Impact der Veröffentlichungen. Marx und Bornmann⁴⁸ heben hervor, dass der h-Index das komplexe Zusammenspiel unterschiedlicher Faktoren übersimplifiziert und zwei Indikatoren mehr oder weniger „willkürlich“ zu einer Maßzahl zusammenfasst. Hirsch selbst steht dem h-Index kritisch gegenüber, auch wenn

⁴⁶ Waltman, Ludo und Nees Jan van Eck, „The Inconsistency of the h-Index“, *Journal of the American Society for Information Science and Technology*, 63 (2012), S. 406–415.

⁴⁷ CWTS, 2018, S. 14.

⁴⁸ Marx, Werner und Lutz Bornmann, 2015, S. 171.

er damit „nicht völlig unzufrieden ist“.⁴⁹ Diese Kritikpunkte führen dazu, dass immer wieder neue Varianten des h-Index vorgeschlagen werden. Der m-Quotient stellt eine dieser Varianten dar, welche von Hirsch selbst entwickelt wurde, um Wissenschaftlerinnen und Wissenschaftler in unterschiedlichen Karrierestadien vergleichen zu können. Hierfür wird der h-Index der jeweiligen Autorinnen und Autoren durch die Anzahl der Jahre seit der ersten Publikation dividiert.⁵⁰

Verschiedene wissenschaftliche Bereiche haben unterschiedliche Zitierpraktiken. Aus diesem Grund gibt es große Unterschiede zwischen den Disziplinen hinsichtlich der Zitationsdichte, d.h. in der durchschnittlichen Anzahl der erhaltenen Zitate pro Veröffentlichung. So ist beispielsweise die Zitationsdichte in der Mathematik etwa eine Größenordnung kleiner als jene in einigen Bereichen der Biowissenschaften.

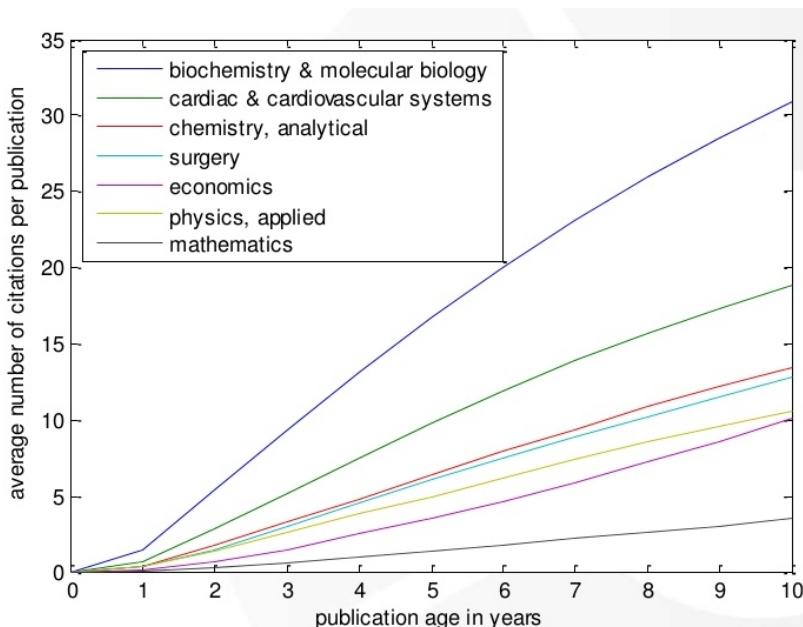


Abbildung 4: Disziplinäre Unterschiede in der Zitationsdichte⁵¹

Wenn eine bibliometrische Analyse des wissenschaftlichen Impacts mehrere Disziplinen umfasst, sollten daher disziplinäre Unterschiede in der Zitationsdichte durch eine Feldnormalisierung (*field normalization*) korrigiert werden. Die Feldnormalisierung erfolgt in der Regel durch

⁴⁹ Hirsch, Jorge E., „Superconductivity, What the H? The Emperor has No Clothes“, *Physics and Society*, 49 (2020), S. 4-9 (S. 4).

⁵⁰ Für eine interessante Kritik des h-Index sowie eine Übersicht über dessen geläufigsten Varianten siehe Niederklapfer, Thomas, 2014.

⁵¹ Waltman, Ludo, „Comparing Scientific Performance across Disciplines: Methodological and Conceptual Challenges“, in *7th International Conference on Information Technologies and Information Society (ITIS2015)* (Novo Mesto, 2015), S. 20.

den Vergleich der Anzahl der Zitationen einer Publikation mit der Anzahl der Zitationen anderer Publikationen aus demselben Gebiet.

Ältere Publikationen hatten mehr Zeit, das Interesse anderer Wissenschaftlerinnen und Wissenschaftler zu wecken als neuere Publikationen. Daher werden ältere Publikationen tendenziell häufiger zitiert als neuere. Auch hier kann eine Normalisierung zur Korrektur der Ungleichheit beitragen, indem man die Anzahl der Zitationen einer Publikation mit der Anzahl der Zitationen anderer Publikationen aus dem gleichen Jahr vergleicht.

Diese beiden Normalisierungen können in einem Indikator vereint werden, indem die Anzahl der Zitationen einer Veröffentlichung durch die durchschnittliche Anzahl der Zitate aller Publikationen desselben Fachbereichs und desselben Jahres dividiert wird. Weiters kann auch hier zwischen großenabhängigen (*Total Normalized Citation Score*, TNCS) und großenunabhängigen (*Mean Normalized Citation Score*, MNCS) Varianten unterschieden werden bzw. der Fokus lediglich auf die am häufigsten zitierten Veröffentlichungen gerichtet werden. Tabelle 3 illustriert mögliche Unterschiede, die sich aus der Verwendung dieser beiden normalisierten Indikatoren ergeben können. Betrachtet man bei den *Leiden Rankings* den großenabhängigen TNCS, so belegen die Universität Wien, die Medizinische Universität Wien und die Technische Universität Wien die ersten drei Ränge (Rot, Orange, Gelb) innerhalb Österreichs. Betrachtet man hingegen den großenunabhängigen MNCS, so findet sich die Medizinische Universität Wien nun auf Rang sechs, die Ränge eins und zwei werden von der Universität Wien, der Universität Graz und der dritte Rang von der Technischen Universität Wien und der Universität für Bodenkultur Wien belegt.

	TNCS	MNCS
Universität Wien	4.613	1,18
Medizinische Universität Wien	4.388	1,14
Technische Universität Wien	2.873	1,08
Universität Innsbruck	2.032	1,04
Universität für Bodenkultur Wien	1.689	1,14
Universität Graz	1.624	1,15
Medizinische Universität Innsbruck	1.587	1,06
Technische Universität Graz	1.540	1,09
Medizinische Universität Graz	1.496	0,98
Universität Linz	1.033	0,93

Tabelle 3: Total Normalized Citation Score (TNCS) und Mean Normalized Citation Score (MNCS) österreichischer Universitäten im Zeitraum 2014-2017; Bedeutung der Schattierung: 1. Reihung Rot, 2. Reihung Orange, 3. Reihung Gelb⁵²

Der MNCS ermöglicht einen Vergleich im internationalen Maßstab. So bedeutet ein Wert über bzw. unter eins, dass die Publikationen der jeweiligen Universität öfter bzw. weniger oft zitiert werden als der Durchschnitt der Publikationen aus demselben Forschungsbereich und gleichem Jahr. Aus Tabelle 3 ist ersichtlich, dass nur Publikationen der Medizinischen Universität Graz und der Universität Linz weniger oft zitiert wurden, und die Universität Wien 18 Prozent über dem Durchschnitt liegt.

Tabelle 4 illustriert die Bedeutung der größenumunabhängigen Indikatoren an einem weiteren Beispiel. Betrachtet man nur den Output der jeweiligen Universitäten, hier aus dem Bereich der Bildungsforschung⁵³, so sticht die Universität Wien mit 144 Veröffentlichungen klar hervor. Der normalisierte größenumunabhängige MNCS zeichnet jedoch ein anderes Bild. Die Veröffentlichungen der Universität Salzburg, wenn auch zahlenmäßig weitaus geringer (P=40), weisen eine beträchtlich größere relative Zitationshäufigkeit gegenüber dem Durchschnitt auf. Die Universität Salzburg liegt 56% über, die Universität Wien 31% unter dem Durchschnitt.

⁵² Adaptiert von CWTS, 2019a.

⁵³ Hesse, Friedrich W. u. a., 2019, S. 16.

Key Players	P	MNCS	MNJS	PP(top 10%)
University of Vienna	144	0,69	0,80	7%
University of Graz	68	1,09	0,90	9%
University of Salzburg	49	1,56	1,11	22%
Johannes Kepler University of Linz	40	0,74	0,76	6%
University of Innsbruck	31	0,53	0,71	6%

Tabelle 4: Vergleich unterschiedlicher bibliometrischer Indikatoren am Beispiel des Fachbereichs „Bildungswissenschaften“ an fünf österreichischen Universitäten (P: Gesamtzahl der Publikationen, MNCS: Mean Normalized Citation Score, MNJS: Mean Normalized Journal Score, PP(top 10%): Anteil der 10% meist zitierten Publikationen)⁵⁴

Die Normalisierung von Impact-Indikatoren für den wissenschaftlichen Bereich erfordert die Wahl eines Feldklassifikationssystems. In der Praxis werden dafür oftmals die Zeitschriftenkategorien gemäß WoS oder Scopus herangezogen. Eine Alternative ist die algorithmische Generierung von Forschungsfeldern auf der Ebene der Publikationen. Ein Beispiel hierfür stellt das vom CWTS generierte Klassifizierungssystem des CI-Systems basierend auf Zitationsbeziehungen dar. Von Klassifizierungssystemen auf Publikationsebene kann erwartet werden, dass sie genauere Ergebnisse liefern; eine Normalisierung auf Grundlage der bestehenden Zeitschriftenkategorien ist demgegenüber transparenter und leichter verständlich. Für das selbstgenerierende Klassifizierungssystem des CWTS wird jede Publikation einem Cluster (*research area*) zugewieilt. Diese können graphisch dargestellt werden, wobei jene Cluster, deren Artikel sich häufig zitieren, als näher zusammengehörig angesehen werden, und jene mit einer schwächeren Zitationsbeziehung eine größere Distanz voneinander aufweisen. Das CI-System gruppiert die Cluster auf vier Ebenen. Abbildung 5 kategorisiert die rund 4000 Cluster auf der Top-Ebene in fünf zentrale Disziplinen. Die Größe der Cluster steigt mit der Anzahl der enthaltenen Artikel an.⁵⁵

⁵⁴ Ebd.

⁵⁵ Noyons, Ed, 2019a.

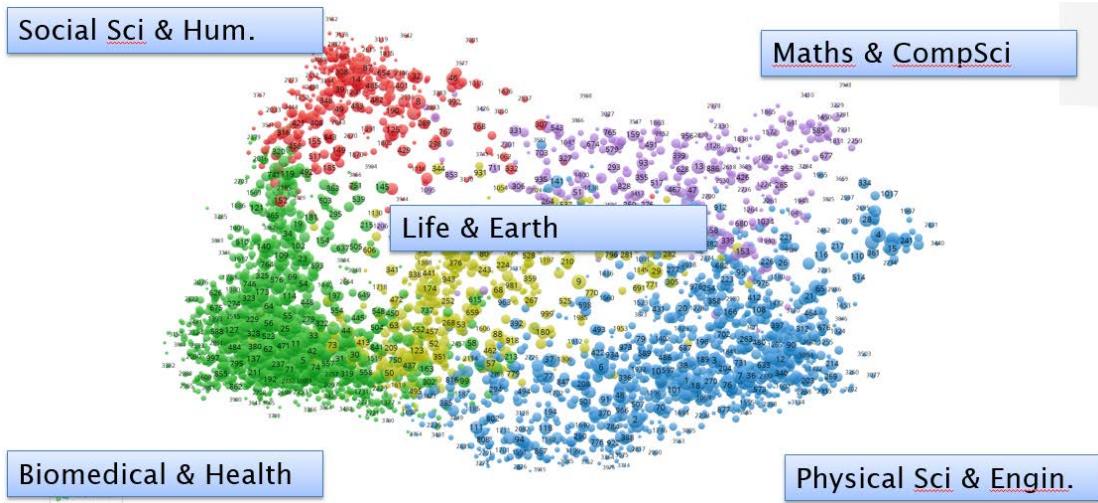


Abbildung 5: Top-Ebene des CWTS-Kategorisierungssystems: Fünf zentrale Wissenschaftsdisziplinen⁵⁶

Im Zuge der Normalisierung der Impact-Indikatoren muss ein Mindestalter der Veröffentlichungen, die einbezogen werden sollen, festgelegt werden. Sehr aktuelle Publikationen erhalten in der Regel keine oder fast keine Zitationen. Die Normalisierung bezüglich des Publikationsalters liefert für diese Veröffentlichungen keine aussagekräftigen Ergebnisse. Sehr aktuelle Veröffentlichungen, meist jene, die jünger als ein Jahr sind, werden daher in der Regel von der Berechnung normalisierter Impact-Indikatoren ausgeschlossen.

Häufig werden Publikationen von mehreren Forscherinnen und Forschern aus unterschiedlichen Forschungseinrichtungen, teilweise auch länderübergreifend, gemeinschaftlich verfasst. Wenn eine Publikation von mehreren Forschungseinheiten gemeinsam verfasst wurde, stellt sich die Frage, wie die Zitationen der Publikation auf die verschiedenen Forschungseinheiten verteilt werden. Die beiden am häufigsten verwendeten Ansätze sind das Voll- und das Bruchzählverfahren. Im Rahmen des Vollzählverfahrens werden die Zitate einer Publikation allen Mitgliedern einer Autorenschaft zugewiesen. Im Rahmen des Bruchzähl-Ansatzes werden die Zitate einer Veröffentlichung anteilsweise jeder der Co-Autoren-Forschungseinheiten zugeordnet. Im Falle von drei Forschungseinheiten erhält jede Autorin und jeder Autor somit ein Drittel der Zitationen der Gemeinschaftsveröffentlichung. Bei der Arbeit mit normalisierten Impact-Indikatoren hat der Bruchzählansatz den Vorteil, dass nicht nur die Unterschiede zwischen den Forschungsfeldern in der Zitationsdichte, sondern auch die Unterschiede zwischen den Forschungsfeldern hinsichtlich Zusammenarbeit korrigiert werden.

⁵⁶ Ebd., S. 16.

Darüber hinaus existieren Ansätze, die auf der Reihenfolge der Autorinnen und Autoren in der Autorenliste einer Veröffentlichung basieren. Sie vergeben beispielsweise die Zitate einer Publikation meist an die/den erst- bzw. letztgenannte/n Autorin/Autor, oder an die korrespondierende Person einer Publikation. Allerdings existieren keine universellen Richtlinien, die die Reihenfolge festlegen. „Die Reihenfolge der Autorenschaft ist eine Kulturfrage“, unterstreicht Ulrike Beisiegel.⁵⁷ Unabhängig von der spezifischen Fächerkultur kann es in allen Disziplinen zu Machtausübung kommen, um einen Platz unter den Autorinnen und Autoren zu erlangen, oder dass anhaltende Traditionen weitergepflegt werden, die den Institutsvorstand ehrenhalber als Autorin oder Autor genannt sehen. Ein Verständnis der Autorenkultur, beziehungsweise ein Bewusstsein der damit einhergehenden Problematiken, ist daher von großer Bedeutung.

Selbstzitationen werden bei der Berechnung von Impact-Indikatoren oft ausgeschlossen, auch wenn viele aus durchaus stichhaltigen Gründen gesetzt werden. Das Ausschließen von Selbstzitaten verhindert die damit mögliche Manipulation der Indikatoren.

Impact-Indikatoren werden auch für Zeitschriften ermittelt. Das bekannteste Beispiel ist der Journal-Impact Faktor (JIF), der im Wesentlichen der durchschnittlichen Anzahl von Zitaten entspricht, die in einem bestimmten Jahr von Publikationen erhalten wurden, die in den beiden Vorjahren in der Zeitschrift erschienen sind. So wird gerne von A+, A, B, C, D und E Zeitschriften gesprochen, oder von Q1, Q2, Q3 und Q4 Journals. Der von *Clarivate Analytics* auf der Grundlage von WoS-Daten berechnete Journal-Impact Faktor wird in den *Journal Citation Reports* (JCR)⁵⁸ veröffentlicht. Weitere Beispiele für Impact-Indikatoren für Zeitschriften sind der 5-Jahres-Wirkungsfaktor und der Artikel-Einfluss-Score, die ebenfalls im JCR veröffentlicht werden, sowie der *CiteScore*, *SCImago Journal Rank* (SJR) und *Source Normalized Impact per Paper* (SNIP), die von *Elsevier* auf der Grundlage von *Scopus*-Daten bereitgestellt werden.

⁵⁷ Lücke, Nicole, „Wer auf den Titel gehört“, *DUZ Magazin*, 2 (2010).

⁵⁸ Der *Journal Citation Report* wird jährlich veröffentlicht. Er enthält zusätzlich zum Journal-Impact Faktor auch detaillierte Informationen über die Anzahl der Publikationen und Zitate einer Zeitschrift. Die Journals werden, basierend auf dem Impact Faktor, Quartilen zugeordnet, wobei Zeitschriften der Top 25% als Q1-Journal bezeichnet werden. Siehe Clarivate Analytics, *InCites Indicators Handbook* (Philadelphia, 2018).

Der JIF ist jedoch nicht unumstritten; so wird z.B. auf die Möglichkeiten dessen Beeinflussung durch die Herausgeber hingewiesen.⁵⁹ Auch gibt es innerhalb einer Zeitschrift bedeutende Unterschiede zwischen den Publikationen hinsichtlich der Anzahl an Zitationen. Daher wird es häufig als problematisch angesehen, einen Journal-Impact Faktor zur Bewertung einzelner Publikationen zu verwenden.⁶⁰

CWTS hingegen lehnt die Verwendung von Journal-Impact Indikatoren zur Bewertung einzelner Publikationen nicht generell ab. Dies wird damit begründet, dass Zeitschriften unterschiedliche Qualitätsstandards haben, was sich zum Teil in den Indikatoren für den Impact von Zeitschriften widerspiegelt. CWTS verwendet *den Mean Normalized Journal Score* (MNJS), der ähnlich dem MNCS einzelner Forscherinnen und Forscher oder Forschungseinheiten berechnet und interpretiert wird.

Tabelle 5 stellt eine Übersicht der gebräuchlichsten bibliometrischen Impact-Indikatoren dar. Neben kurzen Erläuterungen werden auch deren bedeutendste Vor- und Nachteile festgehalten. An dieser Stelle soll auf einige generelle Punkte hingewiesen werden, welche auf alle Indikatoren gleichermaßen zutreffen: Eine Variation zwischen den Indikatoren ist gegeben, je nachdem welche Datenbank zur Berechnung verwendet wird. Es ist ebenso von Bedeutung, wie Zitationen im Falle von kooperativ verfassten Publikationen gezählt werden. Um aussagekräftige Interpretationen im Zuge einer Forschungsevaluation anstellen zu können, ist ein Wissen über diese Aspekte unumgänglich. Weiterhin ist festzuhalten, dass keiner der Indikatoren ein direkter Repräsentant für die Qualität von Forschung ist. Eine derartige Interpretation ist weitgehend umstritten.

⁵⁹ Falagas, Matthew und Vangelis Alexiou, „The Top-Ten in Journal Impact Factor Manipulation“, *Archivum Immunologiae et Therapiae Experimentalis*, 56 (2008), S. 223–226.

⁶⁰ Die *San Francisco Declaration on Research Assessment* lehnt eine derartige Anwendung des Journal-Impact Faktors ab, siehe Kapitel 3.3.

Indikator	Kurzbeschreibung	Repräsentant für	Vorteile	Nachteile
Total Citation Score (TCS)	Anzahl der Zitate; häufig dargestellt als P.	Den gesamten wissenschaftlichen Impact einer Einzelperson oder größerer Analyseeinheit.	<ul style="list-style-type: none"> – Einfache Berechnung 	<ul style="list-style-type: none"> – Größenabhängig; größere Analyseeinheiten tendieren zu besseren Resultaten – Ausreißerpublikationen mit vielen Zitaten wirken verzerrnd – Anfällig für Manipulation – Interdisziplinäre Vergleichbarkeit eingeschränkt – Benachteiligung neuerer Publikationen
Mean Citation Score (MCS)	Durchschnittliche Anzahl der Zitate.	Den durchschnittlichen wissenschaftlichen Impact einer Einzelperson oder größerer Analyseeinheit.	<ul style="list-style-type: none"> – Größenunabhängig; große wie kleine Analyseeinheiten können gute Resultate erzielen – Ausreißerpublikationen wirken weniger verzerrnd 	<ul style="list-style-type: none"> – Interdisziplinäre Vergleichbarkeit eingeschränkt – Benachteiligung neuerer Publikationen
h-Index	Anzahl der Publikationen (h) mit h Zitaten.	Den wissenschaftlichen Output in Relation zum Impact einer Einzelperson.		<ul style="list-style-type: none"> – Größenabhängig – Benachteiligt jene mit hohem Output und niedrigem Impact (insbesondere kleinere Disziplinen und Spezialgebieten) – Interdisziplinäre Vergleichbarkeit eingeschränkt – Benachteiligung neuerer und daher noch wenig häufig zitierter Publikationen – Inkonsistenzprobleme

Indikator	Kurzbeschreibung	Repräsentant für	Vorteile	Nachteile
Total Normalized Citation Score (TNCS)	Anzahl der Zitate, normalisiert für disziplinäre Unterschiede und Publikationsalter.	Den gesamten wissenschaftlichen Impact einer Einzelperson oder größerer Analyseeinheit.	<ul style="list-style-type: none"> - Leichtere interdisziplinäre Vergleichbarkeit - Keine Benachteiligung neuerer Publikationen 	<ul style="list-style-type: none"> - Größenabhängig - Ausreißerpublikationen wirken verzerrnd
Mean Normalized Citation Score (MNCS)	Anteil der Zitate über/unter dem Durchschnitt, normalisiert für disziplinäre Unterschiede und Publikationsalter.	<p>Den durchschnittlichen wissenschaftlichen Impact einer Einzelperson oder größerer Analyseeinheit.</p> <p>MNCS über/unter 1 deutet auf einen über-/unterdurchschnittlichen Impact hin.</p>	<ul style="list-style-type: none"> - Größenunabhängig - Ausreißerpublikationen wirken weniger verzerrnd - Erhöhte interdisziplinäre Vergleichbarkeit - Keine Benachteiligung neuerer Publikationen 	
Proportion of Highly Cited Publications ($PP \geq x$)	Anteil der Publikationen, welche x oder mehr Zitate haben.	Den Anteil der höchstzitierten Publikationen von allen Publikationen einer Einzelperson oder größerer Analyseeinheit.	<ul style="list-style-type: none"> - Größenunabhängig - Ausreißerpublikationen wirken weniger verzerrnd 	<ul style="list-style-type: none"> - Interdisziplinäre Vergleichbarkeit eingeschränkt - Benachteiligung neuerer Publikationen
Mean Normalized Proportion of Highly Cited Publications (PP($top\ x\%$))	Anteil der $x\%$ meist zitierten Publikationen, normalisiert für disziplinäre Unterschiede und Publikationsalter.	Den Anteil der Publikationen einer Einzelperson oder größerer Analyseeinheit an den Top $x\%$ aller Publikationen eines Feldes.	<ul style="list-style-type: none"> - Größenunabhängig - Ausreißerpublikationen wirken weniger verzerrnd - Erhöhte interdisziplinäre Vergleichbarkeit - Keine Benachteiligung neuerer Publikationen 	

Indikator	Kurzbeschreibung	Repräsentant für	Vorteile	Nachteile
<i>Journal-Impact Factor (JIF)</i>	Durchschnittliche Anzahl der Zitate von Publikationen einer Zeitschrift erschienen in den zwei Vorgängerjahren.	Den durchschnittlichen wissenschaftlichen Impact der Veröffentlichungen eines Journals.	<ul style="list-style-type: none"> – Ermöglicht die Einschätzung von Zeitschriften als ganzes – Geringere Benachteiligung neuerer Publikationen, da zeitlich auf zwei Jahre begrenzt 	<ul style="list-style-type: none"> – Interdisziplinäre Vergleichbarkeit eingeschränkt – Ausreißerpublikationen wirken verzerrnd – Manipulationsmöglichkeiten durch Steuerung von Editoren – Rückschlüsse auf Impact/Qualität einzelner Publikationen einer Zeitschrift umstritten
<i>Mean Normalized Journal Score (MNJS)</i>	Durchschnittliche Anzahl der Zitate von Publikationen der Zeitschriften erschienen in den zwei Vorgängerjahren, normalisiert für disziplinäre Unterschiede und Publikationsalter.	Den durchschnittlichen normalisierten Impact der Zeitschriften, in denen eine Analyseinheit veröffentlicht wurde. Ein MNJS über/unter 1 deutet auf einen über-/unterschnittlichen Impact hin.	<ul style="list-style-type: none"> – Ausreißerpublikationen wirken weniger verzerrnd – Erhöhte interdisziplinäre Vergleichbarkeit – Keine Benachteiligung neuerer Publikationen 	<ul style="list-style-type: none"> – Manipulationsmöglichkeiten auch durch Steuerung von Editoren – Rückschlüsse auf Impact/Qualität einzelner Publikationen einer Zeitschrift sind umstritten

Tabelle 5: Übersicht der gebräuchlichsten bibliometrischen Indikatoren⁶¹

⁶¹ CWTS, 2018; Hesse, Friedrich W. u. a., 2019; Marx, Werner und Lutz Bornmann, 2015.

3.3 Anforderungen und Grenzen bibliometrischer Methoden

Zur Ergänzung der Darstellung bibliometrischer Methoden, Datenquellen und Indikatoren widmet sich dieser Abschnitt den Anforderungen an die Bibliometrie, ihren Chancen, aber auch Grenzen. Als Grundlage hierfür wurde die *San Francisco Declaration on Research Assessment* (DORA), das *Leiden Manifesto* sowie die *Responsible Metrics* und *Recommendations* des *Metric Tide*-Reports herangezogen.⁶² Der Fokus richtet sich auf diese drei Dokumente, da sie bis dato die wohl einflussreichsten darstellen, auch wenn alle drei bereits etwas in die Jahre gekommen sind. Dies soll nicht die Bedeutung anderer derartiger Dokumente dämpfen, oder gar die Existenz neuerer, auch bedeutender fachspezifischer Initiativen wie jene von *Informatics Europe*⁶³, negieren.

Der Fokus von DORA ist auf die Verwendung des Journal-Impact Faktors (JIF) gerichtet, wobei die beiden anderen Dokumente versuchen, eine umfassendere Perspektive zu vermitteln. Gemeinsam wird dabei die Forderung nach einem reflektierten Umgang mit Indikatoren zur Evaluierung wissenschaftlichen Outputs hervorgehoben. Die Komplexität wissenschaftlicher Forschung verlangt demnach auch entsprechend differenzierte Methoden der Evaluierung, angepasst an den jeweiligen Kontext. Diese Methoden sollten daher nicht von vorneherein ausschließlich quantitativ oder qualitativ sein, die Kriterien nicht statisch oder einseitig. Vielmehr sollen sie flexibel, transparent bezüglich der Datenerfassung und Analyseprozesse und dadurch auch robust und aussagekräftig sein. Dieser Grundtenor spiegelt sich in der Ausgestaltung der *Responsible Metrics* des *Metric Tide*-Reports wider.

Die Auswahl der genannten Dokumente sowie deren Kernaussagen deuten bereits an, dass an dieser Stelle lediglich bestehende und fundierte Erkenntnisse zusammengetragen werden sollen. Mit unterschiedlicher Gewichtung richten sich die Dokumente an Bereitsteller sowie Bezieher bibliometrischer Analysen, nicht zuletzt aber auch Forscherinnen und Forscher, deren Arbeit die Grundlage derartiger Analysen liefert. Sie alle tragen Verantwortung, bibliometrische Methoden als ein tragfähiges und aussagekräftiges Mittel zur Evaluierung von Forschungsleistung zu ermöglichen.

⁶² DORA, „Declaration on Research Assessment“ <<https://sf-dora.org/>> [zugegriffen 4. September 2019]; Hicks, Diana u. a., 2015; Wilsdon, James u. a., 2015.

⁶³ Informatics Europe, „Joint Statement on Informatics Research Evaluation“ (2020) <<https://www.informatics-europe.org/news/545-joint-statement-on-informatics-research-evaluation.html>> [zugegriffen 22. Juli 2020].

Aussagen, wonach quantitative Methoden Peer Review-Verfahren nicht ersetzen können, bibliometrische Methoden die Vielfalt der Forschungsdisziplinen sowie Karrierepfade berücksichtigen sollten, oder jenseits von Publikationen in wissenschaftlichen Journals vielfältige Formen des Forschungsoutputs erfasst werden können, sind schwerlich als bahnbrechend zu rubrizieren. Auch die Empfehlung, die Leitlinien und Kriterien von Evaluierungsverfahren transparenter zu kommunizieren, oder die Aufforderung, den JIF weitgehend zu vermeiden, insbesondere zur Evaluierung von Einzelpersonen, bzw. ihn nur in Kombination mit anderen Performance-Indikatoren zu verwenden, erscheinen selbsterklärend. Auch das Einmahnen verantwortungsvoller Publikationspraktiken oder die Empfehlung der Verwendung einer ORCID Identifikation mögen redundant wirken. Und doch sollen diese Kernaussagen der drei Dokumente aufs Neue ins Treffen geführt werden, bilden sie doch die Rahmenbedingungen, um bibliometrische Methoden gewinnbringend einsetzen zu können.

Auf Grundlage der Aufarbeitungen des aktuellen Wissensstands, Gesprächen mit Expertinnen und Experten des Forschungsfeldes sowie Vertreterinnen und Vertretern nationaler wie internationaler Hochschuleinrichtungen wurde die nachfolgende Tabelle 6 erstellt. Diese enthält eine nicht als abschließend zu betrachtende Zusammenfassung (*living document*) der Anwendungsbereiche, Chancen, Grenzen und auch Gefahren bibliometrischer Methoden auf unterschiedlichen Ebenen der Evaluierung (Individualebene, Institutionelle Ebene und Systemebene) aus Sicht des Wissenschaftsrates.

Individualebene (einzelne Wissenschaftler, wissenschaftliches Personal)			
Anwendungsbereiche	Chancen	Grenzen und Gefahren	
<ul style="list-style-type: none"> Berufungen Beförderungen Entscheidungen über Rufabwendung Belohnungen (Boni, Leistungszulagen) Preise Beurteilung der Vorleistung bei Fördermittelanträgen 	<ul style="list-style-type: none"> Übersichtliche Darstellung des Forschungs-Outputs Verbesserte Möglichkeiten der objektiven Beurteilung Gute Vergleichbarkeit des wissenschaftlichen Outputs von Kandidatinnen und Kandidaten innerhalb eines Fachgebiets Wertvolle quantitative Ergänzung von Peer Review-Verfahren Ermöglicht Einblicke in die zeitliche Entwicklung wissenschaftlicher Aktivitäten Zitierhäufigkeit als wertvolle quantitative Ergänzung zur Beurteilung einzelner Veröffentlichungen Einblick in internationale und interdisziplinäre Kooperationen von Kandidatinnen und Kandidaten Nutzung von Daten als Mittel zur Erhöhung der Gerechtigkeit in Bezug auf Gender. 	<ul style="list-style-type: none"> Negativanteile für langfristige, risikoreiche Forschung mit längeren Publikationspausen (welche aber eventuell bahnbrechende Erkenntnisse bringen könnte; <i>Blue Sky</i>) Unzureichende Beurteilung bibliometrisch schlechter messbarer Leistungsbereiche (z.B. Lehre, Wissenstransfer, Öffentlichekeitsarbeit, Wissenschaftsmanagement) Crowding-out-Effekte: Betonung extrinsischer Motivation, Unterbewertung intrinsischer Motivation Hindernis für (bislang) wenig ausgewiesene Nachwuchswissenschaftler; Verstärkungseffekt Matthäus-Effekt 	

Institutionelle Ebene (Fakultäten, Departments, Zentren, Cluster, Netzwerke, Fächergruppen)		
Anwendungsbereiche	Chancen	Grenzen und Gefahren
<ul style="list-style-type: none"> Leistungsbezogene Mittelvergabe Stellen- und Professurzuweisungen Infrastrukturinvestitionen Beurteilung der Forschungsumgebung (z.B. bei Beantragung von Verbundforschungsfördermitteln) Evaluation zur Entscheidung über Fördermittel, institutionelle Mittelallokation, strategische Mitte Benchmarking zur Einschätzung des Fächerportfolios und regionaler Schwerpunkte für potentielle Kooperationspartner oder zur Beurteilung der Attraktivität für mobile Doktoranden Marketingzwecke 	<ul style="list-style-type: none"> Überblick über die wissenschaftliche Ausstrahlung einer Institution in ihrem Fachgebiet Erleichterung von Vergleichen zwischen Institutionen Stärken- und Schwächenanalysen und Einschätzung von Fachprofilen Einblicke in die zeitliche Entwicklung der wissenschaftlichen Performance von Fachbereichen Potentialanalyse für sichtbare, interdisziplinäre Schwerpunkte sowie für neue Forschungsfelder, Themen und Schnittstellen Identifikation von gelebten, aber auf Leistungsebene nicht erkannten (interdisziplinären) wissenschaftlichen Netzwerken Identifikation möglicher neuer Kooperationspartner und thematischer Bereiche 	<ul style="list-style-type: none"> Unzureichende Beurteilung interdisziplinärer Zusammenarbeit bei fachspezifischen Analysen Überbetonung von Forschungsleistungen bestimmter, gut messbarer Bereiche Unzureichende Beurteilung anderer schlechter messbarer Bereiche Vergleichbarkeit bibliometrischer Daten zwischen Disziplinen eingeschränkt Verstärkung der Macht von <i>High-Impact Journals</i>; Kostentreiber für die Bibliotheken Stärkere Orientierung an Leistungsvolumen/Forschungs-Output anstatt an Qualität

Systemebene (Hochschulsystem und Wissenschaftspolitik)		
Anwendungsbereiche	Chancen	Grenzen und Gefahren
<ul style="list-style-type: none"> Institutionelle Differenzierung des Hochschulsystems durch Rankings und Ratings im Bereich der Forschung Stärken-, Schwächen- und Potentialanalyse für eine differenzierte Standortentwicklung 	<ul style="list-style-type: none"> Potentialanalyse in neuen Wissenschaftsfeldern als Grundlage für die Wissenschaftspolitik und Fördermaßnahmen Überblick über Stärken und Schwächen einzelner Forschungseinrichtungen innerhalb des Systems Identifikation von überregionalen wissenschaftlichen Netzwerken Rascher Überblick über thematische Schwerpunktsetzungen Fächerspezifisch differenzierte Einschätzung der Leistungsfähigkeit und deren zeitlicher Entwicklung im Vergleich mit anderen, vergleichbaren Ländern 	<ul style="list-style-type: none"> Frage nach Vergleichbarkeit der Datenerhebung zwischen verschiedenen Institutionen Verengung wissenschaftlicher Leistung auf ihren messbaren Bereich kann risikoreiche Forschung mit zeitweilig weniger publicationsintensiven Zeiten und einer langfristigen Orientierung unterminieren Kultureller Bias in bestimmten Fachgebieten zu Ungunsten nicht englischsprachiger Zeitschriften Vorschnelle Entscheidung nationaler politischer Fördermittelgeber über „Ausdifferenzierung“ von Fächerprofilen statt dem Erlauben von Doppelungen zur Verstärkung der kritischen Masse

Tabelle 6: Zusammenfassung der Anwendungsbereiche, Chancen sowie Grenzen und Gefahren bibliometrischer Methoden auf unterschiedlichen Ebenen der Evaluierung (Individualebene, Institutioneller Ebene und Systemebene).

4 Rolle bibliometrischer Methoden

Anknüpfend an die Ausführungen zur Steuerung und Evaluierung von Forschung im Allgemeinen und den Grundlagen der Bibliometrie im Besonderen widmet sich das folgende Kapitel nun der Frage, welche Rolle bibliometrische Methoden an österreichischen Universitäten sowie in ausgewählten internationalen Beispielen einnehmen und welche Daten und Indikatoren tatsächlich erhoben werden. Dies ermöglicht ein Verständnis des gegenwärtigen Stellenwerts der Bibliometrie.

4.1 Umfrage zur Bibliometrie an österreichischen Universitäten (2010)

Einer Umfrage aus dem Jahr 2010 zufolge werden in Österreich bibliometrische Methoden seit 1992 angewandt, wobei eine signifikante Steigerung der Nutzung bibliometrischer Daten ab 2004 festgestellt wurde. In erster Linie machen die Forschenden selbst, hier vor allem jene aus den Bereichen Natur- und Ingenieurwissenschaften sowie Medizin und Wirtschaftswissenschaften, von bibliometrischen Daten und Dienstleistungen Gebrauch, gefolgt von Forschungsförderungsorganisationen und politischen Entscheidungsträgern.⁶⁴ Bibliometrische Indikatoren waren der Studie zufolge in allen Fachbereichen durchwegs bekannt, allen voran der Journal-Impact Faktor, und nachgeordnet *Citations per Paper* und der h-Index. Vor allem in der Medizin und den Naturwissenschaften wurden bereits damals bibliometrische Analysen als Messinstrument wissenschaftlicher Produktion für geeignet erachtet. Nichtsdestotrotz sprachen sich die meisten Befürworter für die Anwendung bibliometrischer Methoden lediglich als zusätzliches und nicht ausschließliches Instrument zur Leistungsbewertung aus. Alternative Messinstrumente, z.B. Umfang und Qualität der Lehrtätigkeit, Einwerbung von Drittmitteln, Teilnahme an Konferenzen, Buchbeiträge und allen voran Peer Review-Verfahren⁶⁵, sollten demnach in die Leistungsermittlung integriert werden. Bereits 2010 zeigt die Studie auf, dass die Berücksichtigung der fächerspezifischen Gegebenheiten als wichtig angesehen wurde: Mit der Begrün-

⁶⁴ Hasitzka, Katharina, Juan Gorraiz und Christian Gumpenberger, „Bibliometrie in Österreich: Ein neues Aufgabenfeld für Bibliotheken“, in *Universitätsbibliotheken im Fokus: Aufgaben und Perspektiven der Universitätsbibliotheken an öffentlichen Universitäten in Österreich*, hg. von Bruno Bauer, Christian Gumpenberger und Robert Schiller (Graz, 2013), S. 216–225.

⁶⁵ Vgl. dazu auch Österreichischer Wissenschaftsrat, 2014.

dung, die Indikatoren seien stark fachbereichsabhängig, lehnten es drei Viertel der Teilnehmerinnen und Teilnehmer der Umfrage (vor allem in den Rechtswissenschaften) ab, quantitative Indikatoren als Vergleichsparameter heranzuziehen. Die vorgebrachten Einwände deckten sich mit jenen, die in der internationalen Debatte um die Grenzen bibliometrischer Instrumente vorherrschend sind.⁶⁶

4.2 Umfrage zur Bibliometrie an österreichischen Universitäten (2019)

Der Österreichische Wissenschaftsrat führte im Februar 2019 eine an die Rektorate gerichtete Umfrage durch. An dieser beteiligten sich 14 Universitäten.⁶⁷ Zusätzlich wurden persönliche Gespräche und Interviews durchgeführt, um ein möglichst detailliertes Verständnis zur Praxis der Forschungsevaluierung an den einzelnen Standorten zu gewinnen. Dabei zeigte sich, dass die Arbeit mit metrischen Mess- und Analysemethoden mittlerweile fest im universitären Forschungsalltag integriert ist. Dennoch sind einige Kritikpunkte und Vorbehalte aus 2010 auch heute noch relevant. Darüber hinaus machen die Proliferation und fortschreitende Internationalisierung der wissenschaftlichen Publikationspraxis sowie die wachsende Bedeutung von Forschungssteuerung und des damit einhergehenden Evaluationsbedarfs die Bibliometrie zu einem Thema von zunehmender Bedeutung, auch im Hinblick auf deren Anwendbarkeit auf bestimmte Fachbereiche, wie etwa die Geistes-, Sozial- und Kulturwissenschaften.⁶⁸

Die Fragen⁶⁹ bzw. Antworten der Umfrage zur gegenwärtigen Anwendung bibliometrischer Methoden an österreichischen Universitäten lassen sich in vier Kategorien unterteilen, welche im Folgenden erörtert werden:

- Organisatorische Einbettung
- Forschungsrelevante Daten und Indikatoren
- Anwendungsfelder bibliometrischer Leistungsindikatoren
- Benchmarking und Austausch zwischen den Universitäten

⁶⁶ Hasitzka, Katharina, Juan Gorraiz und Christian Gumpenberger, 2013.

⁶⁷ Donau-Universität Krems, Universität für Bodenkultur, Medizinische Universität Graz, Medizinische Universität Innsbruck, Medizinische Universität Wien, Montanuniversität Leoben, Universität Mozarteum, Universität für Musik und darstellende Kunst Wien, Technische Universität Wien, Universität Graz, Universität Innsbruck, Universität Klagenfurt, Universität Wien, Wirtschaftsuniversität Wien.

⁶⁸ Bayer, Florian u. a., *Sichtbarkeitssteigerung in den Geistes-, Sozial- und Kulturwissenschaften (GSK): Ergebnisse einer Befragung an der Universität Wien* (Wien, 2017).

⁶⁹ Der für die Umfrage verwendete Fragebogen befindet sich im Anhang des Berichts.

Organisatorische Einbettung

Die Erfassung von Forschungsdaten und ihre Aufbereitung mit metrischen Mess- und Analysemethoden ist an vielen Universitäten im Organisationsbereich „Forschungsservice und -information“ verortet. An manchen Universitäten werden bibliometrische Daten auch im Rahmen des Leistungs- und Qualitätsmanagements untersucht. Des Öfteren befassen sich speziell dafür eingerichtete Einheiten der Bibliotheken mit dieser Thematik. So ist beispielsweise an der Bibliothek der Universität Wien eine eigene Abteilung für Bibliometrie und Publikationsstrategien eingerichtet, welche immer wieder wissenschaftliche Projekte (auch in Kooperation mit nationalen und internationalen Partnern) durchführt.

An den meisten Universitäten werden allerdings aktuell keine wissenschaftlichen Projekte zum engeren Thema Bibliometrie durchgeführt. Am Institut für Wissenschafts- und Technikforschung der Universität Wien ist das Messen von Forschungsleistung immer wieder Gegenstand von Forschungsprojekten. Es bestehen auch Kooperationen im Rahmen der Forschungsplattform <https://rri.univie.ac.at/>. An der Universität Graz befasst sich das Institut für Informationswissenschaft und Wirtschaftsinformatik mit dem Forschungsthema Infometrie und Szentometrie. Des Weiteren ist eine universitätsübergreifende Abstimmungsgruppe mit der Technischen Universität Graz (TU Graz) und der Medizinischen Universität Graz (Med Uni Graz) eingerichtet. Eine Arbeitsgruppe zum Thema Qualitätskriterien und Maßnahmen zur Sicherung der Erreichung von Qualitätskriterien in den einzelnen Wissenschaftsdisziplinen ist in Planung.

Forschungsrelevante Daten und Indikatoren

Universitäten benutzen Forschungsdaten zur Erstellung und zum Monitoring der Zielvereinbarungen zwischen Rektorat und einzelnen Fakultäten ebenso wie zur Erstellung der jährlich an das Bundesministerium für Bildung, Wissenschaft und Forschung zu erbringenden Wissensbilanz. Darüber hinaus werden die Daten teilweise auch für das universitätseigene Forschungsinformationssystem sowie zur forschungsbezogenen Leistungsbewertung erhoben. Zu den erhobenen Daten gehören unter anderem Publikationen, Vorträge, Organisation von Veranstaltungen, Drittmittelprojekte und Stipendien - gegliedert nach Förderinstitution - Funktionen in Zeitschriften, wissenschaftlichen Institutionen und Kommissionen, Gastwissenschaftlerinnen und Gastwissenschaftler, Gastaufenthalte der eigenen Forschenden, Grants, Patente und Gutachten. Welche Daten genau erhoben werden, variiert zwischen den Universitäten. Zur Anwendung kommen metrische Indikatoren wie Zitationen und daraus abgeleitete Indikatoren, Drittmittelindikatoren, JIF bzw. *Journal Citation Report* etc. Herangezogen werden die Datenbanken *Web of Science*, *Scopus* und *Google Scholar*. Einige Universitäten (z.B. TU Wien) geben an, dass met-

rische Indikatoren (insbesondere Zitationsindikatoren, h-Index) im Allgemeinen nur eine ergänzend-informative Rolle spielen und nicht standardisiert ausgewertet werden. Vereinzelt (z.B. an der Universität Graz) werden jedoch auch über übliche Indikatoren, wie h-Index und M-Quoten, hinausgehende Daten erfasst. Dazu zählen z.B. der Anteil der Top 10% bzw. 1% Publikationen (PP (top 10/1%)) im Forschungsfeld, der MNCS und der Anteil der Publikationen mit internationalen Co-Autoren.

Bibliometrische Auswertungen werden häufig im Rahmen von *Informed Peer Reviews* zur Evaluierung von Fakultäten, Berufungsprozessen und Laufbahnstellenverfahren sowie zur Analyse der Qualitätsentwicklung des Publikationsoutputs eingesetzt. In naturwissenschaftlich ausgerichteten Fakultäten werden sie teilweise auch zur Mittelverteilung auf Fakultätsebene herangezogen. Szientometrische Indikatoren (Wissenschaftspreise, Drittmittelprojekte, Rankingergebnisse) werden an der Universität Wien genutzt, um die Identifikation ihrer Stärkefelder zu unterstützen. An der TU Wien werden internationale Rankings, z.B. das *QS-Ranking* und das *THE-Ranking*, für das Monitoring der Universitätsentwicklung im internationalen Kontext herangezogen. Dies geschieht sowohl für die gesamte Institution, als auch heruntergebrochen auf fachspezifische Sub-Rankings für die einzelnen Fakultäten. Vor dem Hintergrund ihres breiten Fächerkanons wird Bibliometrie an der TU Wien, wie bei den meisten befragten Universitäten, lediglich als ein (kleinerer) Teil der Bewertungskriterien von Forschungsoutput gesehen.

Unter den antwortenden Universitäten herrscht Einvernehmen darüber, dass bibliometrische Methoden in bestimmten Fachbereichen (vor allem in den Geistes- und Sozialwissenschaften) nicht standardisiert und vor allem nicht in gleicher Weise wie z.B. in den Naturwissenschaften angewendet werden können und sollten.⁷⁰ Die Universität Wien kritisiert an dieser Stelle den zunehmenden Rückgriff auf Bibliometrie (z.B. in Rankings oder auch in der Wissenschaftspolitik), welcher oft vereinheitlicht und fachkulturell unreflektiert erfolgt, ohne Rücksichtnahme auf Datenlage, Publikations- und Zitationskultur der betrachteten Felder. Die meisten Universitäten pflegen Kataloge von Leistungsindikatoren, die unterschiedliche, an die jeweiligen Fachbereiche und den Gepflogenheiten der spezifischen Scientific Community entsprechend angepasste quantitative, aber auch qualitative Leistungsindikatoren, wie z.B. Tätigkeiten in Fachgesellschaften, Gutachtertätigkeiten für Fachzeitschriften etc. enthalten. Bibliometrische Variablen (Zitationen) spielen dabei eine unterschiedlich große Rolle.

An der TU Wien werden verpflichtend Daten zu Projekten und dem Publikationsoutput erhoben, während weitere Leistungen, wie beispielsweise Gremienaktivität, Gutachtertätigkeit, Beratung,

⁷⁰ Bayer, Florian u. a., 2017.

Impact in der Öffentlichkeit etc. optional in persönlichen Forschungsprofilen („Forschungsportfolio“) eingetragen werden können. Für Fächer, in denen übliche bibliometrische Indikatoren nicht oder nur schwer anzuwenden sind, wurden an der Universität Wien Listen qualitativ hochwertig eingeschätzter Zeitschriften sowie Verlage für Forschungsmonographien erstellt. Zudem gibt es Projekte, um das Prestige von (Buch)-Verlagen, z.B. über *Scholarly Publishers Indicators* (SPI)⁷¹, als akademischen Qualitätsstandard heranzuziehen.

Mit der Frage sinnvoller bibliometrischer Kennzahlen und Indikatoren für den internationalen Vergleich von Fachbereichen, die nicht traditionell bibliometrisch erfass- und bewertbar sind, beschäftigt sich auch die Universität Graz im Rahmen des Verbundprojekts EQA⁷² der Universitäten Siegen, Luzern, Riga und Graz.

Die Universität für Musik und darstellende Kunst Wien (mdw) erhebt mittels ihres Qualitätsmanagements künstlerische Publikationen (Druckwerke, Ton-, Bild- und Datenträger). Bibliometrische Auswertungen gibt es hier keine. Auch die Universität Mozarteum Salzburg unterstreicht, dass die bestehenden Konzepte der Leistungsevaluation für die Geistes-, Sozial- und Kulturwissenschaften und speziell für die Musikuniversitäten zu wenig passgenau seien. Es wird daher derzeit eine beobachtende Position eingenommen.

In manchen Fachbereichen, wie z.B. der Informatik, liegt die Schwierigkeit der Nutzung klassischer bibliometrischer Indikatoren darin, dass diese weitgehend auf Zeitschriftenpublikationen beschränkt sind und z.B. Konferenzbeiträge, welche im Bereich Informatik eine große Rolle spielen, nicht hinreichend erfasst werden. Auch die Montanuniversität Leoben (MUL) weist darauf hin, dass in den montanistischen Wissenschaften Forschungsergebnisse häufig in Form von Vorträgen und Konferenzbänden veröffentlicht werden. Diese Art der Veröffentlichung wird allerdings in den Forschungsinformationssystemen aller Universitäten erhoben, da sie ohnehin Teil der Wissensbilanz sind.

An der Wirtschaftsuniversität Wien (WU Wien) wird bei der Erfassung bibliometrischer Kennzahlen ein internes Journalranking herangezogen. Für die betriebswirtschaftlich ausgerichteten

⁷¹ Giménez-Toledo, Elea, Carlos Tejada-Artigas und Jorge Mañana-Rodríguez, „Scholarly Publishers Indicators (SPI)“ (2014) <<http://ilia.cchs.csic.es/SPI/indexEn.html>> [zugegriffen 4. September 2019].

⁷² Universität Siegen, „European Quality Audit“ (2019) <https://www.uni-siegen.de/start/die_universitaet/qualitaetsmanagement/european_quality_audit/?lang=de> [zugegriffen 4. September 2019].

Institute wurde weiters ein strengeres, spezifisch für diesen Fachbereich zugeschnittenes Qualitätsranking der wissenschaftlichen Fachzeitschriften (VHB-JOURQUAL 3⁷³) etabliert, das im Verband der Hochschullehrer für Betriebswirtschaft im deutschsprachigen Raum koordiniert wird. Daneben werden eine Reihe anderer Forschungsleistungen, wie z.B. Gesetzeskommentare, Urteilsbesprechungen in den juristischen Fächern, und weitere Indikatoren, wie die Durchführung von Veranstaltungen und die Förderung des wissenschaftlichen Nachwuchses, berücksichtigt.

Die Universität Innsbruck erhebt zweimal jährlich einen relativ breiten Katalog an Leistungsdaten, die auch mit internationalen Datenbanken (z.B. *Web of Science*) verlinkt und abgeglichen werden. In jenen Fachbereichen, für welche Publikationsmetriken nicht im entsprechenden Ausmaß zur Verfügung stehen, oder wo aufgrund unterschiedlicher Publikationskulturen keine wirkliche Vergleichbarkeit gewährleistet werden kann, wird behutsam mit entsprechenden Auswertungen verfahren. Als Beispiel wurden auch hier die Rechtswissenschaften genannt, für die zwar ein Grundgerüst an Zahlen (z.B. Anzahl der Publikationen) verwendet, jedoch kein Vergleich auf breiter Basis im (inter)nationalen Kontext gezogen wird.

An der Universität Klagenfurt können sich die einzelnen Institute seit 2012 an einem Klassifikationsprojekt beteiligen. Die darin angelegten Qualitätskriterien für die Forschungsleistungen orientieren sich an den Maßstäben der jeweiligen Forschungskultur. Zudem haben die Institute bzw. Organisationseinheiten spezifische Klassifikationsraster für ihre spezifische Forschungskultur entwickelt.

Die aktuellen Entwicklungen im Forschungsbereich Bibliometrie werden von der Mehrheit der österreichischen Universitäten aufmerksam verfolgt und die verwendeten Indikatoren gegebenfalls angepasst. Eine Notwendigkeit für eine Neu- oder Weiterentwicklung von Indikatorensets wird bei den meisten Universitäten nicht gesehen. Lediglich die TU Wien empfiehlt im Zuge der erkennbaren Transformation der Publikationskultur hin zu Open Access und der zunehmenden Bedeutung von Open Science eine Weiterentwicklung metrik-basierter Kennzahlen für die österreichischen Universitäten und im internationalen Kontext. Dies sei auch eine Chance für eine sinnvolle Abstimmung zwischen den Forschungsstätten. Die TU Wien verweist auch auf die Notwendigkeit, sich hinsichtlich der stetig verändernden Publikationskultur mit einer Überarbeitung bis hin zu einer Neudeinition der Leistungsmessung von Forschungsoutput

⁷³ VHB, „VHB-JOURQUAL3“ (2019) <<https://vhbonline.org/vhb4you/vhb-jourqual/vhb-jourqual-3>> [zugegriffen 10. Februar 2020].

im Sinne von *Responsible Metrics*⁷⁴ zu befassen. Die Universität Graz gibt an, universitätsinterne, fachspezifische Qualitätskriterien für die Forschung zu erarbeiten, welche künftig in die Beurteilung von Forschungsleistung einfließen sollen. Auch die Universität für Bodenkultur Wien (BOKU) entwickelt das Indikatorenset hinsichtlich neuer Auswertungsmöglichkeiten rund um das eigene Forschungsinformationssystem kontinuierlich weiter.

Alternative Metriken kommen zurzeit nur vereinzelt zur Anwendung. Die BOKU betont jedoch, dass Altmetrics, wenn auch aktuell nicht für die Leistungsbewertung berücksichtigt, als ein integraler Bestandteil im eigenen Forschungsinformationssystem gilt; so wird bei zitierten Publikationen der *Altmetric Donut* samt Details angezeigt. Die Medizinische Universität Innsbruck (MUI) hingegen steht Altmetrics eher skeptisch gegenüber, da angezweifelt wird, dass Soziale Medien als Informationsquellen im Kontext der Forschungsevaluation geeignet sind.

Interdisziplinäre Forschung erfordert besondere Aufmerksamkeit bei der Leistungsmessung und der Auswertung bibliometrischer Daten. So werden an der Universität Wien und der TU Wien im Rahmen eines regelmäßigen Forschungsmonitorings fächerübergreifende Forschungsfelder separat, z.B. im Hinblick auf gemeinsame Publikationen und Drittmitteleinwerbungen, ausgewertet. Somit kann interdisziplinäre Zusammenarbeit in einem gewissen Rahmen sichtbar gemacht werden. Ähnlich verfahren auch die Donau-Universität Krems (DUK), die Universität Klagenfurt und die BOKU. An der WU Wien wird interdisziplinäre Zusammenarbeit über die Anteile der „*Subject Areas*“ im Online Analyse-Werkzeug *Scival* erfasst. In ähnlicher Weise, allerdings mit Hilfe des Netzwerk-Analyseprogramms *Pajek* in Verbindung mit dem bibliometrischen Analysewerkzeug *BibExcel*, werden die Potentiale interdisziplinärer Zusammenarbeiten an der Medizinischen Universität Wien (MedUni Wien) analysiert.

Anwendungsfelder bibliometrischer Leistungsindikatoren

Bibliometrische Auswertungen werden an vielen Universitäten zur Unterstützung des Rektorats im Zuge von internen Evaluations- und auch Berufungsprozessen herangezogen. Auch hier zeigen die Antworten, dass die Nutzung metrischer Leistungsindikatoren bei den meisten Universitäten im Hinblick auf die unterschiedlichen Fächerkulturen sehr stark differenziert erfolgt.

Dies betrifft etwa die Unterstützung von Berufungsverfahren durch die Bereitstellung bibliometrischer Daten aus der universitätsinternen Forschungsdatenbank. An fast allen der antwortenden Universitäten werden für Berufungen bibliometrische Analysen, meist auf Anfrage,

⁷⁴ Siehe Kapitel 3.3.

durchgeführt, sofern der betroffene Fachbereich dafür geeignet ist und die verfügbaren Daten tatsächlich eine sinnvolle und akzeptierte Datenquelle darstellen.

An der Universität Graz werden in Zielvereinbarungen mit neuberufenen Professorinnen und Professoren, je nach Ausrichtung und fachlicher Publikationskultur, auch Zielwerte für Publikationen festgelegt. Beim Abschluss der Qualifizierungsvereinbarungen ist die Erfüllung der Qualifizierungsziele an vielen Universitäten (z.B. TU Wien, WU Wien, Universität Wien, Universität Innsbruck) an quantitative Indikatoren (z.B. Zahl der Q1-Journalartikel, Monographien bzw. Buchbeiträge in renommierten Verlagen sowie Einwerbung von Drittmittelprojekten) geknüpft. Generell werden bibliometrische Daten auf der Ebene der einzelnen Wissenschaftlerin bzw. des einzelnen Wissenschaftlers allerdings immer nur als Zusatzinformation neben anderen Leistungsdimensionen (z.B. Drittmittelprojekten), persönlichen Gesprächen bzw. Stellungnahmen von Gutachtern und nicht als prioritäres Kriterium genutzt.

Die Mittelverteilung innerhalb der Fakultäten erfolgt im Allgemeinen nach eigenen, in den Fakultäten selbst entwickelten Schlüsseln, die, je nach Fachbereich differenziert, vielfach auch metrische Leistungsindikatoren einbeziehen (z.B. an den Universitäten Graz, Wien und Innsbruck, der TU Wien und der DUK).

Die internen Qualitätskontrollen sehen regelmäßige Zielvereinbarungen mit den Fakultäten vor, die grundsätzlich sowohl auf quantitativen wie auch qualitativen Indikatoren beruhen. Bibliometrische Indikatoren werden dabei nur teilweise direkt genutzt (z.B. WU Wien), an der Universität Graz werden sie den Fakultäten im Rahmen des Monitorings der Zielvereinbarungen zur Verfügung gestellt. Im Rahmen von *Informed Peer Review*-Verfahren bei der regelmäßig erfolgenden Evaluierung der Fakultäten an den Universitäten Innsbruck und Wien werden den externen Gutachtern neben dem Selbstbericht und den Zukunftsperspektiven der Fakultät auch Auszüge aus der Forschungsleistungsdatenbank aus den zurückliegenden Jahren zur Verfügung gestellt.

Die Medizinischen Universitäten Graz, Innsbruck und Wien geben an, metrische Leistungsindikatoren in allen Bereichen (Berufungen, Vergabe von Qualifizierungsstellen, im Rahmen des internen Qualitätsmanagements sowie bei der leistungsorientierten Mittelvergabe) einzusetzen. Die Med Uni Graz nutzt für ihre Berufungen eine multifaktorielle, bibliometrische Analyse und für alle weiteren Bereiche standardmäßig bibliometrische Daten aus dem internen Forschungsportal. Bei der Verteilung des Sach- und Investitionsmittelbudgets des wissenschaftlichen Bereichs gehen an der Med Uni Graz bibliometrische Indikatoren mit ca. 40% in die Leistungsparameter ein, die restlichen Kriterien sind eingeworbene Forschungsförderungsmittel (40%) und Leistungen in der Lehre (20%).

An den teilnehmenden Kunstuiversitäten spielt die Bibliometrie im engeren Sinne eine untergeordnete Rolle. Das Mozarteum setzt aktuell vorrangig auf andere Leistungsindikatoren wie z.B. *Peer Review* und Interviews, wobei die mdw angibt, Publikationen als Teil der Leistungserfassung für die personenbezogene Evaluierung heranzuziehen.

Benchmarking und Austausch zwischen den Universitäten

Im Rahmen des österreichischen Forschungsinformationssystem-Netzwerks FIS/CRIS wird ein regelmäßiger Austausch zwischen den an den Universitäten für die Forschungsdokumentation zuständigen Expertinnen und Experten zu übergreifenden Fragestellungen zum Umgang mit bibliometrischen Daten, zu Open Access Publikationen aber auch *Predatory Journals* etc. gepflegt. Die Universität Wien, die MedUni Wien und die TU Wien sind zudem auch im internationalen Netzwerk EuroCRIS⁷⁵ engagiert. Die Grazer Universitäten (Med Uni Graz, Universität Graz und TU Graz) haben das lokale Netzwerk „Erbsenzähler-AG“ zum Thema Bibliometrie ins Leben gerufen. Eine ähnliche Kooperation haben auch die BOKU, die Med Uni Graz und die Veterinärmedizinische Universität Wien aufgebaut, um das FIS gemeinsam weiterzuentwickeln.

Die Universität Wien betreibt die AG Szentometrie, ist zudem in der uniko-AG „Hochschulrankings“ aktiv, und pflegt über die Abteilung für Bibliometrie und Publikationsstrategien ein ausgedehntes internationales Netzwerk (Deutschland, Belgien, Spanien, UK). Das universitäre Forschungsinformationssystem u:cris war an der Gründung von FIS/CRIS Austria beteiligt.

Die BOKU schlägt vor dem Hintergrund eines zukünftigen verstärkten Einsatzes von bibliometrischen Indikatoren (z.B. im Zuge von ausgewählten Evaluationsverfahren oder strategischen Planungen) zur Abstimmung betreffend Datenerhebung, Aufbereitung, Auswertung, Interpretation etc. eine österreichweite, universitätsübergreifende Arbeitsgruppe vor.

Aktuell ist ein Benchmarking im Sinne einer metrik-basierten vergleichenden Performance-Evaluation mit ähnlich ausgerichteten nationalen und internationalen Fachbereichen an den meisten österreichischen Universitäten kein Thema. Die Universität Graz allerdings plant Benchmarks mit internationalen Institutionen im Rahmen des CWTS-Monitors⁷⁶. An der WU Wien

⁷⁵ euroCRIS, <<https://www.eurocris.org/>> [zugegriffen 4. September 2019].

⁷⁶ Die Universität Graz hat ein Evaluierungstool von CWTS lizenziert, in dem bibliometrische Auswertungen für die Universität auf Ebene der Forschenden und der organisatorischen Einheiten zur Verfügung stehen.

kommen *Key Performance Indicators* (KPI) zum Einsatz, die im Rahmen der internen Qualitätskontrolle herangezogen werden, und auch einen Vergleich mit ähnlich ausgerichteten internationalen Forschungsorganisationen sowie, in Teilbereichen, auch mit nationalen Organisationen enthalten. Im Rahmen der Evaluierung von Forschungsschwerpunkten setzt die MedUni Wien *InCites*-Analysen⁷⁷ für das Benchmarking mit ähnlich ausgerichteten nationalen und internationalen Fachbereichen ein. An der Med Uni Graz wird die Integration von normierten Zitierdaten (NIH ICite⁷⁸) als vereinfachte Form des internationalen Daten-Benchmarkings gesehen. Es wird an dieser Stelle betont, dass ein Benchmarking anhand von Prozessen (gegenüber einem Benchmarking anhand von Indikatoren) als wesentlich zielführender erachtet wird.

Die Überlegung, eine einheitliche Datenerhebung auf nationaler Ebene mittels einer gemeinsamen (nationalen) Forschungsdokumentation zur Vereinfachung der Datenerhebung (z.B. nach dem norwegischen Vorbild CRISTIN⁷⁹) zu erstellen, sehen die antwortenden Universitäten insgesamt eher kritisch. Es wird mehrheitlich darauf hingewiesen, dass durch die Wissensbilanz-Verordnung bereits Vorgaben zur Erhebung und Klassifizierung von Publikationen an österreichischen Universitäten zu erfüllen sind. Zudem ist eine Datenerhebung rein zu Dokumentationszwecken nicht zielführend. Derzeit haben die Universitäten jeweils ihr eigenes Erfassungssystem, das, abgestimmt auf die spezifischen Bedürfnisse der Disziplinen und mit zusätzlichen Funktionen ausgestattet, weit über eine Forschungsdokumentation hinausgeht. Die Erfassungssysteme beruhen bei einigen Universitäten (z.B. Universität Innsbruck) auf Eigenentwicklungen, andere (Universität Wien, TU Graz) verwenden kommerzielle Systeme (z.B. PURE⁸⁰). Aufgrund der unterschiedlichen strategischen Ausrichtung haben auch nicht alle Universitäten die gleichen bibliometrischen Datenbanken lizenziert, was einen Abgleich erschwert. Eine gemeinsame Datenhaltung hätte am besten bereits mit der Implementierung des UG 2002 starten sollen, inzwischen ist jedoch zu bezweifeln, dass ein einheitliches Globalsystem die spezifischen Bedürfnisse der einzelnen Universitäten allumfassend abdecken kann. Die Universität Wien hebt hervor, dass nicht die gemeinsame Architektur das Ziel sein sollte, sondern dass die Systeme unabhängig von der technologischen Plattform einen gemeinsamen und akzeptierten Datenstand erfüllen sollten. Es wird für ein gemeinsames Austauschformat im Sinne eines Da-

⁷⁷ Clarivate Analytics, „InCites: An Objective Analysis of People, Programs and Peers“ (2019) <<https://clarivate.com/webofsciencegroup/solutions/incites/>> [zugegriffen 4. September 2019].

⁷⁸ National Institute of Health, „iCite“ <<https://icite.od.nih.gov/>> [zugegriffen 6. September 2019].

⁷⁹ CRISTIN, <<https://www.cristin.no/english/>> [zugegriffen 4. September 2019].

⁸⁰ Elsevier, „Pure“ (2020) <<https://www.elsevier.com/solutions/pure>> [zugegriffen 23. Jänner 2020].

tenstandards nach den Modellen von Großbritannien, den Niederlanden oder Deutschland plädiert. Die MUI steht einer nationalen Datenerhebung kritisch gegenüber, betont jedoch die Sinnhaftigkeit eines zentral organisierten und auch finanzierten Zugangs zu den kommerziellen bibliometrischen Datenbanken für alle Universitäten.

Ein Beispiel für eine systemübergreifende Vernetzung ist die Initiative des FWF, den Output von geförderten Projekten zu evaluieren. Hierzu wird derzeit an einer Verlinkung der universitären Forschungsinformationssysteme mit der zentralen Datenbank des FWF gearbeitet, um die aus geförderten Projekten hervorgegangenen Publikationen zu erfassen. Dies erfordert allerdings Ergänzungen der eingegebenen Publikationsdaten um Informationen zu Fördergeber, Projektnummern etc.

4.3 Ausgewählte internationale Beispiele

Im Folgenden werden Fragen der Forschungsevaluation und die Rolle bibliometrischer Methoden im internationalen Kontext anhand einiger ausgewählter Beispiele erläutert. Grundlage hierfür sind Gespräche mit Vertreterinnen und Vertretern der Universitäten Hamburg und Heidelberg (Deutschland), der Maastricht University (Niederlande), der Universität Zürich (Schweiz), der National University Singapore (NUS), der Tongji University (China) und der UC Berkeley (USA) über die jeweiligen Strategien der Forschungssteuerung auf drei Ebenen: der Individualebene mit dem Fokus auf individuellen Wissenschaftlerinnen und Wissenschaftlern, der institutionellen Ebene, die Fakultäten, Departments, Cluster bzw. Fächergruppen etc. umfasst, und der Systemebene, die übergeordnet das Hochschulsystem und die Wissenschaftspolitik in den Blick nimmt.

Individualebene

An der Universität Hamburg gelten Berufungen als der Kern einer erfolgreichen Universitätsleitung und unterliegen demgemäß einer besonderen Aufmerksamkeit von Seiten der Hochschulleitung. Die Nutzung bibliometrischer Daten in Berufungsverfahren wird disziplinspezifisch gehandhabt. So nehmen bibliometrische Indikatoren bei Berufungsverfahren z.B. in der Physik eine bedeutsamere Rolle ein als in den Geisteswissenschaften, in welchen vielmehr monographische Schlüsselwerke als Referenz für eine Person gelten. Als für die Berufungsqualität ganz wesentliche Verfahrenselemente werden darüber hinaus die direkte Ansprache von Kandidatinnen und Kandidaten durch die Fachbereiche, die Beteiligung zweier direkt vom Präsidenten ausgewählter Gutachterinnen bzw. Gutachter sowie die Einbindung externer *Assessment Center* genannt. Bei der Gewichtung bibliometrischer Daten kommen sowohl der Journal-Impact Faktor

als auch personenbezogene Indikatoren zum Tragen. Dabei werden fünf Aufsätze aus internationalen Journals mit strikter Peer Review-Qualitätssicherung zur Bewertung herangezogen; Konferenzbeiträge werden als weniger bedeutsam gewichtet.

Ähnlich verhält es sich an der Universität Heidelberg, an der ebenfalls disziplinspezifische Maßstäbe bei der Forschungsevaluation angelegt werden. Im Bereich der Medizin wird stark auf Impact-Parameter gesetzt; auch in den Lebenswissenschaften haben *High-Impact*-Publikationen in den jeweiligen Top Journals zentrale Bedeutung. In den Geisteswissenschaften wiederum wird monografischen Schlüsselwerken die größte Bedeutung zugewiesen. Bei der Rekrutierung von Professorinnen und Professoren in Heidelberg werden bibliometrisch hinterlegte Kriterien als notwendig, jedoch nicht hinreichend erachtet; vielmehr wird auch dem internationalen Ruf der Kandidatinnen und der Kandidaten besondere Beachtung geschenkt. 2/3 bis 3/4 der Berufungen kommen aus dem Ausland, teilweise auch als Rückkehrer. Nach Ansicht der Universitätsleitung sind gute Forschungsbedingungen die entscheidende Grundlage für die Rekrutierung von internationalen Spitzenvissenschaftlerinnen und Spitzenvorschern.

Auch an der University of California in Berkeley sind die Zahl der Publikationen oder der h-Faktor der Bewerberinnen und Bewerber nur eines von mehreren Kriterien, wobei allerdings ein Mindestmaß an Veröffentlichungen vorausgesetzt wird. Wichtiger als die Zahl der Publikationen sind Innovationspotential und Zukunftsideen. In einem äußerst aufwendigen Tenure-Track-Verfahren, in das sieben bis zehn externe, hochrangige Gutachterinnen und Gutachter eingebunden werden, werden ausgewählte Publikationen detailliert evaluiert und im Hinblick auf die wissenschaftliche Innovation bewertet.

Die Forschungsleistung der einzelnen Wissenschaftlerinnen und Wissenschaftler wird an der Maastricht University anhand der Publikationsleistung sowie weiterer Leistungen (z.B. Patente, Mitgliedschaften in Gutachtergremien etc.) bewertet. Die Wertigkeit der Publikationen richtet sich dabei auch nach dem Impact Faktor der Zeitschriften, in denen der Wissenschaftler veröffentlicht hat. Jedoch werden auch an der Maastricht University fächerspezifische Unterschiede berücksichtigt. In Fachbereichen wie den Rechts- und Sozialwissenschaften wird z.B. deutlich mehr Wert auf die Veröffentlichung von Monographien gelegt, wobei hier wiederum die Qualität des Verlages Berücksichtigung findet.

Die Universität Zürich ist Mitunterzeichnerin der San Francisco Vereinbarung über die Forschungsbewertung (DORA).⁸¹ Die Bibliometrie wird an der Universität Zürich stark nach Fächerkultur, selbst innerhalb von Fakultäten, ausdifferenziert angewendet. So wird beispielsweise in der Volkswirtschaftslehre im Zuge von Berufungsverfahren⁸² mindestens eine Publikation in einem A+-Journal (aus ca. vier dafür in Frage kommenden Journals) erwartet, die Betriebswirtschaftslehre stützt sich auf das bereits erwähnte VHB-JOURQUAL 3. In der Mathematisch-Naturwissenschaftlichen Fakultät werden wiederum die drei wichtigsten Veröffentlichungen der Bewerberinnen und Bewerber eingehend qualitativ beurteilt. Personenbezogene Indikatoren, wie der h-Index, werden in manchen Fakultäten ergänzend herangezogen.

Eine verhältnismäßig wichtige Rolle spielen bibliometrische Bewertungsmethoden in Berufungsverfahren an der Tongji University, Shanghai, sowie an der National University of Singapore (NUS). Die Kriterien zur Evaluation der Forscherinnen und Forscher im Rahmen von Berufungs- bzw. Tenure-Track-Prozessen an der Tongji University werden dabei laufend angepasst. Derzeit müssen fünf Artikel aus den vergangenen drei Jahren vorgelegt werden, die gemäß dem *Scientific Journal Ranking* (JCR) als Q1 gereiht sind. Neben bibliometrischen Daten wird jedoch die Einschätzung der internationalen Scientific Community immer bedeutsamer. An der NUS werden Bewerberinnen und Bewerber primär über das Tenure-Track-Verfahren rekrutiert. Bei der Vorselektion der Kandidatinnen und Kandidaten werden bibliometrische Daten erfasst und bewertet, ohne dabei disziplinspezifische Kriterien außer Acht zu lassen. Großes Augenmerk wird darauf gelegt, in welchen Journals publiziert, wie viele Keynote-Lectures gehalten wurden, oder welche wissenschaftlichen Auszeichnungen vorliegen. Im Rahmen einer externen Evaluation werden, geleitet von dem Ziel „*Find the best in the world*“, bibliometrische Informationen (Publikationen, Impact- und h-Faktoren) durch ein aufwändiges Peer Review-Verfahren ergänzt. In diesem zweiten Verfahrensschritt spielen neben der Bewertung der internationalen Reputation in der Scientific Community auch weitere Aspekte, wie z.B. Veröffentlichungen von internationalen Standardwerken, Patente, Mitgliedschaften in Scientific Boards oder das Transferpotential der Forschung, eine wichtige Rolle.

⁸¹ DORA wurde von 2013 Institutionen und 16.311 Einzelpersonen unterzeichnet. Siehe <https://sf-dora.org/signers/> [zugegriffen 17. August 2020].

⁸² Die Wirtschaftswissenschaftliche Fakultät hat ein Reglement hinsichtlich des Tenure-Verfahrens erlassen. Dieses besagt, es sollen bei einer Verlängerung der Anstellung und der Beförderung grundsätzlich die gleichen Kriterien wie bei einem ordentlichen Berufungsverfahren angewendet werden.

Institutionelle Ebene

Evaluation auf der Ebene der Fakultäten bzw. Departments erfolgt in allen hier beispielhaft adressierten Universitäten durch externe Peers. An der NUS wird darüber hinaus regelmäßig eine Selbst-Evaluation der Departments durch sorgfältige, vergleichende bibliometrische Analysen mit ähnlich ausgerichteten Departments auf internationaler Ebene durchgeführt, um die internen Prozesse der Departments zu verbessern. Dafür werden fünf *Peer Departments*, die als gleichwertig erachtet werden, sowie drei *Target Departments*, die als Vorbild und Maßstab für die zukünftige Entwicklung dienen, definiert. Beispielsweise kann die Frage geklärt werden, warum bei gleicher Anzahl an Publikationen ein wesentlich geringerer Impact Faktor gegenüber dem *Peer Department* erzielt wird. Dabei werden auch die Leistungen der Professorinnen und Professoren verglichen, um vor allem Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftlern Maßstäbe für die eigene Forschungsentwicklung zur Hand zu geben. Ziel dieser vergleichenden Analyse ist es, die Leitkultur des Departments weiterzuentwickeln, um in einem Zeitraum von ungefähr fünf Jahren das Niveau der Vorbilder zu erreichen.

An der Tongji University wird bei der Evaluation auf Ebene der Departments auf ähnliche Kriterien wie im Berufungsverfahren zurückgegriffen. Zusätzlich werden in stark wettbewerbsorientierten Programmen (vergleichbar mit Exzellenzprogrammen) bibliometrische Daten herangezogen und internationale Rankings (*Shanghai Ranking*, *THE-Ranking* etc.) berücksichtigt.

An der Universität Hamburg erfolgt die Begutachtung auf Fakultätsebene über externe Evaluation, bei der die Leistungsvereinbarungen der Universität mit dem Ministerium (ein Katalog von ca. 30 Kennzahlen, welcher das gesamte Leistungsspektrum umfasst) auf die einzelnen Fakultäten heruntergebrochen werden. An der Universität Heidelberg wiederum wird die Frage nach der Evaluation von Fakultäten als vergleichsweise nachrangig eingestuft. Eine grobe Orientierung über die Leistungsfähigkeit der Fakultäten bieten Rankings, wie z.B. der Deutschen Forschungsgesellschaft (DFG). Stärker im Vordergrund steht dagegen, inwiefern sich an den Fakultäten kreative Forschungsansätze und Zukunftsideen entwickelt haben oder entwickeln können.

An der Maastricht University wird die Forschung typischerweise auf der Ebene von größeren Forschungseinheiten wie Fakultäten, Forschungsverbünden etc. bewertet, die in der Praxis aus mehreren Departments bestehen. Auf nationaler Ebene regelt das sogenannte *Standard Evaluation Protocol* (SEP) die Bewertung von Forschungseinheiten. Darin ist zwar festgelegt, dass die Qualität der Forschung einer Einheit bewertet werden *muss*, das SEP schreibt jedoch nicht vor, *wie* dies zu erfolgen hat. Die Forschungseinheiten können selbst geeignete Indikatoren auswählen. Typische Indikatoren, die von den Forschungseinheiten verwendet werden, sind die Anzahl der (peer reviewed) Publikationen, die durchschnittlichen Impact Faktoren der Zeitschriften, in

denen diese Publikationen veröffentlicht wurden, die Anzahl oder der Anteil der Publikationen in den Top 1%, Top 10% (manchmal sogar Top 25%) Publikationen. Zuweilen wird auch der h-Index der *Principal Investigators* berücksichtigt. In zunehmendem Maße werden jedoch auch qualitative Faktoren in die Bewertung der Forschungsleistung einbezogen.

Seit dem Jahr 2018 werden an der Universität Zürich die Bewertungskriterien in einer Evaluationsvereinbarung festgehalten. Die Kriterien orientieren sich dabei an den strategischen Grundsätzen der Universitätsleitung und der Fakultätsleitungen. Zwischen Drittmitteln, Publikationen oder anderen Leistungen im Prozess zur Ermittlung des Forschungsoutputs gibt es keine Gewichtung. Die Beurteilung der Leistungen erfolgt durch eine externe Expertengruppe in Form eines *Informed Peer Reviews*. Bei bibliometrischen Analysen werden verschiedene normalisierte Indikatoren (im Hinblick auf Forschungsdisziplin und Publikationsalter) herangezogen.

Systemebene

Im Auftrag des Präsidiums der Universität Hamburg wurde der Standort Hamburg nach Kriterien des deutschen Wissenschaftsrates extern begutachtet. Hierbei nahmen bibliometrische Daten (Publikationen, Zitationsindex etc.) eine wichtige Rolle ein, die je nach Fächerkultur adaptiert wurden. Auf Grundlage dieser Evaluation wurde ein Struktur- und Entwicklungsplan erstellt, der sich in den Ziel- und Leistungsvereinbarungen widerspiegelt.

Auch bei der Evaluation der NUS durch das „*Quality Assurance Framework for Universities*“ spielt die Bibliometrie eine bedeutsame Rolle, ist jedoch nur ein Kriterium unter vielen. Dabei geht es auch darum, den „Wert“ der Universität über die Wissenschaft hinaus für die Gesellschaft, die Wirtschaft und die Industrie zu erfassen.

Die Evaluation der Tongji University beruht auf der Begutachtung von Departments. Das zuständige Ministerium beurteilt die Universität auf Grundlage der Leistungsdaten, die von den Departments zur Verfügung gestellt werden; die daraus resultierende Mittelzuwendung basiert auf diesen Daten. Demgemäß kann die finanzielle Unterstützung für die Departments sehr unterschiedlich ausfallen. Diese Unterschiede müssen dann auf Universitätsebene strategisch gemanagt werden.

Qualitätsvereinbarungen zwischen dem Ministerium und den Universitäten in den Niederlanden – so auch der Universität Maastricht – sehen vor, dass die Universitäten über ein sogenanntes *Sektor-Dashboard* (wird derzeit vom Verband der niederländischen Universitäten (VSNU) entwickelt) transparente Informationen über ihre Kernaufgaben Bildung, Forschung und Impact lie-

fern. Welche Indikatoren im Forschungsbereich genutzt werden, wird derzeit untersucht. Implizit werden dabei auch internationale Rankings (THE, ARWU (*Shanghai-Ranking*), QS) als Indikatoren für Forschungsleistung herangezogen.

5 Möglichkeiten bibliometrischer Analysen wissenschaftlicher Disziplinen

Um die Möglichkeiten moderner bibliometrischer Methoden zu illustrieren, werden in diesem Kapitel drei bibliometrische Studien zusammengefasst. Der Fokus ist auf die methodischen Details zweier Studien zur Situation der Bildungsforschung und der Informatik in Österreich gerichtet. Neben Auswertungen der zentralen bibliometrischen Indikatoren (*key performance indicators, KPI*) im Ländervergleich und im Kontext der österreichischen Universitäten, werden auch weitere Analysemethoden, wie Cluster-, *Strength, Potential and Risk* (SPR)-Analysen, bibliometrische Visualisierungsmethoden sowie zeitliche Trendanalysen exemplarisch dargestellt. Vorher soll jedoch die Aufmerksamkeit auf die erste bibliometrische Studie des Österreichischen Wissenschaftsrats (ÖWR) im Bereich der Klinischen Forschung in Österreich gerichtet werden.

Das Kapitel repräsentiert zu einem gewissen Grad die sich immer weiter vertiefende Auseinandersetzung des ÖWR mit der Thematik „Bibliometrie“, aber auch der sich ständig weiterentwickelnden Methodik. Tabelle 7 bietet einen Überblick zu den Grundlagen der jeweiligen Fallbeispiele; sie verdeutlicht ebenfalls die eben angesprochenen Weiterentwicklungen. So nahm die Bedeutung der Bibliometrie bei jedem Fallbeispiel zu. Die Datenbasis bzw. die Zahl der Analyseinheiten wurde ausgeweitet. Indikatoren wurden gezielter eingesetzt, und die Studien wurden durch neue bibliometrische Methoden bereichert.

	Klinische Forschung	Bildungsforschung	Informatik
Durchführung der Analyse	2016	2018	2019
Kontext	Bibliometrie macht nur einen kleinen Aspekt aus, zusätzlich zu einer quantitativen Erhebung von Leistungskennzahlen, Interviews, einer Onlinebefragung und Begehung.	Bibliometrie nahm eine zentrale Rolle ein, zusätzlich zu Interviews und einer Onlinebefragung.	Bibliometrie rückt in den Vordergrund. Feedbackscheiben durch Expertinnen und Experten sowie von Teilnehmerinnen und Teilnehmern der ÖWR-Tagung.
Ländervergleich	9 Vergleichsländer: Dänemark, Deutschland, Frankreich, Großbritannien, Niederlande, Österreich, Schweden, Schweiz und USA.	10 Vergleichsländer: Belgien, Dänemark, Deutschland, Finnland, Großbritannien, Niederlande, Norwegen, Österreich, Schweden und die Schweiz.	11 Vergleichsländer: Belgien, Dänemark, Deutschland, Finnland, Großbritannien, Israel, die Niederlande, Norwegen, Österreich, Schweden und die Schweiz.
Datenbasis	WoS Core Collection – Zeitschriftenartikel und Reviews zwischen 2010–2014.	WoS Core Collection – Zeitschriftenartikel und Reviews zwischen 2000–2016 und deren Zitate bis Ende 2017.	<ol style="list-style-type: none"> 1. WoS Core Collection – Zeitschriftenartikel und Reviews zwischen 2000–2017 und deren Zitate bis Ende 2018. 2. WoS Conference Proceedings Citations Index - Analysezeitraum 2014–2017.
Einschränkung des Feldes	Selbstzitate wurden bei allen Studien entfernt und die Publikationsdatensätze mit dem CWTS-Kategorisierungssystem abgestimmt.	Feldabgrenzung basierend auf 4 WoS subject categories: <i>education & educational research, education, scientific disciplines, education, special sowie psychology, educational.</i>	<ol style="list-style-type: none"> 1. Feldabgrenzung in 2 Schritten: <ul style="list-style-type: none"> – 7 WoS subject categories: <i>artificial intelligence, cybernetics, hardware & architecture, information systems, interdisciplinary applications, software engineering sowie theory & methods.</i> 2. Clusteranalyse <ul style="list-style-type: none"> – Beiträge der wichtigsten Konferenzen der 26 Unterbereiche der Informatik basierend auf www.csrankings.org.

	Klinische Forschung	Bildungsforschung	Informatik
Bibliometrische Indikatoren	Normierte und nicht normierte KPI: P, MCS, MNCS, TCS, TNCS, P(uncited), PP(uncited), P(top 1% und 10%), PP(top 1% und 10%).	Fokus auf normierte KPI: P, MNCS, MNJS, PP(top 10%).	<ol style="list-style-type: none"> 1. Fokus auf normierte KPI: P, MNCS, MNJS, PP(top 10%), PP(collab), PP(int. collab), PP(industry). 2. Nicht normierte Indikatoren aufgrund von Datenlimitierung: P, MCS
Fokus der österreichischen Analyse	Alle 3 staatlichen Medizinischen Universitäten Weitere bibliometrische Methoden	5 aktivste Universitäten Zeitliche Entwicklung, <i>term maps</i> aller Vergleichsländer sowie der Einzelländer	<ol style="list-style-type: none"> 1. Cluster Analyse, SPR-Analyse für 5 österreichische Universitäten (Fokus auf <i>strength areas</i> und zeitliche Entwicklung). 2. <i>Term maps</i> aller Vergleichsländer, durchschnittliches Publikationsjahr und Impact der häufigsten Begriffe (alle Vergleichsländer) und der Einzelländer.

Tabelle 7: Übersicht über die bibliometrischen Fallbeispiele zur Klinischen Forschung, Bildungsforschung und Informatik in Österreich

5.1 Klinische Forschung

In der Stellungnahme und den Empfehlungen zur Klinischen Forschung in Österreich hat sich der ÖWR in einer ersten stichprobenhaften Analyse bereits 2016 zum Stellenwert der Bibliometrie (neben anderen Messmethoden) in der klinischen Medizin geäußert. Für eine detaillierte Darstellung der bibliometrischen Analysen wird auf diesen Bericht verwiesen. Die Einbeziehung bibliometrischer Daten in einem breiteren Spektrum von Qualitätsindikatoren war in der klinischen Medizin in anderen Ländern zu diesem Zeitpunkt seit langem gebräuchlich. Sie wird z.B. in den Niederlanden vom CWTS-Institut der Universität Leiden regelmäßig und systematisch analysiert, wobei alle acht niederländischen Universitätsklinika miteinander verglichen werden. Diese Daten werden allen Leitungsorganen der universitätsmedizinischen Einrichtungen zur Verfügung gestellt und dienen deren eigenen strategischen Planungen, z.B. im Rahmen der Schwerpunktbildung und der Definition von kritischer Masse. Die bibliometrische Analyse der Klinischen Forschung in Österreich im Jahre 2016 hatte im Gegensatz zu den ausführlichen Untersuchungen in den Niederlanden einen orientierenden Charakter, um dieses Instrument erstmalig in der ganzen Breite in der österreichischen Universitätsmedizin anzuwenden.

Für diese Zwecke beauftragte der ÖWR das CWTS, bibliometrische Daten aus 28 Fächergruppen für folgende Länder zu erheben und zu analysieren: Dänemark, Deutschland, Frankreich, Großbritannien, Niederlande, Österreich, Schweden, Schweiz und USA. Die verwendeten bibliometrischen Daten stammten aus der WoS Core Collection und schlossen Artikel und Reviews in Fachzeitschriften mit ein. Die gemessenen Indikatoren sowie die dazugehörende Datenbank wurden den Medizinischen Universitäten Graz, Innsbruck und Wien zur weiteren eigenen Verwendung zur Verfügung gestellt.

Die Auswertung dieser Studie zeigte für alle Medizinischen Universitäten wissenschaftliche Schwerpunkte im internationalen Spitzengeschehen. Dennoch ergaben sich aus der Diskussion der Studienergebnisse in einem Workshop mit Vertreterinnen und Vertretern der Universitätsmedizin auch Fragen und Kritik, vor allem im Hinblick auf die adäquate Zuordnung von Publikationen und deren Impact zu klinischen Fachabteilungen und Subdisziplinen. Auch zeigten sich Diskrepanzen zu wissenschaftlichen Leistungsanalysen mittels anderer Methoden. Aus den Diskussionen wurde deutlich, dass zu einer vergleichenden Analyse der Forschungsleistung der österreichischen klinischen Medizin weitere, tiefergehende Untersuchungen sowie eine bessere Zuordnung von Wissenschaftsgebieten zu den Organisationseinheiten in der Klinik (z.B. bis auf das Abteilungsniveau) erforderlich sind. Zum Beispiel kann eine Klinik inadäquat bewertet werden, wenn die institutionelle Zuordnung, z.B. zum Bereich der Gastroenterologie, aber der wissenschaftliche Publikations-Output einer anderen Disziplin (z.B. Onkologie) zugeordnet wurde.

Ebenso wurde im Rahmen der Diskussion, auch von den Experten des CWTS, darauf hingewiesen, dass Bibliometrie gerade in der Medizin nur ein Parameter ist, der in einer Matrixanalyse zur Messung des wissenschaftlichen Outputs neben anderen herangezogen und nie isoliert betrachtet werden sollte. Darüber hinaus wurde deutlich gemacht, dass bibliometrische Analysen in der Medizin nur mit Kenntnis detaillierter methodischer Grundlagen interpretiert werden sollten. Die aus der Betrachtung der klinischen Forschung abgeleiteten Fragen und Probleme der bibliometrischen Analysen in diesem Bereich haben dazu geführt, dass sich der Wissenschaftsrat im Folgenden ausführlicher und disziplinenübergreifend mit dem Thema der Bibliometrie befasst hat, was letztendlich zum hier vorliegenden Bericht führte.

5.2 Bildungsforschung

Um den Status Quo der Bildungsforschung in Österreich sowie im europäischen Vergleich zu erheben, wurde von der Innovationsstiftung für Bildung (ISB) die Studie „*Determination of the State of Educational Research in Austria*“⁸³ (ISB-Studie) durchgeführt. Im Auftrag der ISB wurde vom FWF und dem ÖWR hierfür die Datengrundlage in drei Schritten geschaffen: eine detaillierte Interview-Studie, eine Online-Umfrage und eine von CWTS durchgeführte bibliometrische Analyse. Auf dieser Datenbasis hat ein internationales externes Expertinnen- und Expertenpanel Empfehlungen formuliert.⁸⁴

Im Folgenden werden die wesentlichen Ergebnisse der bibliometrischen Analyse der Bildungsforschung in Österreich sowie in den neun europäischen Vergleichsländern zusammengefasst. Tabelle 8 enthält einen Vergleich des Publikationsoutputs (P) sowie den prozentualen Anteil der Bildungsforschung am gesamten Publikationsoutput der Länder. Dabei ist zu berücksichtigen, dass der Anteil der wissenschaftlichen Veröffentlichungen in Medien, die von WoS erfasst werden (interne Abdeckung) eine relativ hohe Bandbreite - zwischen 55% (Niederlande) und 36% (Großbritannien) - aufweist. Österreich findet sich im mittleren Bereich (46%) wieder. In Großbritannien werden Publikationen überwiegend in Medien veröffentlicht, welche nicht von WoS erfasst sind. Die zweite Spalte zeigt, dass die Bildungsforschung in den untersuchten Ländern insgesamt nur einen geringen Anteil an allen (über das WoS erfassten) Publikationen

⁸³ Hesse, Friedrich W. u. a., 2019.

⁸⁴ Der vollständige CWTS-Bildungsforschungsreport „Research Performance and Benchmark Analysis of Educational Sciences in Austria“ befindet sich im Anhang der ISB-Studie (Hesse, Friedrich W. u. a., 2019). Im Folgenden wird auf den CWTS-Bildungsforschungsreport und die darin verwendeten Seitenangaben verwiesen, um die exakte Zurückverfolgung ebendort zu erleichtern.

beisteuert. Lediglich vier Länder, Großbritannien (1,39%), Niederlande (1,19%), Finnland (1,12%) und Norwegen (1,08%), weisen einen Prozentsatz knapp über eins auf; in Österreich liegt dieser Wert bei 0,28%. Die erste Spalte in Tabelle 8 gibt Auskunft über den Publikationsoutput (P) der jeweiligen Länder. Da es sich um absolute Zahlen handelt, ist eine länderübergreifende Vergleichbarkeit nur sehr schwer möglich.

Land	P	%	Interne Abdeckung (%)
Belgien	2043	0,74	52
Dänemark	778	0,38	40
Deutschland	6557	0,45	43
Finnland	1891	1,12	44
Großbritannien	21650	1,39	36
Niederlande	5912	1,19	55
Norwegen	1622	1,08	41
Österreich	522	0,28	46
Schweden	2423	0,71	40
Schweiz	1210	0,34	44

Tabelle 8: Übersicht über alle 10 Vergleichsländer: Publikationsoutput (P), Prozentanteil der Bildungsforschung am gesamten Publikationsoutput (%), Interne Abdeckung in Prozent; Bedeutung der Schattierung: 1. Reihung Rot, 2. Reihung Orange, 3. Reihung Gelb, letzte Reihung Grün⁸⁵

Hingegen stellt Tabelle 9 die absoluten Publikationszahlen (P) in Zusammenhang mit den größtenteils unabhängigen und daher aussagekräftigeren bibliometrischen Indikatoren MNCS, MNJS und PP(top 10%) der 10 Vergleichsländer. Hier ist ersichtlich, dass, obwohl absolut gesehen, Österreich den niedrigsten Publikationsoutput aufweist, die anderen Indikatoren in durchaus ähnlichen Größenordnungen wie bei den Vergleichsländern liegen. Österreich weist, ebenfalls wie Dänemark, einen MNCS von 0,9 auf, die Schweiz und Deutschland liegen mit 0,87 und 0,84 niedriger, die Niederlande mit einem Wert von 1,26 über dem Weltdurchschnitt. Vergleicht man den absoluten Publikationsoutput (P) mit dem normierten MNCS, so lässt dies erkennen, dass Veröffentlichungen aus Deutschland und Großbritannien trotz eines hohen Publikationsoutputs im Bereich der Bildungsforschung eine vergleichsweise niedrige Wirkung in Bezug auf Zitationen entfalten.

⁸⁵ Adaptiert von CWTs-Bildungsforschungsreport, S. 13.

<i>Unit</i>	<i>P</i>	<i>MNCS</i>	<i>MNJS</i>	<i>PP</i> (top 10%)
<i>Austria</i>	522	0,9	0,86	9%
<i>Belgium</i>	2043	1,1	1,06	11%
<i>Denmark</i>	778	0,9	0,93	7%
<i>Finland</i>	1891	0,97	0,97	9%
<i>Germany</i>	6557	0,84	0,81	8%
<i>Netherlands</i>	5912	1,26	1,19	14%
<i>Norway</i>	1622	1	0,97	9%
<i>Sweden</i>	2423	0,96	0,96	7%
<i>Switzerland</i>	1210	0,87	0,85	8%
<i>United Kingdom</i>	21650	1,05	1,02	10%

Tabelle 9: Bibliometrische Analyse der Bildungsforschung - *P*, *MNCS*, *MNJS* und *PP*(top 10%) Indikatoren der 10 Vergleichsländer⁸⁶

Die beiden übrigen Indikatoren (*MNJS* und *PP*(top 10%)) folgen einem ähnlichen Trend wie der *MNCS*. Hier liegen die Niederlande mit einem *MNJS* von 19% über dem Weltdurchschnitt deutlich voran, wobei auch hier alle anderen Länder viel näher an den Weltdurchschnitt heranrücken. Die Niederlande weisen mit 14% auch den höchsten *PP*(top 10%)-Wert auf.

Tabelle 9 oben hält für Österreich leicht unterdurchschnittliche Werte für den *MNCS* (0,9) und *MNJS* (0,86) fest, sowie einen *PP*(top 10%) von 9%. All diese Werte sind niedriger als jene der Spitzensreiter. Betrachtet man die einzelnen Analyseperioden im Detail, so zeichnet sich ein interessantes Bild ab, welches in Abbildung 6 veranschaulicht ist. Hier ist ersichtlich, dass nach einem anhaltenden Abwärtstrend der beiden Indikatoren *MNCS* und *MNJS* eine positive Entwicklung einsetzt. Der Impact der österreichischen Bildungsforschungspublikationen liegt zwischen den Analyseperioden 2005-2009 und 2008-2012 mit *MNCS*-Werten zwischen 1,03 und 1,15 leicht über dem Weltdurchschnitt. Auch der *PP*(top 10%) mit Werten zwischen 11% und 14% nähert sich jenen der Tabellenführer.

⁸⁶ Ebd., S. 15.

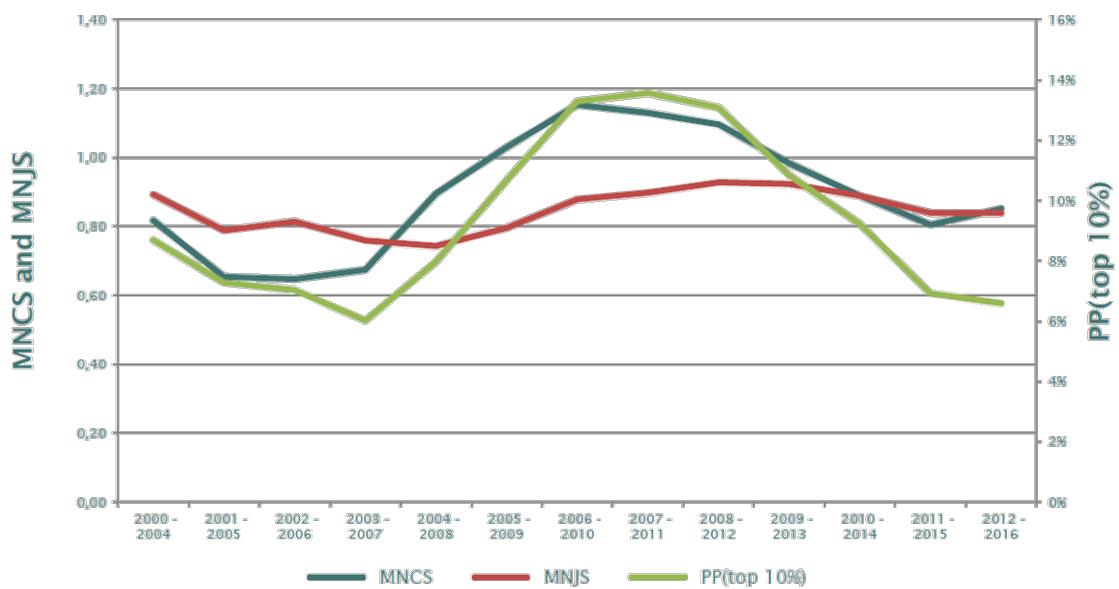


Abbildung 6: Zeitliche Entwicklung von MNCS, MNJS und PP(top 10%) in Österreich zwischen 2000-2016⁸⁷

Die bibliometrische Analyse zur Performance der einzelnen Institutionen in der österreichischen Bildungsforschung zeigt, dass die fünf aktivsten Einrichtungen im Sinne des Publikationsoutputs (P) die Universität Wien (144), die Universität Graz (68), die Universität Salzburg (49), die Johannes Kepler Universität Linz (40) und die Universität Innsbruck (31) sind. Die Gegenüberstellung von Publikationsoutput und MNCS in Abbildung 7 zeigt jedoch, dass die Publikationen der Universität Salzburg trotz des relativ niedrigen absoluten Outputs einen überdurchschnittlichen normierten Zitations-Impact aufweisen. Auch in der Gegenüberstellung mit den anderen Vergleichsländern nimmt die Universität Salzburg in Bezug auf den normierten Index MNCS mit 1,56 eine Spitzenstellung ein; der beste länderbezogene Wert (Niederlande) beträgt 1,26. Dies trifft auch auf den PP(top 10%) Indikator zu, welcher für die Universität Salzburg bei 22% (gegenüber 14% in den Niederlanden) liegt.

⁸⁷ Ebd., S. 20.

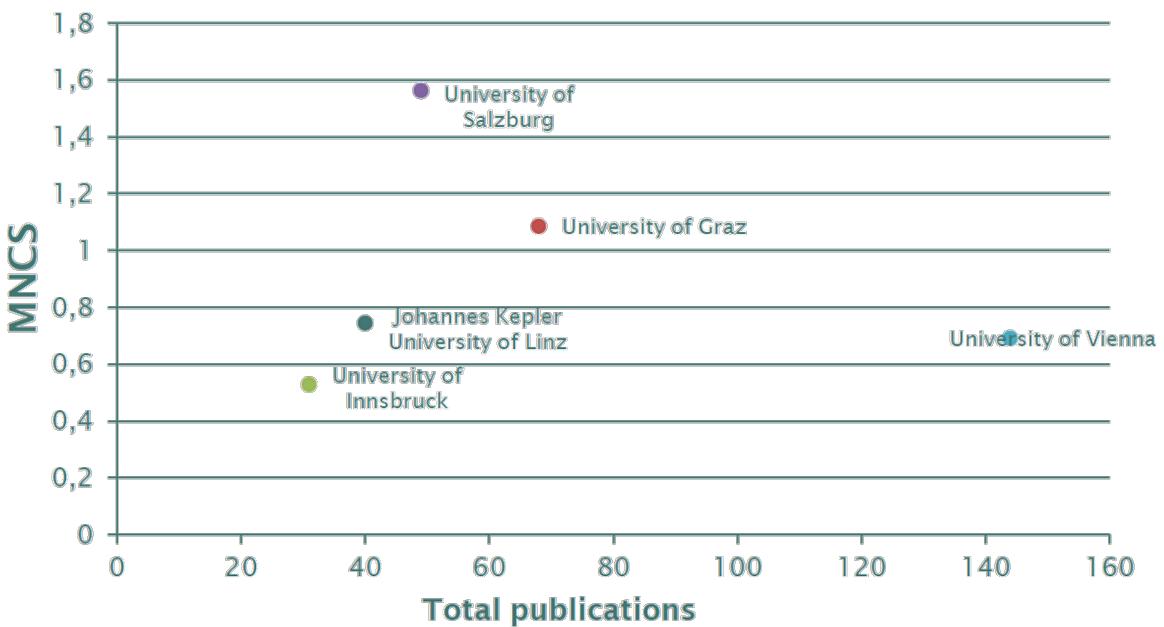


Abbildung 7: Zusammenhang zwischen Publikationsoutput (P) und MNCS fünf relevanter Akteure in der österreichischen Bildungsforschung⁸⁸

Neben rein bibliometrischen Daten wurden zur Visualisierung der Studienergebnisse sogenannte *term maps* von allen Vergleichsländern erstellt. Hierfür wurden die häufigsten Begriffe aus den Titeln und Abstracts aller Publikationen aller Vergleichsländer extrahiert und netzwerk-analytisch untersucht. Die Begriffe wurden in einem Koexistenznetzwerk angeordnet, sodass der Abstand zwischen den Kreisen die Beziehungen der häufigsten Begriffe, basierend auf der Häufigkeit deren gemeinsamen Auftretens, zueinander widerspiegelt. Je öfter ein Begriff kommt, desto größer wird er in der Begriffskarte ausgewiesen. Somit bieten die *term maps* einen anschaulichen Überblick über die wichtigsten in der Literatur behandelten Themen und deren Beziehungen zueinander.

Die Analyse der wichtigsten Begriffe der Publikationen aller 10 Vergleichsländer und die darauf basierende *term map* (Abbildung 8) erlaubt die Unterscheidung von vier Subbereichen innerhalb der Bildungsforschung: Rot repräsentiert den Bereich Bildungspolitik, Grün jenen der allgemeinen bildungswissenschaftlichen Forschung, Gelb umfasst Begriffe, die mit spezifischen Bildungsbedürfnissen in Zusammenhang stehen und Blau klinische Messungen und Methoden. Die Größe der einzelnen Kreise repräsentiert die Häufigkeit eines Begriffs im Gesamtkorpus. Solche *term maps* müssen jedoch mit Vorsicht interpretiert werden, da z.B. in diesem Fall der

⁸⁸ Ebd., S. 22.

Publikationsoutput von Großbritannien annähernd der Summe der restlichen Vergleichsländer entspricht und daher Abbildung 8 stark zu beeinflussen vermag.

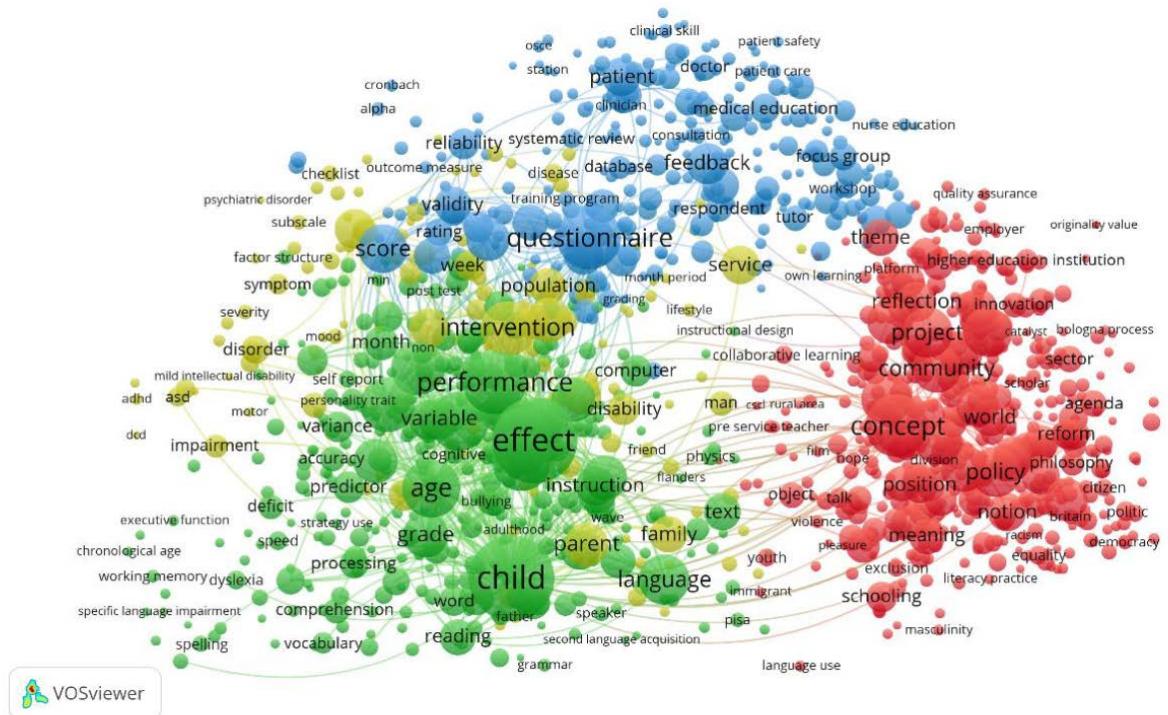


Abbildung 8: Term map aller Vergleichsländer⁸⁹

Die Größe der Kreise, basierend auf dem Gesamtkorpus aller Vergleichsländer, werden auch für die Generierung der *term maps* für die einzelnen Länder übernommen, um einen Vergleich der länderspezifischen Aktivität mit jener aller Vergleichsländer gesamt zu ermöglichen. Die Farbcodierung der Länderkarten variiert jedoch; hier signalisiert Gelb eine hohe Intensität der Publikation, diese nimmt über Grün bis hin zu Blau ab. Die Abbildungen 9, 10 und 11 enthalten beispielhaft die für Österreich, Finnland und Großbritannien generierten *term maps*.⁹⁰

⁸⁹ Ebd., S. 24.

⁹⁰ Die *term maps* aller Vergleichsländer befinden sich im CWTS-Bildungsforschungsreport, S. 25ff.

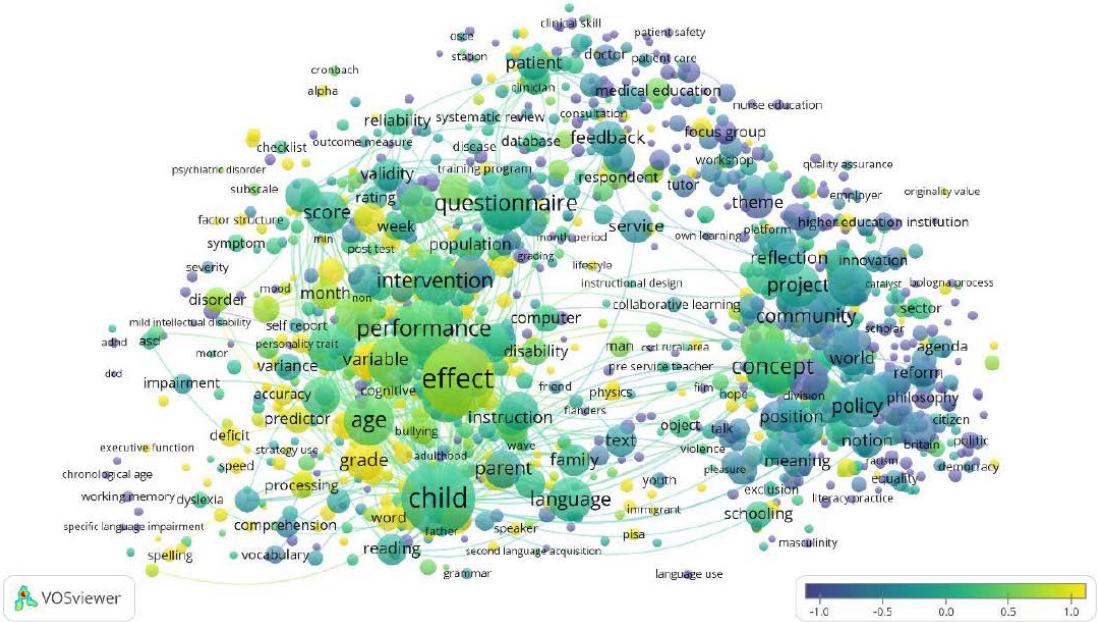


Abbildung 9: Term map der österreichischen Bildungsforschung⁹¹

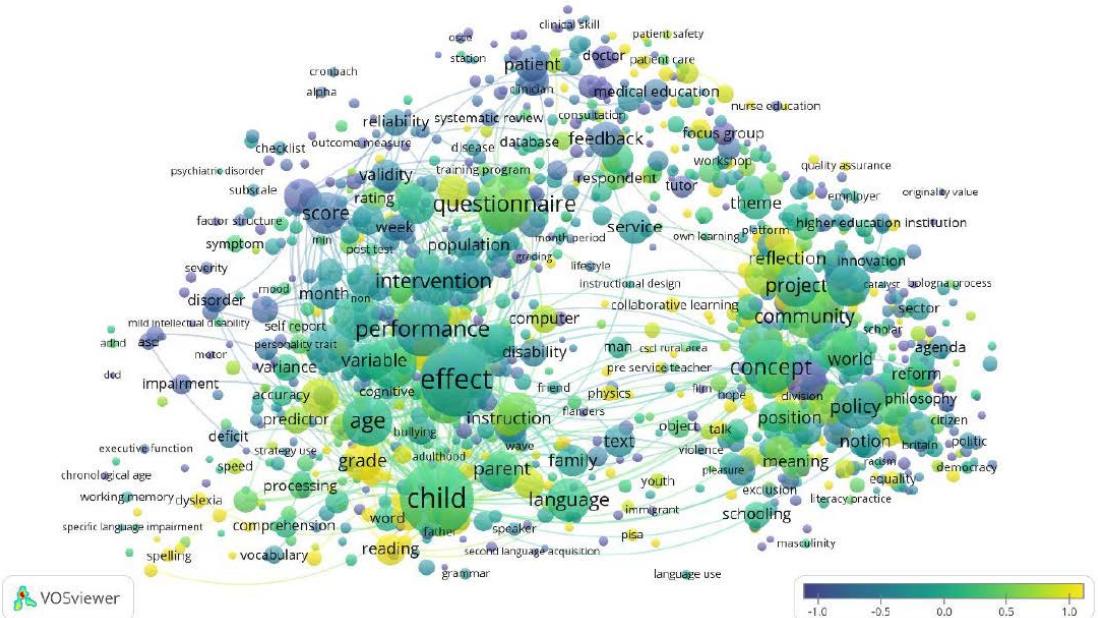


Abbildung 10: Term map der Bildungsforschung Finnlands⁹²

⁹¹ Ebd., S. 25.

⁹² Ebd., S. 27.

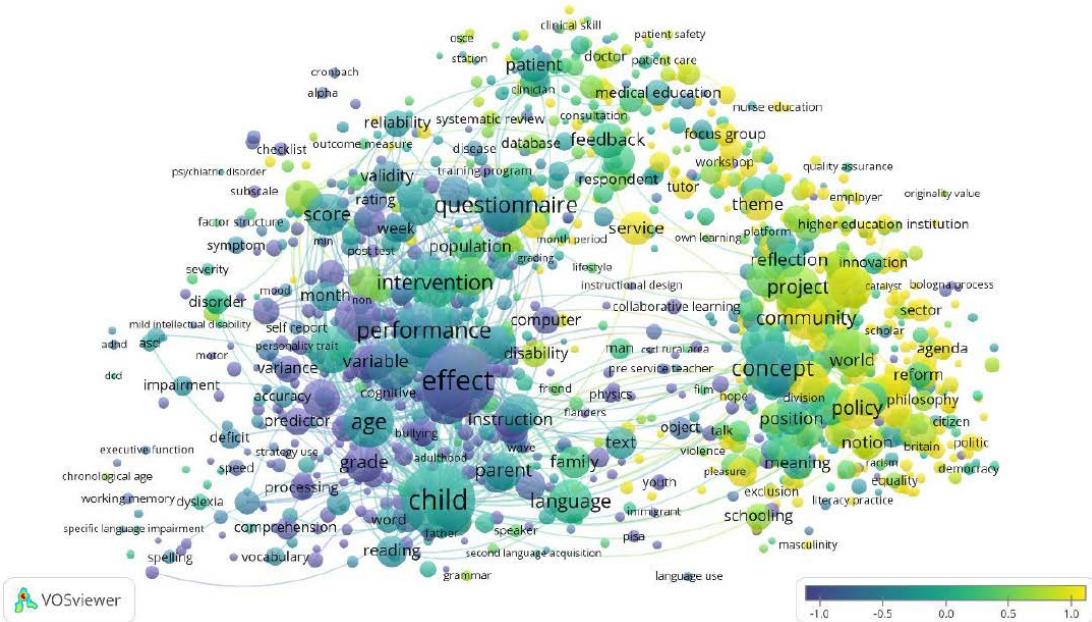


Abbildung 11: Term map der Bildungsforschung Großbritanniens⁹³

Der Vergleich der *term maps* der einzelnen Länder zeigt, dass in Österreich die Bildungswissenschaften stärker auf den Subbereich der allgemeinen bildungswissenschaftlichen Forschung fokussiert sind, Finnland hingegen eine eher ausgewogene Verteilung über alle Subbereiche hinweg aufweist. Großbritannien andererseits zeigt einen hohen Grad an Aktivität im Bereich Bildungspolitik auf, was möglicherweise den aus Tabelle 8 ersichtlichen niedrigen Anteil an Publikationen in WoS gelisteten Fachzeitschriften (interne Abdeckung von 36%) begründet.

5.3 Informatik

Der Wissenschaftsrat weist noch einmal darauf hin, dass im folgenden Kapitel Zwischenergebnisse präsentiert werden. Dies geschieht vornehmlich, um den Arbeitsprozess und die Methodik zu veranschaulichen, und somit Potentiale und Grenzen der Bibliometrie aufzuzeigen.

2019 beauftragte der Wissenschaftsrat CWTS Leiden mit einer bibliometrischen Analyse der Informatik in Österreich, deren Zwischenergebnisse im Rahmen der Veranstaltung „Status quo der Informatik in Österreich“ präsentiert wurden.⁹⁴ Die Studie durchlief zwei zentrale Schritte;

93 Ebd., S. 30.

⁹⁴ Noyons, Ed, 2019b; Noyons, Ed, 2019c.

der erste bezieht sich ausschließlich auf Zeitschriften- und Reviewartikel der sieben WoS *computer science subject categories*. Der zweite bezieht sich auf die bislang von CWTS nicht verwendeten Daten aus dem *Conference Proceedings Citation Index* (CPC-Index), welche im Vergleich zu den anderen WoS Daten bisher relativ unausgereift waren. Die Exklusion des CPC-Index stellt in einigen Disziplinen, wie der Informatik, jedoch eine klare Limitierung des Analyseprozesses dar. Daher wurden für diese Studie Tagungsbeiträge von ausgewählten Fachtagungen und Konferenzen aus dem CPC-Index eingebunden. Eine Liste mit den wichtigsten Konferenzen, angeführt auf www.csrankings.org, wurde CWTS von Expertinnen und Experten bereitgestellt.

Im Folgenden werden die Ergebnisse der Studie vorgestellt, welche im Rahmen der Tagung des Österreichischen Wissenschaftsrates „Informatik in Österreich: Perspektiven und Strategien“ im November 2019 präsentiert wurden. Hier durchliefen sie eine fachkritische Feedbackschleife durch die Teilnehmerinnen und Teilnehmer. Die hier präsentierten Ergebnisse stellen lediglich einen Entwicklungsschritt in der bibliometrischen Analyse der Informatik in Österreich dar. Zwischenzeitlich wurde der CPC-Index vollständig in das CWTS-eigene CI-System integriert und CWTS mit der Überarbeitung der Analyse beauftragt. Der erste Teil der Ergebnisse bezieht sich ausschließlich auf die Analyse von Zeitschriftenartikel und Reviews, gefolgt von der Analyse der Konferenzbeiträge.

Bibliometrische Analyse der Zeitschriftenartikel und Reviews

Als anfänglichen Schritt galt es, das relativ breite akademische Feld der Informatik (*computer science*) abzustecken und eine Datenbasis für die bibliometrische Analyse zu schaffen. Hierfür wurden die Zeitschriftenartikel und Reviews der WoS *computer science subject categories*, die sogenannten *seed publications*, innerhalb der fünf Disziplinen der Top-Ebene des publikationsbezogenen Kategorisierungssystems des CWTS sowie innerhalb der rund 4000 Forschungsbiete (Cluster) der niedrigsten Ebene verortet. Wie bereits auf Seite 29 festgehalten, basiert das CWTS-Kategorisierungssystem auf einer netzwerkanalytischen Untersuchung aller in der WoS-Datenbank erfassten Zeitschriftenartikel und Reviews sowie deren Zitationsbeziehungen. Die Verortung der Informatik *seed publications* innerhalb des Kategorisierungssystems ermöglicht Einsichten darüber, ob die Publikationen lediglich auf die Disziplin *Mathematics & Computer Science* beschränkt sind, oder ob Publikationsaktivitäten auch in anderen Disziplinen

feststellbar sind. Die Clusteranalyse wurde in dieser Studie auch zur publikationsbasierten Abgrenzung des Feldes Informatik herangezogen. Hierfür wurde ein Mindestwert von 100 *seed publications* pro Cluster gesetzt, um zum Gebiet der Informatik gezählt zu werden.⁹⁵

Abbildung 12 stellt die „Informatik“ eingebettet in der Welt aller Wissenschaften dar. Die Größe der einzelnen Cluster spiegelt die Zahl aller WoS-Publikationen wider, ein großer Kreis bedeutet eine höhere Anzahl an Publikationen. Die Einfärbung repräsentiert den Anteil der Informatik *seed publications* und reicht von Gelb (ein hoher Anteil) bis Blau (niedriger Anteil). Die Darstellung illustriert deutlich, dass gelbe Cluster großteils in der Kerndisziplin *Mathematics & Computer Science* angesiedelt sind. Es ist jedoch auch ersichtlich, dass andere Disziplinen ebenso gelbe Cluster beheimaten und grün-türkise Cluster in allen Disziplinen zu finden sind. Es kann weiters festgestellt werden, dass auch außerhalb der Kerndisziplin einige relativ große Kreise gelb und gelb-grün codiert sind. Diese Ergebnisse deuteten auf eine hohe Interdisziplinarität der Informatik hin, was wiederum eine weithin akzeptierte Abgrenzung des Feldes Informatik erschwert. Gleichzeitig ermöglicht dies aber auch eine neue Perspektive auf das Feld über gewohnte Fachgrenzen hinaus. Eine Beurteilung insbesondere auffälliger, fragwürdiger bzw. grenznaher Cluster durch Expertinnen und Experten aus unterschiedlichen Bereichen der Informatik ist für eine zuverlässige Abgrenzung des Feldes jedoch unumgänglich und wird für die Überarbeitung der Analyse derzeit in Kooperation zwischen CWTS und ÖWR durchgeführt.

⁹⁵ Dieser Grenzwert wurde von CWTS gesetzt. Ein niedrigerer bzw. höherer Grenzwert zeigte kaum Auswirkungen auf die Clusteranalyse (E-Mail Kommunikation mit Ed Noyons, CWTS, 17. April 2020).

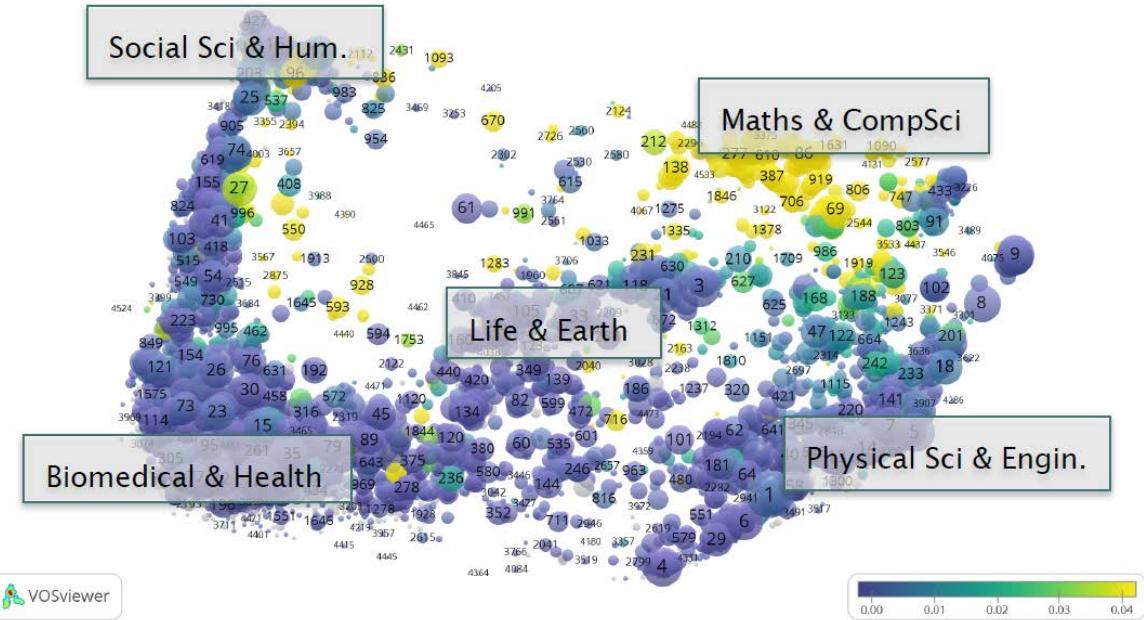


Abbildung 12: Verteilung der Informatik seed publications im CWTS-Kategorisierungssystem (Gelb: hoher Anteil, Blau: niedriger Anteil)⁹⁶

Tabelle 10 enthält eine Übersicht über die wichtigsten bibliometrischen Indikatoren (*key performance indicators, KPI*) der Informatik in Österreich sowie der anderen 10 Vergleichsländer. Hinsichtlich des Publikationsoutputs P ist eine große Bandbreite zwischen 80.398 in Großbritannien und 6.710 in Norwegen ersichtlich. Österreich findet sich mit 10.351 Publikationen an drittletzter Stelle. Eine große Bandbreite kann auch in Bezug auf den MNCS und den MNJS festgestellt werden. In beiden Fällen liegt Österreich innerhalb der Vergleichsländer an der letzten Stelle. Dabei muss jedoch unterstrichen werden, dass die Werte beider Indikatoren (1,04 und 1,02 respektive) auch für Österreich knapp über dem Weltdurchschnitt liegen. Als Spitzenreiter bei beiden Indikatoren liegt die Informatik in der Schweiz mit 39 bzw. 26% über dem Weltdurchschnitt. Der Indikator PP(Top 10%) weist eine geringere Varianz auf und liegt zwischen 12% (Deutschland und Österreich) und 16% (Schweiz).

⁹⁶ Noyons, Ed, 2019c, S. 7

Land	P	MNCS	MNJS	PP(top 10%)
Belgien	14.324	1,10	1,10	12%
Dänemark	7.574	1,34	1,19	13%
Deutschland	62.239	1,06	1,03	12%
Finnland	10.343	1,17	1,09	14%
Großbritannien	80.398	1,23	1,16	14%
Israel	15.991	1,15	1,22	13%
Niederlande	24.422	1,30	1,20	14%
Norwegen	6.710	1,15	1,11	13%
Österreich	10.351	1,04	1,02	12%
Schweden	13.658	1,13	1,13	13%
Schweiz	15.905	1,39	1,26	16%

Tabelle 10: *Bibliometrische Analyse der Informatik: Indikatoren P, MNCS, MNJS und PP(top 10%) der 11 Vergleichsländer; Bedeutung der Schattierung: 1. Reihung Rot, 2. Reihung Orange, 3. Reihung Gelb, letzte Reihung Grün*⁹⁷

Abbildung 13 zeigt den Zusammenhang zwischen dem absoluten Publikationsoutput (P) und dem MNCS-Indikator. Man erkennt, dass Österreich mit 1,04 den geringsten normierten Zitations-Wert aufweist, und dass Norwegen und Dänemark mit einer deutlich geringeren absoluten Zahl wissenschaftlicher Zeitschriftenpublikationen höhere Zitationswerte erreichen. Bemerkenswert ist, dass Deutschland mit einem 6-fachen absoluten Publikationsoutput gegenüber Österreich beim MNCS praktisch gleichauf steht.

⁹⁷ Adaptiert von ebd., S. 14.

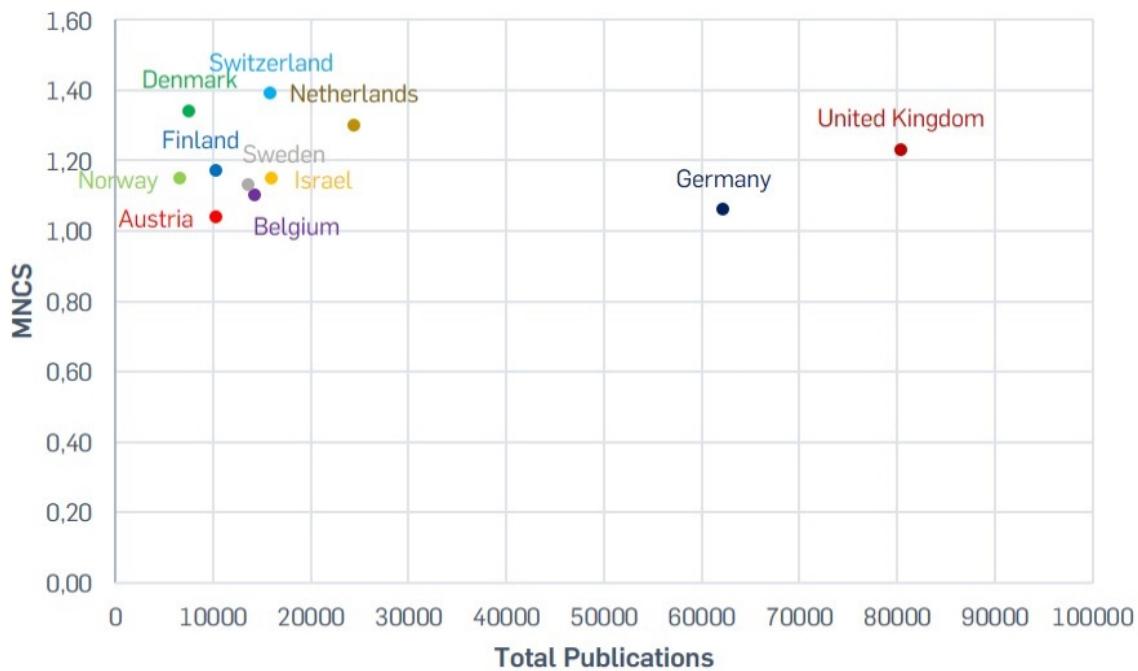


Abbildung 13: Zusammenhang zwischen absolutem Publikationsoutput (P) und MNCS⁹⁸

Der Indikator $PP(collab)$ erlaubt Aussagen über den Anteil der Publikationen, die in Zusammenarbeit unterschiedlicher Institute entstanden sind. Dies ist bei allen Ländern in einem ähnlichen Ausmaß (zwischen 60% der Publikationen in Deutschland und 71% in Norwegen) gegeben. Die Daten erlauben weiterhin eine Differenzierung, ob es sich um internationale Zusammenarbeiten $PP(int. collab)$ oder um Kooperationen mit der Industrie $PP(industry)$ handelt. Bei den internationalen Ko-Autorenschaften ist die Spannbreite etwas größer, zwischen 46% der Publikationen in Deutschland und Finnland, und 61% in der Schweiz. Die Zusammenarbeit mit der Industrie weist hingegen die geringsten Unterschiede auf, zwischen 9% aller Publikationen in Belgien und Norwegen, und 17% der Veröffentlichungen der Schweiz (siehe Abbildung 14).

⁹⁸ Adaptiert von ebd., S. 14.

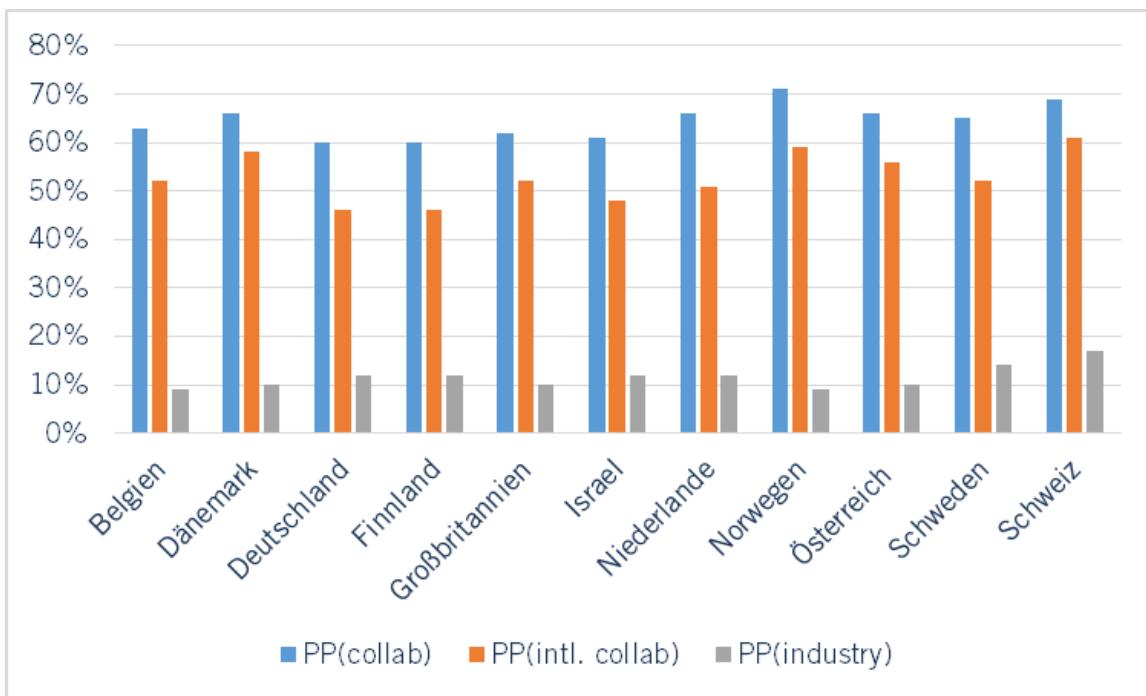


Abbildung 14: Indikatoren $PP(\text{collab})$, $PP(\text{intl. collab})$ und $PP(\text{industry})$ zur kooperativen Autorenschaft in den 11 Vergleichsländern⁹⁹

Richtet man den Fokus auf die Situation innerhalb Österreichs, so ermöglichen bibliometrische Analysen Einsichten zur Aktivität und dem Forschungsimpact der wichtigsten Akteure im Bereich Informatik. Die fünf aktivsten österreichischen Einrichtungen im Sinne des Publikationsoutputs sind die Technische Universität Wien (2.526), Technische Universität Graz (1.318), Universität Wien (1.279), Johannes Kepler Universität Linz (1.202) und die Universität Innsbruck (690).

Institution	P	MNCS	MNJS	PP(top 10%)	PP(industry)
Technische Universität Wien	2526	1,07	1,05	11%	7%
Technische Universität Graz	1318	1,15	0,98	13%	8%
Universität Wien	1279	1,02	1,07	11%	6%
Johannes Kepler Universität Linz	1202	0,95	0,96	9%	9%
Universität Innsbruck	690	1,26	1,16	14%	11%

Tabelle 11: Indikatoren P, MNCS, MNJS, PP (top 10%) und PP(industry) der fünf aktivsten österreichischen Institutionen¹⁰⁰

⁹⁹ Adaptiert von ebd., S. 14.

¹⁰⁰ Adaptiert von ebd., S. 28.

Aus Tabelle 11 geht hervor, dass die Universität Innsbruck trotz des vergleichsweisen niedrigen Publikationsoutputs von 690 den höchsten MNCS Wert (1,26) aufweist, gefolgt von der Technischen Universität Graz (1,15). Ein MNJS von 1,16 bedeutet, dass die Veröffentlichungen der Universität Innsbruck in Zeitschriften mit einem höheren Wirkungsgrad publiziert wurden, als jene der anderen vier Institutionen (MNJS zwischen 0,96 und 1,07). Hinsichtlich des Anteils an den Top 10% zitierten Publikationen liegen die Universität Innsbruck und die Technische Universität Graz mit PP(top 10%)-Werten von 14 und 13 Prozent näher aneinander und setzen sich von den anderen drei Einrichtungen leicht ab (zwischen 9 und 11%). Der Index PP(industry) repräsentiert den Anteil der Publikationen in Ko-Autorenschaft mit der Industrie; dieser liegt bei der Universität Innsbruck bei 11%, bei den übrigen Institutionen zwischen 6 und 9 Prozent.

Die zu Beginn des Kapitels in Abbildung 12 illustrierte Clusteranalyse stellt die „globale“ Situation der Informatik dar. Eine solche Analyse kann jedoch auch auf die nationale oder institutionelle Ebene fokussiert werden und in Kombination mit anderen zentralen Indikatoren wie z.B. P, MNCS, MNJS, PP(top 10%) oder PP(collab) zur Erhebung der sogenannten *strength*, *promising & risk areas* (SPR-Analyse) genutzt werden. Als *Strength areas* werden jene Cluster definiert, die sowohl einen hohen Output (P) als auch Impact (MNCS) aufweisen. *Promising areas* stellen Cluster mit ebenfalls hohem Impact dar, sind jedoch (noch) von einem relativ geringen Output gekennzeichnet. *Risk areas* hingegen zeigen einen hohen Output auf, ohne großen Impact erzielen zu können. Im Rahmen der Informatik-Studie erstellte CWTS zunächst ein Profil der fünf aktivsten österreichischen Institutionen. Nach Abschluss der detaillierten Informatik-Studie wird den einzelnen Universitäten ein Dashboard zur Verfügung gestellt werden, um die Durchführung einer internen Analyse zu ermöglichen. Dies kann insbesondere bei strategischen Überlegungen neue Perspektiven eröffnen und die Weiterentwicklung der Informatik am Standort fördern.

Um die SPR-Analyse zu illustrieren, sollen die Stärken der „österreichischen Informatik“ als Beispiel dienen. Die Grenzwerte und verwendeten Indikatoren können im CWTS-Dashboard individuell angepasst werden. Zur Definition eines Clusters als *key research area* wurden für dieses Beispiel folgende Kriterien festgelegt:

- Der Publikationsoutput P muss über die gesamte Analyseperiode (2000-2017) betrachtet mindestens 20 betragen und in den letzten vier Analysejahren mindestens 10,
- Der MNCS muss mindestens 20% über dem Weltdurchschnitt liegen (≥ 1.20),

- Die Wachstumsrate des Forschungsbereichs muss größer als 1 sein.¹⁰¹

An den fünf aktivsten österreichischen Institutionen erfüllen folgende Cluster die genannten Kriterien. Aufgrund der definierten hohen Forschungs- und Publikationsaktivität können sie, bibliometrisch betrachtet, als *key research areas* bezeichnet werden:¹⁰²

- *3D shape*
- *Brain computer interface*
- *Clique width*
- *Cloud computing*
- *Graph*
- *Graph transformation*
- *Music therapy*
- *Quantum walk*
- *Runge Kutta method*
- *Single image super resolution*
- *Software process improvement*
- *Unconstrained optimization*
- *Vehicle routing problem*

Diese Liste verwendet die Labels des CWTS-Kategorisierungssystems. Zum Zeitpunkt der Beauftragung der Analyse bildeten ausschließlich Fachzeitschriften und Reviews die Datengrundlage; eine Einbeziehung von Konferenzbeiträgen war nicht möglich. Dies führte zu teils überraschenden Ergebnissen; in Diskussionen erwiesen sich vor allem zwei Cluster (*music therapy* und *Runge Kutta method*) als besonders auffallend.

Die stetige Erweiterung der Datengrundlage macht die Einbeziehung von Konferenzbeiträgen mittlerweile möglich, weshalb der Wissenschaftsrat zunächst eine Persönlichkeit des Faches mit der Begutachtung der zugrundeliegenden Daten und in weiterer Folge CWTS Leiden mit der Überarbeitung der Analyse beauftragt hat.

¹⁰¹ Ebd., S. 30.

¹⁰² Adaptiert von Noyons, Ed, 2019b, S. 11-15.

Die von CWTS erstellten Profile der einzelnen Universitäten beinhalten nicht nur die zentralen Indikatoren des Clusters, sie erlauben auch ein vertieftes Verständnis der Top-Autorinnen und -Autoren sowie der Top-Begriffe der Publikationstitel. Ebenfalls können die einzelnen Artikel eingesehen werden. Das Wissen von Expertinnen und Experten ist hier zentral, um die Beiträge hinreichend bewerten und die Bedeutung des Forschungsgebietes für strategische Planungen eines Institutes gesichert beurteilen zu können.

Bibliometrische Analyse der Konferenzbeiträge

Eine Schwierigkeit bibliometrischer Analysen für die Informatik besteht darin, dass in diesem Fachgebiet, im Gegensatz zu den meisten naturwissenschaftlichen und technischen Disziplinen, die wissenschaftliche Relevanz von Tagungsbeiträgen sehr viel stärker ausgeprägt ist und die von Fachzeitschriften häufig übersteigt. Daher wurden die Analyseverfahren im Zuge dieser Studie teilweise adaptiert, sodass die Indikatoren P und *Mean Citation Score* (MCS) nunmehr sowohl für Fachzeitschriften als auch für Konferenzbeiträge ausgewertet werden können. Aus Abbildung 15 wird ersichtlich, dass die Bedeutung von Tagungsbeiträgen jedoch nicht in allen Ländern gleich stark ist; die Reihung der Vergleichsländer ändert sich markant, wenn diese dem Publikationsoutput basierend auf Konferenzbeiträgen folgt. So wechseln Großbritannien und Deutschland die top Plätze, da Deutschland hier einen weitaus höheren Output aufweist, und die Schweiz rückt von Platz fünf auf Rang drei vor.

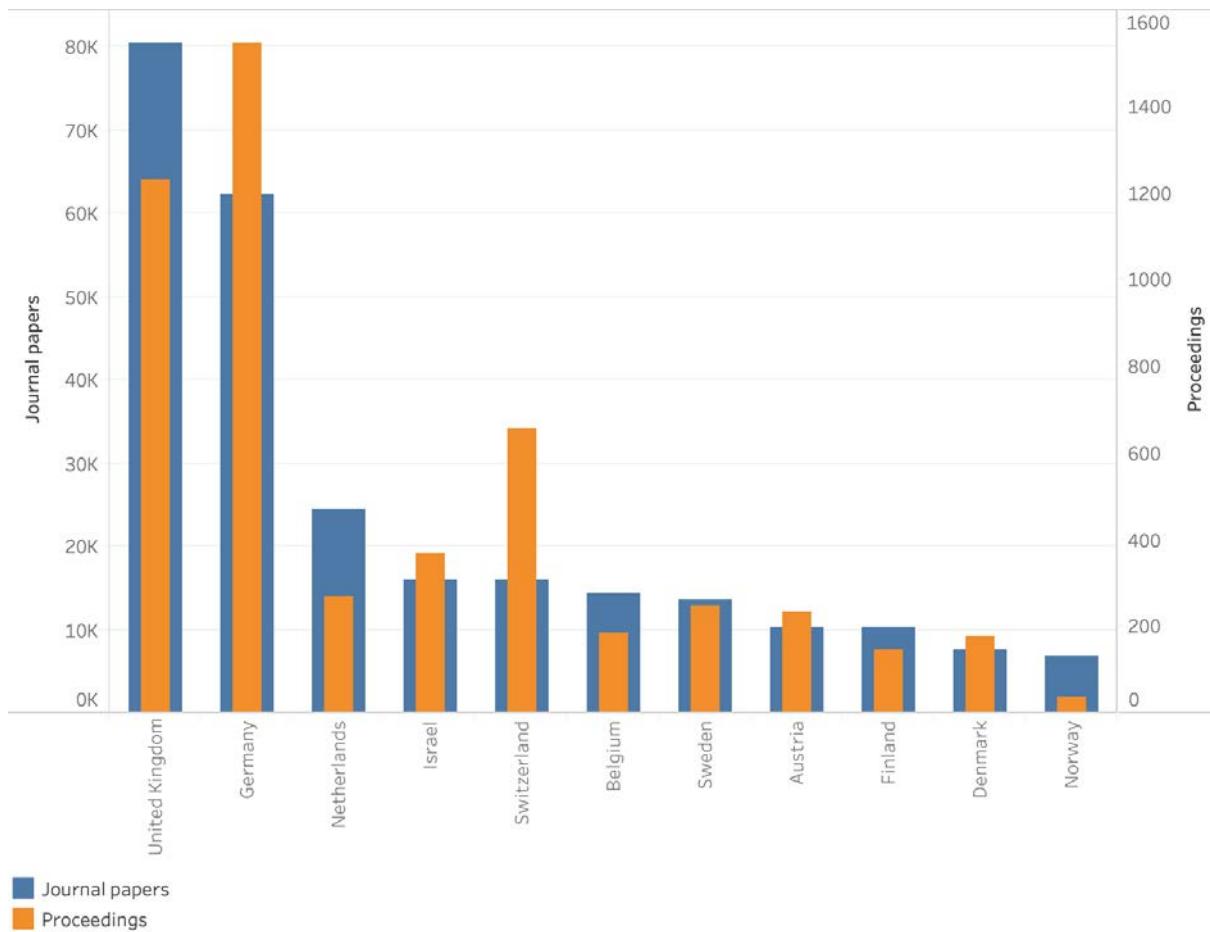


Abbildung 15: *Publikationsoutput (P) nach Fachzeitschriftenbeiträgen und Konferenzbeiträgen*¹⁰³

Der fachspezifisch normierte Impact-Indikator MNCS konnte für Tagungsbeiträge in dieser Studie noch nicht berücksichtigt werden. Um zumindest einen vagen Eindruck über die relative Bedeutung der Tagungsbände in der Informatik in Bezug auf die Zitationshäufigkeit zu geben, enthält Abbildung 16 eine Reihung der Vergleichsländer gemäß dem MNCS für die Fachzeitschriftenbeiträge und eine Gegenüberstellung mit dem nicht normierten MCS, ausgewertet nur für Tagungsbeiträge. Man erkennt, dass sich die Reihung der Länder auch hier signifikant ändert, je nachdem, ob man Fachzeitschriften oder Tagungsbeiträge betrachtet. So rücken Schweden und Belgien von Platz 8 und 9 respektive vor auf Platz 1 und 2. Österreich bildet das Schlusslicht unter den Vergleichsländern, wenn der MNCS betrachtet wird, rückt aber auf Platz 3, wenn der Impact der Konferenzbeiträge als Kriterium verwendet wird.

¹⁰³ Noyons, Ed, 2019c, S. 18.

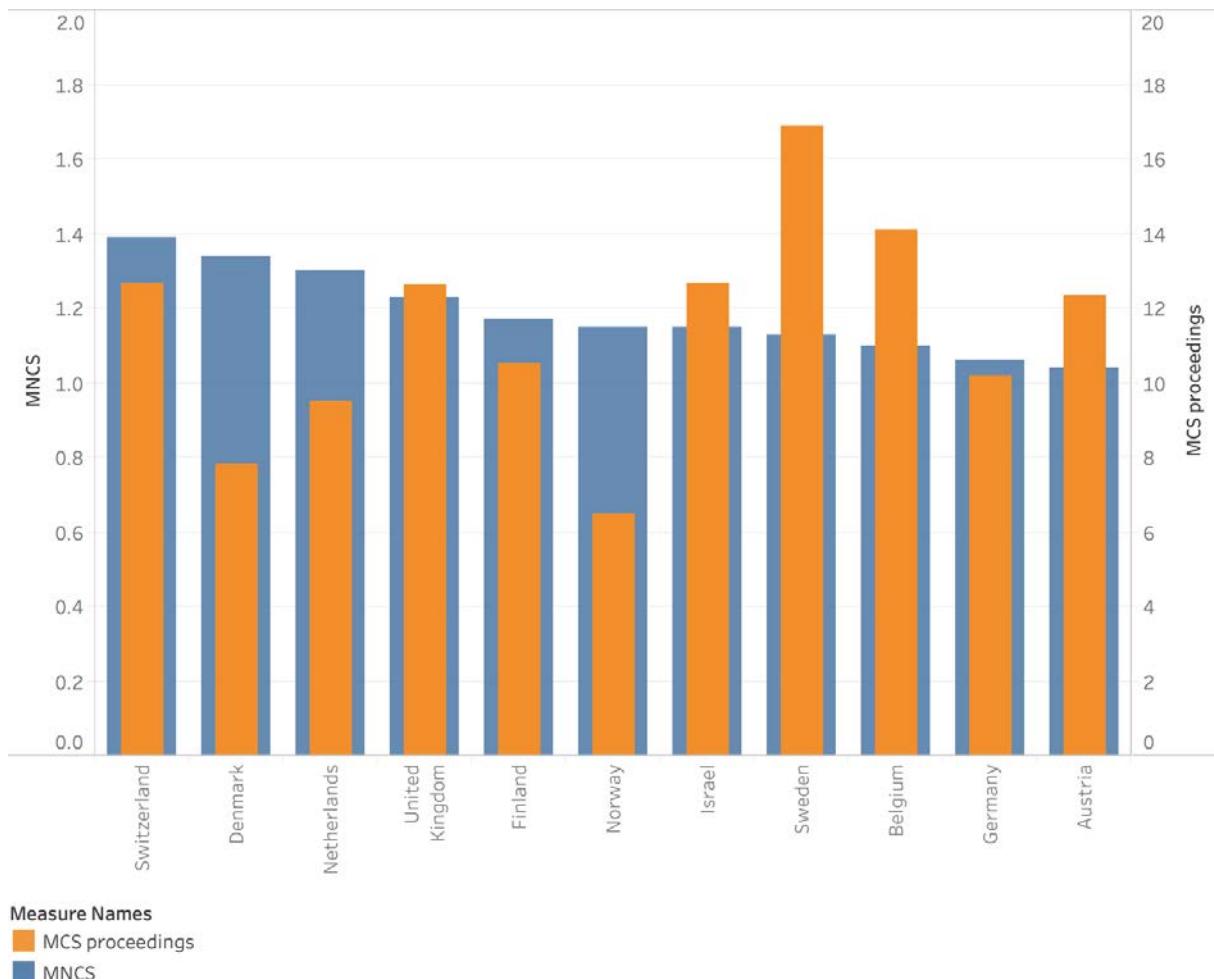


Abbildung 16: Impact-Indikatoren für Konferenzbeiträge (MCS) und für Fachzeitschriftenbeiträge (MNCS)¹⁰⁴

Wie für die Bildungsforschung, so wurden auch für die Informatik sogenannte *term maps* erstellt. Diese Koexistenznetzwerke basieren auf einer netzwerkanalytischen Untersuchung der am häufigsten verwendeten Begriffe der Titel und Abstracts der Konferenzbeiträge. Die dabei entstandenen fünf thematischen Subbereiche wurden durch Expertinnen und Experten im Bereich Informatik mit folgenden Labeln versehen: *systems & hardware*, *imaging & artificial intelligence*, *user interface*, *robotics* und *security & theory*. Abbildung 17 stellt die Verteilung der häufigsten Begriffe aller Vergleichsländer dar, wobei die unterschiedlichen Farben die thematischen Subbereiche repräsentieren, die Größe des jeweiligen Kreises die Häufigkeit des Begriffs in den Konferenzbeiträgen und die Distanz zwischen den Kreisen die Häufigkeit deren gemeinsamen Auftretens. Auf den folgenden Seiten bieten fünf *term maps* einen anschaulichen

¹⁰⁴ Ebd., S. 19.

Überblick über die am häufigsten verwendeten Begriffe aller Konferenzbeiträge der Vergleichsländer und deren Beziehungen zueinander (Abbildung 17), sowie deren zeitliche Entwicklung (Abbildung 18) und Impact (Abbildung 19). Diesen nachkommend dienen die *term maps* von Israel (Abbildung 20) und Österreich (Abbildung 21) als Beispiele, um die netzwerkanalytische Untersuchung der häufigsten Begriffe auf Länderebene zu illustrieren.

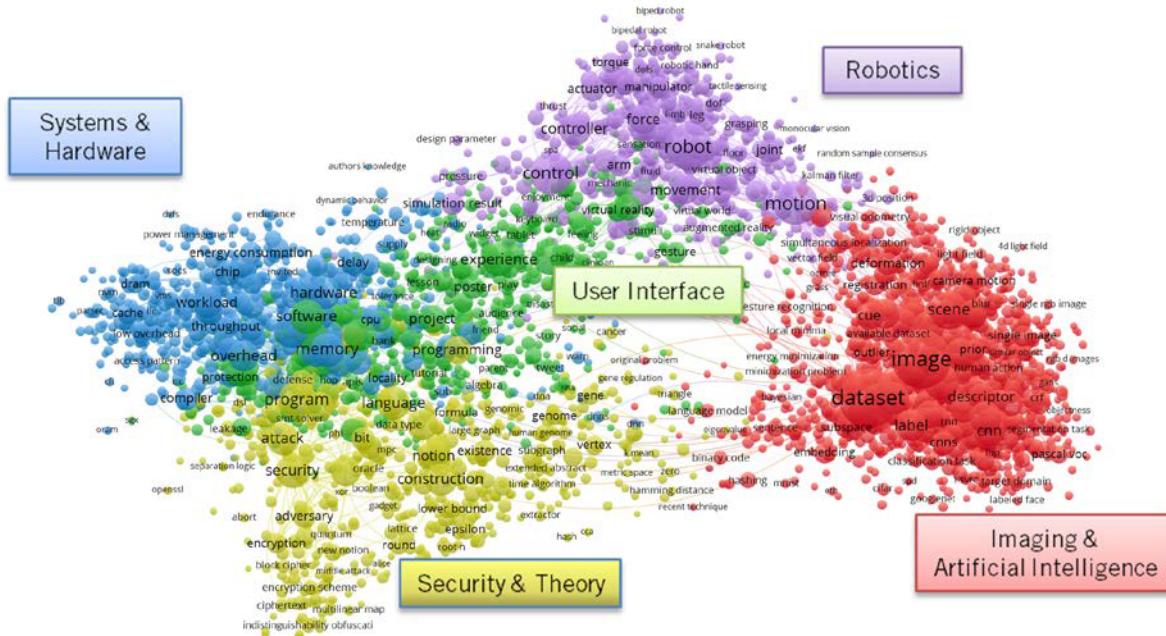


Abbildung 17: Term map der Konferenzbeiträge aller Vergleichsländer¹⁰⁵

Die Analyse der Publikationsjahre ermöglicht eine interessante Einsicht in die zeitliche Entwicklung der erwähnten Begriffe. Abbildung 18 bezieht sich auf alle Vergleichsländer und zeigt das durchschnittliche Publikationsjahr der in den jeweiligen Kreisen beinhalteten Konferenzbeiträge. Die Farbcodierung reicht hier von Blau (ältere Publikationen) bis Gelb (neuere Publikationen). Die Visualisierung zeigt, dass viele Begriffe des Gebiets *robotics* durchschnittlich ältere Konferenzbeiträge umfassen und kein gelber Kreis hervorsticht. Die anderen vier thematischen Subbereiche spiegeln das gesamte Spektrum der Farbkodierung wider. Eine größere

¹⁰⁵ Ebd., S. 21.

Anzahl gelber Kreise, besonders auch größerer gelber Kreise, ragen im thematischen Subbereich *imaging & artificial intelligence* hervor.

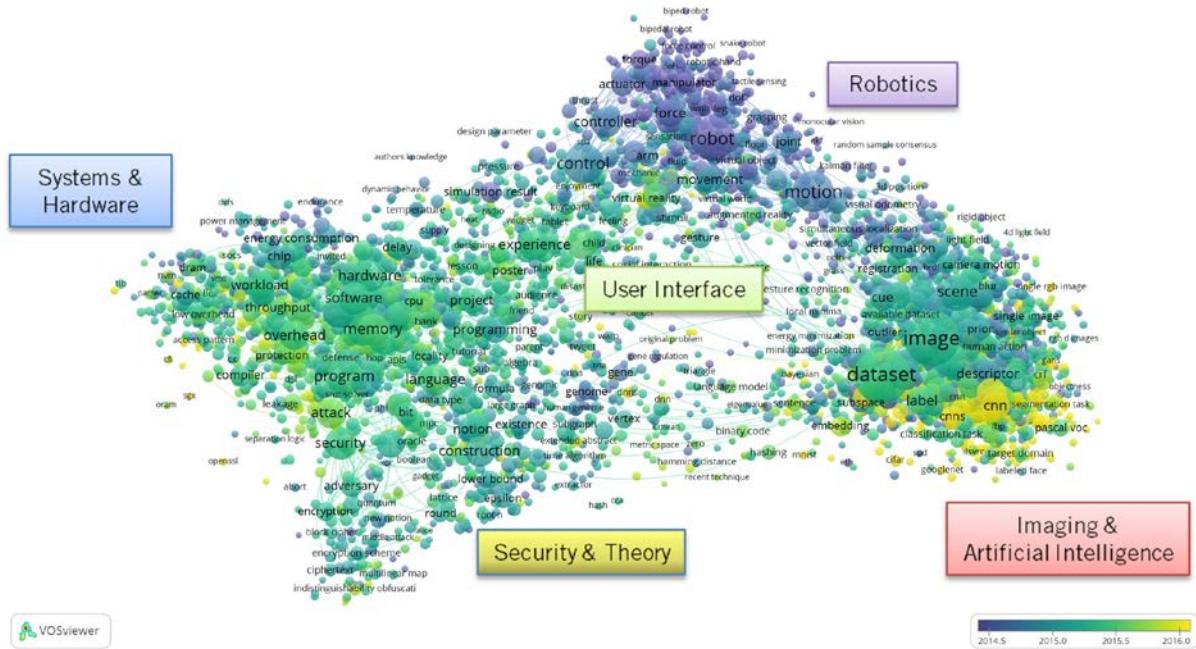


Abbildung 18: Term map der Konferenzbeiträge alle Vergleichsländer – durchschnittlicher Publikationszeitpunkt der häufigsten Begriffe (Gelb: neuere Publikationen, Blau: ältere Publikationen)¹⁰⁶

Abbildung 19 illustriert den Impact der einzelnen Begriffe aller Vergleichsländer, gemessen anhand des nicht-normierten, größtenunabhängigen MCS-Indikators. Die Farbskalierung reicht von Blau (niedriger MCS) bis Gelb (hoher MCS). Hier ist ersichtlich, dass besonders der thematische Subbereich *imaging & artificial intelligence* einen verhältnismäßig großen Impact erzielen konnte, und dass mehrere kleinere Kreise existieren, die ebenfalls einen hohen MCS aufweisen. Ferner ist auffällig, dass, gesamt betrachtet, viele der großen Kreise einen relativ niedrigen MCS erzielen und im blauen Farbspektrum liegen.

¹⁰⁶ Ebd., S. 22.

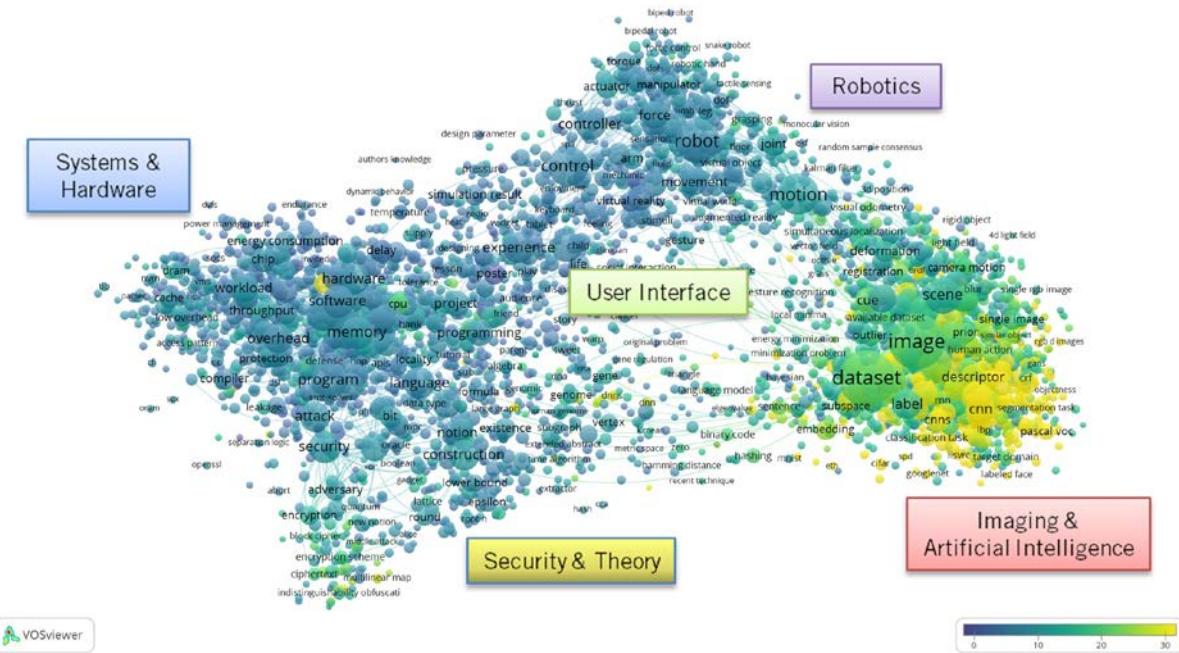


Abbildung 19: Term map der Konferenzbeiträge alle Vergleichsländer- Impact der häufigsten Begriffe (Gelb: hoher MCS, Blau: niedriger MCS)¹⁰⁷

Die Abbildungen 20 und 21 zeigen die Anteile der Konferenzbeiträge von Israel und Österreich an der *term map* aller Vergleichsländer (Abbildung 17). Die Größe der Kreise repräsentiert die Häufigkeit der jeweiligen Begriffe basierend auf allen Vergleichsländern insgesamt, die Farbcodierung repräsentiert den Anteil der Publikationen Israels bzw. Österreichs und reicht von Gelb (hoher Anteil) bis Blau (niedriger Anteil). Während für Israel eine klare Schwerpunktsetzung im thematischen Subbereich *security & theory* erkennbar ist, so ist die Informatik-Forschung in Österreich gleichmäßiger verteilt, wobei aber auch hier eine erhöhte Publikationsaktivität in den thematischen Subbereichen *imaging & artificial intelligence* und *security & theory* festzustellen ist.

¹⁰⁷ Ebd., S. 25.

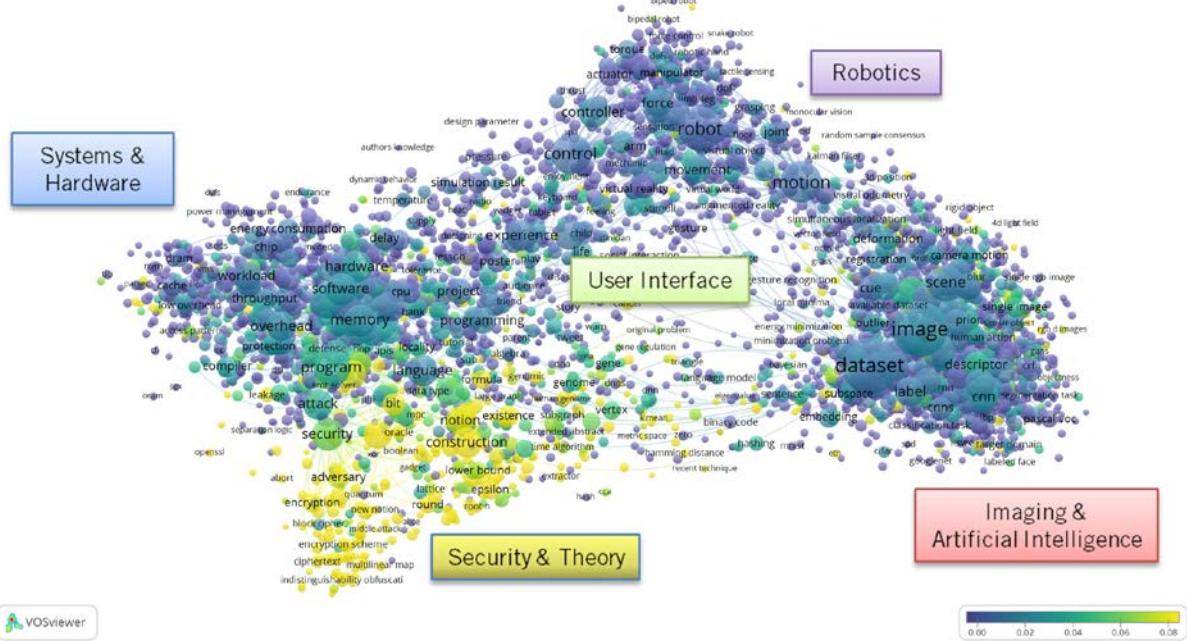


Abbildung 20: Term map der Konferenzbeiträge Israels: Anteil an allen Konferenzbeiträgen der Vergleichsländer (Gelb: hoher Anteil, Blau: niedriger Anteil)¹⁰⁸

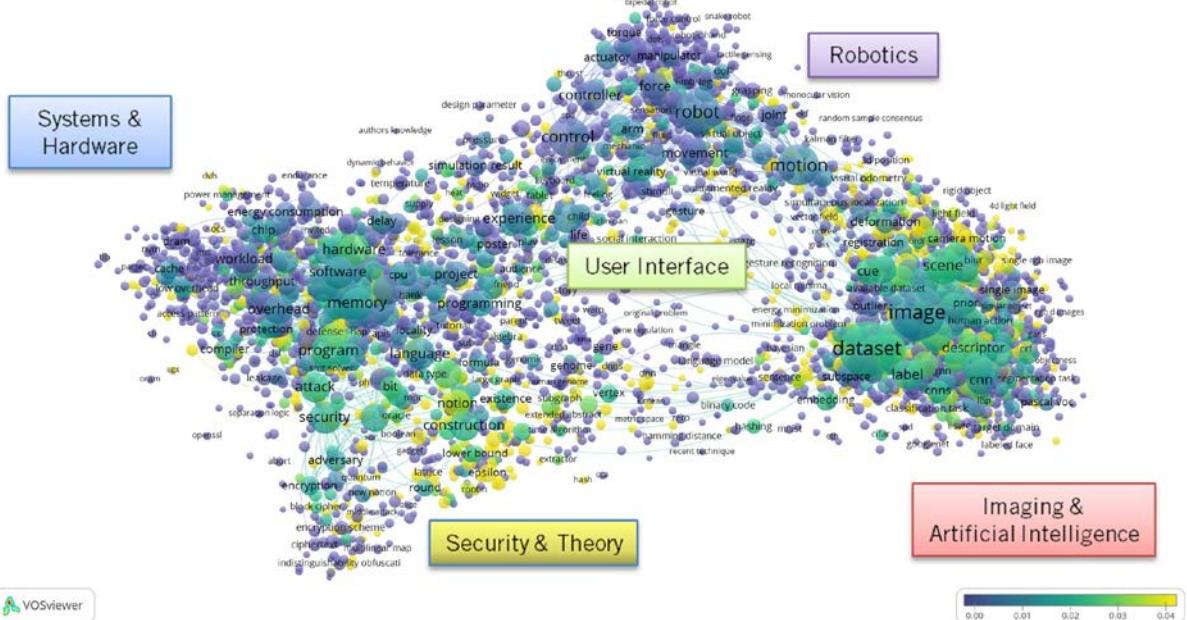


Abbildung 21: Term map der Konferenzbeiträge Österreichs: Anteil an allen Konferenzbeiträgen der Vergleichsländer (Gelb: hoher Anteil, Blau: niedriger Anteil)¹⁰⁹

¹⁰⁸ Ebd., S. 23.

¹⁰⁹ Ebd., S. 24.

Die hier präsentierten Daten bezüglich der Konferenzbeiträge müssen mit einer gewissen Vorsicht betrachtet werden. Es handelt sich zu diesem Zeitpunkt lediglich um die Anfänge der Erweiterung bibliometrischer Analyseverfahren auf Konferenzbeiträge. Die verwendeten *proceedings*-Rohdaten wurden zwischenzeitlich weiter bearbeitet, um die Vergleichbarkeit und Integration der Konferenzbeiträge zu ermöglichen, ebenso wie die Anwendung normierter Indikatoren. Diese methodischen Weiterentwicklungen werden einen weitaus aussagekräftigeren Einblick in die Forschungsleistung der Informatik in Österreich zulassen. Die in diesem Kapitel enthaltenen Fallbeispiele zeigten, dass die Möglichkeiten der Bibliometrie zur Evaluierung von Forschungsfeldern, selbst auf der Ebene einzelner Institute, mit der wachsenden Raffinesse der Methoden sowie der stetig wachsenden Datenmenge zunehmen. Die kritische Auseinandersetzung des ÖWR mit der breiteren Thematik der Bibliometrie führte zu einigen Erkenntnissen und Empfehlungen, welche Gegenstand des folgenden Kapitels sind.

6 Erkenntnisse und Empfehlungen

Zu den Problemen einer unkritischen Nutzung metrischer Methoden zur Forschungsevaluation wurde in *The Metric Tide* bereits ausführlich Stellung bezogen; auch der Wissenschaftsrat hat sich zur Vermessung der Wissenschaft bereits geäußert. Zudem fasst das *Leiden Manifesto* wesentliche Empfehlungen zum verantwortungsvollen Umgang mit bibliometrischen Daten zusammen, weshalb diese hier nicht wiederholt werden.¹¹⁰ Ein verantwortungs- und sinnvoller Einsatz metrik-basierter Methoden setzt voraus, deren Beschränkungen und Potentiale gleichermaßen zu kennen und zu würdigen, weshalb nachfolgend eine kritische Reflexion zentraler Erkenntnisse hinsichtlich bibliometrischer Analysen sowie deren Bedeutung für die Forschungsevaluation erfolgt. Gleichzeitig behält sich der Wissenschaftsrat vor, den Fokus in den daraus abgeleiteten Empfehlungen auf die Potentiale dieser Methoden zu legen.

Datengrundlage

Die Aussagekraft bibliometrischer Analysen hängt nicht zuletzt von einer Reihe von Parametern ab, deren Bestimmung vertiefte Einsichten in die Forschungskulturen der jeweiligen Forschungsdisziplinen erfordern. Dies betrifft u.a. die Auswahl von WoS *journal subject categories* eines Fachgebietes sowie die Auswahl der berücksichtigten Publikationsmedien.

Entscheidend für den Nutzen bibliometrischer Daten zur Bewertung von Fachgebieten sind die Verfügbarkeit und der Abbildungsgrad der verwendeten bibliometrischen Daten, welche die für das jeweilige Fachgebiet relevanten Publikationsmedien umfassen. Dies ist in den Natur-, den Lebens- und den Ingenieurwissenschaften, in denen internationale Zeitschriftenpublikationen relevant sind, weitgehend unproblematisch, in der Informatik wegen des hohen Stellenwerts von Konferenzen schon etwas schwieriger. In den Sozial-, Wirtschafts- und Geisteswissenschaften sind standardisierte Auswertungsmethoden noch nicht generell etabliert, wiewohl auch hier Ansätze zu erkennen sind, auf engere Fachbereiche zugeschnittene länderübergreifende Qualitätsstandards zu schaffen.

Die einer bibliometrischen Auswertung zugrundeliegende Datenbasis sollte durch Expertinnen und Experten aus dem jeweiligen Fachgebiet überprüft werden, um zu gewährleisten, dass den Spezifika dieses Fachbereichs Rechnung getragen wird. Dies trifft insbesondere auf die Klassifizierung von Wissenschaftsbereichen innerhalb von Fachdisziplinen zu, die als Grundlage für feldspezifisch normierte Indikatoren herangezogen werden. Eine solche Klassifizierung erfordert

¹¹⁰ Hicks, Diana u. a., 2015; Österreichischer Wissenschaftsrat, 2014; Wilsdon, James u. a., 2015.

tiefgehende Kenntnisse über die Struktur des betrachteten Wissenschaftsgebiets und kann daher nicht von außerhalb erfolgen. Nur dann sind indikator-basierte Datenanalysen eine sinnvolle, weithin akzeptierte Ergänzung zu qualitativen Informationen und subjektiven Urteilen.

Besondere Aufmerksamkeit sollte der Beurteilung interdisziplinärer Forschung zukommen, deren Eingliederung in disziplinäre Fachzusammenhänge mit entsprechenden Feldnormierungen eventuell starke Verzerrungen nach sich zieht. Gerade, um den vielfältigen negativen Anreizen im Hinblick auf die Weiterentwicklung interdisziplinärer Forschung Einhalt zu gebieten, bedürfen interdisziplinäre Forschungsansätze in bibliometrischen Analysen des besonderen Schutzes durch gezielte Berücksichtigung und kritisches Hinterfragen der Datenbasis.

Empfehlung

Der Wissenschaftsrat empfiehlt den wissenschaftlichen Communities, sich über fachinterne Qualitätsstandards im Zusammenhang mit Publikationsmedien zu verständigen, um die Interpretation von Indikatoren zu erleichtern, zu deren angemessener Weiterentwicklung beizutragen und deren Nutzen für Evaluationsprozesse zu erhöhen.

Datenzugang und -analyse

Bibliometrische Daten werden an den österreichischen Universitäten seit Jahren genutzt, um Forschungsleistungen auf verschiedenen Ebenen, beginnend bei der einzelnen Wissenschaftlerin bzw. dem einzelnen Wissenschaftler, über Forschungsgruppen, Fachbereiche und Fakultäten bis hin zur Universitätsebene zu messen. Die hierfür erforderlichen Daten werden systematisch, wenn auch an den verschiedenen Standorten in unterschiedlicher Weise, vorgehalten. Hierbei werden durchwegs disziplinspezifische Unterschiede berücksichtigt; quantitative, metrik-basierte Methoden werden fast durchwegs als Ergänzung bzw. Unterstützung von Evaluierungsprozessen bzw. Peer Review-Verfahren genutzt.

Die Aufbereitung und Analyse bibliometrischer Daten auf Basis interner und internationaler Datenbanken, wie im Bericht anhand von Beispielen dargestellt, kann einen erheblichen Aufwand verursachen, insbesondere wenn internationale Vergleiche gewünscht sind. Eine österreichweite Koordination des Zugangs zu externen Anbietern, die sich auf die Erstellung solcher Analysen spezialisiert haben, würde die Kosten für die Institutionen reduzieren und Abstimmungen erleichtern. In diesem Zusammenhang verweisen wir auch auf die spätere Empfehlung „Kompetenzzentrum“.

Empfehlung

Der Wissenschaftsrat empfiehlt einen gemeinschaftlich organisierten Zugang zu (externen) Dienstleistungen, um Aufwand und Kosten zu minimieren sowie die Abstimmung zu verstärken.

Indikatoren

Bibliometrische Analysen erfordern Einsichten zu den Stärken und Schwächen der angewandten Indikatoren. Feldspezifisch normierte Indikatoren sind in jedem Fall absoluten Indikatoren vorzuziehen, da diese die unterschiedlichen Wissenschaftskulturen der verschiedenen Fachdisziplinen berücksichtigen. Eine Vergleichbarkeit über Disziplinengrenzen hinweg ist in manchen Fällen trotz feldnormierter Indikatoren nur sehr eingeschränkt möglich, da der Abbildungsgrad durch bibliometrische Indikatoren disziplinspezifisch stark variiert.

Bibliometriebasierte Rankings klammern diese Unterschiede und methodischen Grenzen weitgehend ein, daher sollte deren Aussagekraft nicht überschätzt werden und ihre Verwendung als Grundlage für Entscheidungen nur in Kombination mit vertiefenden Analysen erfolgen. Der schnelle und unkritische Vergleich von Institutionen mit Hilfe von Rankings führt zu verzerrten Einschätzungen.

Empfehlung

Um die unterschiedlichen Wissenschaftskulturen der verschiedenen Fachdisziplinen in angemessener Weise zu berücksichtigen, sind feldspezifisch normierte Indikatoren absoluten Indikatoren vorzuziehen.

Auswertung bibliometrischer Daten

Fachspezifisch normierte und damit großenunabhängige Indikatoren (z.B. MNCS) und Gegenüberstellungen zu absoluten Indikatoren, wie z.B. zum Publikationsoutput, lassen gewisse Rückschlüsse auf den Impact der Forschung auf Institutions- wie auch auf Länder-Ebene zu. Durch fachspezifische Normierung lassen sich z.B. Forschungsschwerpunkte mit überdurchschnittlichem internationalen Impact, wiederum auf Länder- oder Institutionsebene, identifizieren. Visualisierte Auswertungen, z.B. in Form von *term maps*, lassen thematische Schwerpunkte der Forschung und deren innere Verflechtungen erkennen - Informationen, die ohne solche Methoden kaum zu gewinnen sind.

Die Analyse der zeitlichen Entwicklungen bibliometrischer Indikatoren und deren vergleichende Auswertung für einzelne Subfelder eines Forschungsgebietes im internationalen Kontext haben

Aussagekraft und lassen interessante Rückschlüsse auf die Entwicklung der Leistungsfähigkeit eines Fachgebietes innerhalb einer Institution bzw. innerhalb einer Region oder eines Landes zu. Insbesondere aus der Verlinkung unterschiedlicher (z.B. Impact und Output-bezogener) Indikatoren sowie aus der Kombination metrik-basierter Methoden (z.B. Bibliometrie, Altmetrics) können wertvolle Informationen zur Wirksamkeit und Sichtbarkeit der Forschung in der Scientific Community und in der Öffentlichkeit auf der Ebene des Forschenden, der einzelnen Institution wie auch auf überregionaler Ebene gewonnen werden.

Indikatoren müssen reflektiert und informiert festgelegt werden. In Verbindung mit fortgeschrittenen Auswertungs- und Visualisierungsmethoden ermöglichen sie wertvolle Einsichten in Stärken, Schwächen und Zukunftspotentiale, die für die Steuerung von Entwicklungsprozessen im Wissenschaftssystem genutzt und weiterentwickelt werden sollten.

Empfehlung

Der Wissenschaftsrat empfiehlt den Institutionen und wissenschaftlichen Communities, die Potentiale laufend verfeinerter Indikatoren und metrik-basierter Methoden zu nutzen und an deren Weiterentwicklung mitzuwirken, um aussagekräftige Informationen zu Schwerpunktsetzungen und Wirksamkeit der Forschung erhalten zu können; dies dient nicht zuletzt der eigenen Sichtbarkeit.

Bibliometrie als Werkzeug im Rahmen von Evaluationsprozessen

Zwei Erwartungshaltungen an bibliometrische Methoden im Kontext der Forschungsevaluierung müssen als Mythen verstanden werden. Die Bibliometrie stellt keine Zeitersparnis dar, da Auswahl und Aufbereitung der Daten sehr aufwendig sind und deren Interpretation das Wissen und den Input von Peers benötigen. Bibliometrische Analysen dürfen daher nicht als schnellere Methode der Evaluation missverstanden werden.

Auch die Erwartungshaltung, Bibliometrie wirke dem subjektiven akademischen Bias, der dem Peer Review-Verfahren angelastet wird, automatisch entgegen, da die Daten sozusagen für sich sprechen, darf keinesfalls unreflektiert hingenommen werden. Demgegenüber steht ein inhärenter methodischer Bias der Bibliometrie, wenn die Datengrundlagen nicht die fachspezifischen Publikationskulturen berücksichtigen bzw. widerspiegeln. Bibliometrische Daten, deren Verfügbarkeit, Auswahl und Interpretation, werden von bestehenden sozialen Konstruktionen beeinflusst und erfordern daher Reflexion, einen offenen Umgang mit alternativen Methoden, sowie das Fachwissen von Expertinnen und Experten.

Zur Unterstützung von Evaluierungsprozessen auf der Individual-Ebene (z.B. bei Berufungs- oder Tenure-Track-Verfahren) sind sinnvoll ausgewählte bibliometrische (z.B. feldspezifisch normierte) Indikatoren für die Bewertung des wissenschaftlichen Oeuvres der Kandidatinnen bzw. Kandidaten im Rahmen einer grobgranularen Vorauswahl hilfreich. Dies trifft besonders auf Kandidatinnen und Kandidaten zu, deren Promotion bereits mehrere Jahre zurückliegt. In den weitergehenden Stufen eines Evaluationsprozesses stellen quantitative, bibliometrische Indikatoren eine sinnvolle Ergänzung zu qualitativ analysierenden Bewertungen im Sinne eines *Informed Peer Review*-Verfahrens dar. Als primäre Grundlage für Entscheidungen sind sie jedoch nicht geeignet. Entscheidungsfindungen, z.B. für Stellenbesetzungen, sollten letztlich auf Basis einer ganzheitlichen, vielfältige qualitative Kriterien umfassenden Analyse der Leistungen und des Entwicklungspotentials der Kandidatinnen und Kandidaten in Forschung und Lehre unter angemessener Berücksichtigung unterschiedlicher Lebensphasen und Karrierestufen erfolgen.

Empfehlung

Der Wissenschaftsrat empfiehlt, bibliometrische Indikatoren als ein heuristisches Werkzeug zur Identifikation von Fragen und Problemfeldern sowie als Möglichkeit einer zusätzlichen Perspektive in Evaluierungsverfahren zu betrachten und zu nutzen.

Potentialanalyse und Unterstützung von Entwicklungsprozessen

Fortgeschrittene Methoden der Analyse bibliometrischer Daten ermöglichen neue Perspektiven auf die Forschungslandschaft, sowohl international, als auch auf der Ebene einzelner Länder und einzelner Institutionen. Dies setzt allerdings voraus, dass es gelingt, einen methodischen Bias bei der Auswahl und Aufbereitung der Daten weitgehend zu vermeiden.

Auf institutioneller Ebene (Fachbereiche, Universitäten, Forschungsinstitute) bzw. auf überregionaler Ebene können datenbasierte Analysemethoden, die auf der Auswertung bibliometrischer Indikatoren und deren Verknüpfung mit weiteren, im Forschungsmanagement verfügbaren Daten beruhen, hilfreiche Erkenntnisse für die strategische Entscheidungsfindung liefern. Dies betrifft beispielsweise vergleichende Analysen auf unterschiedlichen Aggregationsebenen (innerhalb einzelner Universitäten bzw. Regionen als auch überregional bzw. international) zu Schwerpunktbildungen, Vernetzungen von Forschungsaktivitäten und fächerübergreifenden Forschungskooperationen, Analysen zum Forschungsimpact und Input-Output-Relationen. Bibliometrische Analysen erlauben es, mögliche *blind spots* mit hoher Publikationsaktivität jedoch

geringem Wirkungsgrad, aber auch bislang unerkannte Potentiale und die Entwicklung emergenter Forschungsfelder zu erkennen.

Empfehlung

Der Wissenschaftsrat empfiehlt, Bibliometrie als Werkzeug zur Potentialanalyse von Forschungsleistung, emergenten Feldern und Netzwerken sowie zur Unterstützung institutioneller Selbstreflexion zu etablieren.

Kompetenzzentrum

Die Einrichtung eines nationalen Kompetenzzentrums für Forschungsevaluation, in welchem Fragestellungen zu Bibliometrie-basierten Verfahren untersucht und datenbasierte Analysen mit dem erforderlichen Know-how im Auftrag von Universitäten, Ministerien und Forschungsorganisationen durchgeführt werden können, würde eine Bereicherung der Forschungslandschaft in Österreich darstellen. Insbesondere für Forschungsförderungsorganisationen wäre eine solche universitätsunabhängige Institution hilfreich, um die Evaluation der Wirksamkeit von Forschungsprogrammen zu unterstützen.

Angesichts der finanziell angespannten Situation des Staatshaushaltes empfiehlt der Wissenschaftsrat, bestehende Einrichtungen und vorhandene Kompetenzen an österreichischen Universitäten zu nutzen, um ein solches Zentrum zu etablieren.

Empfehlung

Ein unabhängiges nationales Kompetenzzentrum für bibliometrische Analysen sollte unter Nutzung und Vernetzung vorhandener Expertisen eingerichtet werden.

7 Literatur

Altmetric, „Discovery and Resupply of Pharmacologically Active Plant-Derived Natural Products: A Review“ (2019) <<https://www.altmetric.com/details/4411153>> [zugegriffen 31. Juli 2019].

Bayer, Florian, Juan Gorraiz, Christian Gumpenberger, Lukas Mitterauer und Steve Reding, *Sichtbarkeitssteigerung in den Geistes-, Sozial- und Kulturwissenschaften (GSK): Ergebnisse einer Befragung an der Universität Wien* (Wien, 2017) <<https://doi.org/10.5281/ZENODO.400965>> [zugegriffen 10. Februar 2020].

Bornmann, Lutz, *Messung und Bewertung von Forschungsleistungen: Ein externes Gutachten für die Arbeitsgruppe des Österreichischen Wissenschaftsrates* (München, 2013).

Clarivate Analytics, „InCites: An Objective Analysis of People, Programs and Peers“ (2019) <<https://clarivate.com/webofsciencegroup/solutions/incites/>> [zugegriffen 4. September 2019].

Clarivate Analytics, *InCites Indicators Handbook* (Philadelphia, 2018) <<http://help.prod-incites.com/inCites2Live/8980-TRS/version/default/part/AttachmentData/data/InCites-Indicators-Handbook - June 2018.pdf>> [zugegriffen 4. September 2019].

CRISTIN, <<https://www.cristin.no/english/>> [zugegriffen 4. September 2019].

CWTS, *Bibliometric Analysis of Dutch University Medical Centers 2004-2015/2016* (Leiden, 2017) <https://www.nfu.nl/img/pdf/18.1111_CWTS_Bibliometric_study_on_Dutch_academic_medical_centres_2004-2015_2016.pdf> [zugegriffen 29. Jänner 2020].

CWTS, *Bibliometrics for Research Management and Research Evaluation: A Brief Introduction* (Leiden, 2018).

CWTS, „Leiden Ranking Listview 2019“ (Leiden, 2019a) <<https://www.leidenranking.com/ranking/2019/list>> [zugegriffen 1. August 2019].

CWTS, „University Profile Maps“ (Leiden, 2019b) <<https://www.vosviewer.com/university-profile-maps>> [zugegriffen 31. Juli 2019].

DORA, „Declaration on Research Assessment“ <<https://sfdora.org/>> [zugegriffen 4. September 2019].

Elsevier, „Pure“ (2020) <<https://www.elsevier.com/solutions/pure>> [zugegriffen 23. Jänner 2020].

euroCRIS, <<https://www.eurocris.org/>> [zugegriffen 4. September 2019].

Falagas, Matthew und Vangelis Alexiou, „The Top-Ten in Journal Impact Factor Manipulation“, *Archivum Immunologiae et Therapiae Experimentalis*, 56 (2008), S. 223–226 <<https://doi.org/10.1007/s00005-008-0024-5>> [zugegriffen 29. Juni 2020].

FWF, „Förderstatistiken: Altmetrics“ <<https://www.fwf.ac.at/de/ueber-den-fwf/foerderungsstatistiken/altmetrics/>> [zugegriffen 4. September 2019].

Garfield, Eugene, „Citation Indexes in Sociological and Historical Research“, *Essays of an Information Scientist*, 1 (1963), S. 43–46 <<http://garfield.library.upenn.edu/essays/V1p043y1962-73.pdf>> [zugegriffen 4. September 2019].

Giménez-Toledo, Elea, Carlos Tejada-Artigas und Jorge Mañana-Rodríguez, „Scholarly Publishers Indicators (SPI)“ (2014) <<http://ilia.cchs.csic.es/SPI/indexEn.html>> [zugegriffen 4. September 2019].

Gorraiz, Juan, Philip Purnell und Wolfgang Gläzel, „Opportunities for and Limitations of the Book Citation Index“, *Journal of the American Society for Information Science and Technology*, 64 (2013), S. 1388–1398 <<https://doi.org/10.1002/asi.22875>> [zugegriffen 10. Februar 2020].

Hammarfelt, Björn, „Beyond Coverage: Toward a Bibliometrics for the Humanities“, in *Research Assessment in the Humanities: Towards Criteria and Procedures*, hg. von Michael Ochsner, Sven E. Hug und Hans-Dieter Daniel (Zürich, 2016), S. 115–131 <https://doi.org/10.1007/978-3-319-29016-4_10> [zugegriffen 10. Februar 2020].

Harzing, Anne-Wil, Satu Alakangas und David Adams, „hla: An Individual Annual h-Index to Accommodate Disciplinary and Career Length Differences“, *Scientometrics*, 99 (2014), S. 811–821 <<https://doi.org/10.1007/s11192-013-1208-0>> [zugegriffen 10. Februar 2020].

Hasitzka, Katharina, Juan Gorraiz und Christian Gumpenberger, „Bibliometrie in Österreich: Ein neues Aufgabenfeld für Bibliotheken“, in *Universitätsbibliotheken im Fokus: Aufgaben und Perspektiven der Universitätsbibliotheken an öffentlichen Universitäten in Österreich*, hg. von Bruno Bauer, Christian Gumpenberger und Robert Schiller (Graz, 2013), S. 216–225 <<http://eprints.rclis.org/24293/>> [zugegriffen 4. September 2019].

Havemann, Frank, *Einführung in die Bibliometrie* (Berlin, 2016) <https://www.researchgate.net/publication/309230452_Einführung_in_die_Bibliometrie> [zugegriffen 4. September 2019].

Hesse, Friedrich W., Ina Matt, Falk Reckling, Thomas Völker und Nikolaus Possanner, *Determination of the State of Educational Research in Austria* (Wien, 2019) <<http://doi.org/10.5281/zenodo.3558401>> [zugegriffen 5. Dezember 2019].

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke und Ismael Rafols, „The Leiden Manifesto for Research Metrics“, *Nature*, 520 (2015), S. 9–11 <<https://doi.org/10.1038/520429a>> [zugegriffen 4. September 2019].

Hirsch, Jorge E., „Superconductivity, What the H? The Emperor has No Clothes“, *Physics and Society*, 49 (2020), S. 4–9 <<http://arxiv.org/abs/2001.09496>> [zugegriffen 23. April 2020].

Hirschauer, Stefan, „Peer Review Verfahren auf dem Prüfstand“, *Zeitschrift für Soziologie*, 33 (2002), S. 62–83 <https://www.researchgate.net/publication/305055760_Peer_Review_Verfahren_auf_dem_Pruftstand_Peer_Review_Research_-_Reviewed> [zugegriffen 4. September 2019].

Informatics Europe, „Joint Statement on Informatics Research Evaluation“ (2020) <<https://www.informatics-europe.org/news/545-joint-statement-on-informatics-research-evaluation.html>> [zugegriffen 22. Juli 2020].

Leitner, Karl-Heinz, *Messung und Bewertung von Forschungsleistung in Österreich: Externes Gutachten* (Wien, 2013).

Lücke, Nicole, „Wer auf den Titel gehört“, *DUZ Magazin*, 2 (2010) <<https://www.duz.de/beitrage/!/id/248/wer-auf-den-titel-gehoert>> [zugegriffen 5. August 2019].

Matthias, Lisa und Holger Breithaupt, *San Francisco Vereinbarung über die Forschungsbewertung* (o.D) <https://sfdora.org/wp-content/uploads/2018/09/DORA_German.pdf> [zugegriffen 9. Juli 2020].

Marx, Werner und Lutz Bornmann, „Bibliometrische Verfahren zur Bewertung von Forschungsleistung: Hintergründe, Aussagekraft und Grenzen“, *Soziale Welt*, 66 (2015), S. 161–179 <https://www.researchgate.net/publication/293027179_Bibliometrics_in_research_evaluation_-_Background_significance_and_limitations> [zugegriffen 4. September 2019].

National Institute of Health, „iCite“ <<https://icite.od.nih.gov/>> [zugegriffen 6. September 2019].

Niederklapfer, Thomas, *Der Hirsch-Index und seine Varianten: Eine Abhandlung zur Theorie und Praxis* (Innsbruck, 2014) <<https://diglib.uibk.ac.at/ulbtirolhs/content/titleinfo/216431>> [zugegriffen 6. September 2019].

Noyons, Ed, *Advanced Bibliometric Analysis in the Context of Research Evaluation - Part 1: Strength, Promising & Risk Areas (SPRA)*, Vortrag für den Österreichischen Wissenschaftsrat (Wien, 2019a).

Noyons, Ed, *Bibliometric Analysis of the Austrian Computer Science Research* (unveröffentlichter Draft) (Leiden, 2019b).

Noyons, Ed, „Bibliometric Analyses of the Austrian Computer Science Research“, Präsentation bei *Informatik in Österreich: Perspektiven und Strategien* (Wien, 2019c).

ÖAW, *Qualitätskriterien für Publikationen in den Geistes-, Sozial- und Kulturwissenschaften: Bericht der AD-Hoc-Kommission der Phil.-Hist. Klasse der ÖAW* (Wien, 2012) <<https://www.oeaw.ac.at/fileadmin/NEWS/2017/PDF/Richtlinie-Praesentation.pdf>> [zugegriffen 4. September 2019].

Österreichischer Wissenschaftsrat, *Die Vermessung der Wissenschaft: Messung und Beurteilung von Qualität in der Forschung* (Wien, 2014) <https://www.wissenschaftsrat.ac.at/downloads/Empfehlungen_Stellungnahmen/2015_2010/Messung_Endversion_inkl-Cover.pdf> [zugegriffen 4. September 2019].

Österreichischer Wissenschaftsrat, *Klinische Forschung in Österreich: Stellungnahme und Empfehlungen* (Wien, 2016a) <https://www.wissenschaftsrat.ac.at/downloads/Empfehlungen_Stellungnahmen/2018_2016/Endversion_Klinische-Forschung-in-Oesterreich.PDF> [zugegriffen 29. Jänner 2020].

Österreichischer Wissenschaftsrat, *Klinische Forschung in Österreich: Stellungnahme und Empfehlungen (Anhang 2)* (Wien, 2016b) <https://www.wissenschaftsrat.ac.at/downloads/Empfehlungen_Stellungnahmen/2018_2016/Endversion-Anhang_2.pdf> [zugegriffen 29. Jänner 2020].

Österreichischer Wissenschaftsrat, *Tagungsbericht zur Veranstaltung „Qualitätsstandard oder leeres Ritual? Begutachtungen in der Diskussion“* (Wien, 2018), <<https://www.wissenschaftsrat.ac.at/downloads/tagungsbericht-zur-veranstaltung-qualitaetsstandard-oder-leeres-ritual-begutachtungen-in-der-diskussion-am-8.-und-9.-november-2018-in-wien.pdf>>[zugegriffen 3. August 2020].

Österreichischer Wissenschaftsrat, „*Informatik in Österreich: Perspektiven und Strategien*“ (Wien, 2019), <<https://www.wissenschaftsrat.ac.at/blog/informatik-in-oesterreich-perspektiven-und-strategien>> [zugegriffen 3. August 2020].

Pritchard, Alan, „Statistical Bibliography or Bibliometrics?“, *Journal of Documentation*, 25 (1969), S. 348–349 <https://www.researchgate.net/publication/236031787_Statistical_Bibliography_or_Bibliometrics> [zugegriffen 4. September 2019].

Schweizerischer Wissenschafts- und Technologierat, *Leistungsmessung und Qualitätssicherung in der Wissenschaft* (Bern, 2013) <https://www.swir.ch/images/stories/pdf/de/SWTR_Schrift_3_2013_D_2_Auflage_Qualitaetssicherung_web.pdf> [zugegriffen 4. September 2019].

Taubert, Niels Christian, „Bibliometrie in der Forschungsevaluation. Zur Konstitution und Funktionslogik wechselseitiger Beobachtung zwischen Wissenschaft und Politik“, in *Quoten, Kurven und Profile. Zur Vermessung der sozialen Welt.*, hg. von Jan Passoth und Joseph Wehner (Wiesbaden: Springer VS, 2013), S. 179–204 <<https://pub.uni-bielefeld.de/record/2393266>> [zugegriffen 10. Februar 2020].

Taubert, Niels Christian und Peter Weingart, „Wandel des wissenschaftlichen Publizierens – eine Heuristik zur Analyse rezenter Wandlungsprozesse“, in *Wissenschaftliches Publizieren - Zwischen Digitalisierung, Leistungsmessung, Ökonomisierung und medialer Beobachtung*, hg. von Peter Weingart und Niels Taubert (Berlin, 2016), S. 3–38 <<https://pub.uni-bielefeld.de/record/2904656>> [zugegriffen 10. Februar 2020].

Tunger, Dirk, Marcel Clermont und Andreas Meier, „Altmetrics: State of the Art and a Look into the Future“, in *Scientometrics*, hg. von Mari Jibu und Yoshiyuki Osabe (London, 2018), S. 123–134 <<https://www.intechopen.com/chapter/pdf-download/61110>> [zugegriffen 4. September 2019].

Universität Siegen, „European Quality Audit“ (2019) <https://www.uni-siegen.de/start/die_universitaet/qualitaetsmanagement/european_quality_audit/?lang=de> [zugegriffen 4. September 2019].

VHB, „VHB-JOURQUAL3“ (2019) <<https://vhbonline.org/vhb4you/vhb-jourqual/vhb-jourqual-3>> [zugegriffen 10. Februar 2020].

de Vrieze, Jop, „Funders Groan under Growing Review Burden“, *Science*, 357 (2017), S. 343 <<https://doi.org/10.1126/science.357.6349.343>> [zugegriffen 4. September 2019].

Waltman, Ludo, „Comparing Scientific Performance across Disciplines: Methodological and Conceptual Challenges“, in *7th International Conference on Information Technologies and Information Society (ITIS2015)* (Novo Mesto, 2015) <<https://de.slideshare.net/LudoWaltman/comparing-scientific-performance-across-disciplines-methodological-and-conceptual-challenges>> [zugegriffen 4. September 2019].

Waltman, Ludo und Nees Jan van Eck, „The Inconsistency of the h-Index“, *Journal of the American Society for Information Science and Technology*, 63 (2012), S. 406–415 <<https://doi.org/10.1002/asi.21678>> [zugegriffen 29. Juni 2020].

Wilsdon, James, Liz Allen, Eleonora Belfiore, Philip Campbell, Stephen Curry, Steven Hill, u. a., *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management* (2015) <<https://www.researchgate.net/publication/279402178> The Metric Tide Report of the Independent Review of the Role of Metrics in Research Assessment and Management> [zugegriffen 4. September 2019].

Wissenschaftsrat, *Begutachtungen im Wissenschaftssystem: Positionspapier* (Berlin, 2017) <<https://www.wissenschaftsrat.de/download/archiv/6680-17.pdf;jsessionid=133F324345FCA6D7A015852B868926DC.delivery1-master?blob=publicationFile&v=2>> [zugegriffen 4. September 2019].

8 Anhänge

1. CWTS, *Bibliometrics for Research Management and Research Evaluation: A Brief Introduction* (Leiden, 2018)
2. DORA: Matthias, Lisa und Holger Breithaupt, *San Francisco Vereinbarung über die Forschungsbewertung (o.D)*
3. “Responsible Metrics” in Wilsdon, James u. a., *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management* (2015), S. 134-135
4. Fragebogen des ÖWR an österreichische Rektorate
5. Internationaler Fragebogen des ÖWR

Anhang 1: CWTS, *Bibliometrics for Research Management and Research Evaluation: A Brief Introduction* (Leiden, 2018)



Bibliometrics for Research Management and Research Evaluation

A Brief Introduction



Universiteit
Leiden



Bibliometrics for Research Management and Research Evaluation

*This manual was written on behalf of the Austrian
Science Board*

Ludo Waltman

Ed Noyons

February, 2018

CWTS BV
P.O. Box 905
2300 AX Leiden, The Netherlands
Tel. +31 71 527 3909
Fax +31 71 527 3911
E-mail info@cwts.leidenuniv.nl

**ÖSTERREICHISCHER
WISSENSCHAFTSRAT**

Table of contents

1 Introduction.....	4
2 Bibliometric data sources.....	6
3 Bibliometric analyses	8
4 Bibliometric indicators of scientific impact	12
5 Responsible use of bibliometrics.....	19
6 Examples of bibliometric analyses.....	22

1 Introduction

Bibliometrics is the quantitative analysis of written documents. It is frequently used to analyze scientific publications. Researchers may for instance use bibliometrics to get an overview of their research field and of the connections with neighboring research areas. Bibliometrics is also often used for research management and research evaluation purposes. In this manual, we discuss the application of bibliometrics in this context. Our institute, the Centre for Science and Technology Studies (CWTS) at Leiden University, has a longstanding experience with the use of bibliometrics to support research management and research evaluation (see Box 1).

Research management involves complex decision making, for instance about setting research priorities and allocating resources. Bibliometrics offers quantitative information that can be used to support decision making in a research management context. This information is often provided using numerical indicators (sometimes also referred to as ‘metrics’, although in this manual we will not use this term). The journal impact factor and the h-index are probably the best-known examples of such indicators. However, bibliometric information can also be made available in other ways, for instance by presenting lists of important publications or by presenting information visually in graphs and other types of visualizations.

Box 1. Centre for Science and Technology Studies (CWTS)

The Centre for Science and Technology Studies (CWTS) is an interdisciplinary research institute at Leiden University that studies scientific research and its connections to technology, innovation, and society. CWTS is a world-leading center of expertise on bibliometrics and scientometrics. It extensively studies the use of bibliometrics and scientometrics for research management and research evaluation purposes.

CWTS BV is a company owned by Leiden University that offers bibliometric and scientometric analyses to support research management and research evaluation. Analyses performed by CWTS BV are for instance used by universities and other research organizations, research funders, governments, and scientific publishers, mostly in Europe but also elsewhere in the world. CWTS BV also provides consultancy and training in the areas of bibliometrics, scientometrics, and research evaluation.

For more information, please visit www.cwts.nl and www.cwtsbv.nl.



Bibliometric indicators

A bibliometric indicator provides a quantitative proxy of a concept of interest. Indicators based on citation counts are frequently used. These indicators are typically interpreted as proxies of scientific impact. They do not provide exact measurements of scientific impact, but they do offer approximate information about the scientific impact of publications, researchers, or research institutions. Indicators of scientific impact are the best-known bibliometric indicators, but bibliometric indicators can also be used to provide other types of information, for instance related to scientific output, scientific collaboration, mobility, interdisciplinarity, gender, and open access publishing.

When using bibliometric indicators, it is essential to keep in mind that indicators usually do not exactly measure the concept of interest, but only provide approximate information. This is one of the reasons why important research management decisions should not be made based solely on bibliometric indicators. Bibliometric indicators provide relevant information that is often helpful to support decision making. However, for most decisions the information provided by bibliometric indicators is too limited to rely exclusively on this information.

Beyond bibliometrics

Bibliometric information can be seen as one element within a broader range of information sources available to support decision making in a research management context. Relevant information for instance can also be obtained from peer review and from other quantitative sources, for instance data on research funding and research staff and altmetric data. However, a discussion of these information sources is outside the scope of this manual.

Organization of this manual

This manual offers a short introduction into the use of bibliometrics for research management and research evaluation purposes. In Chapter 2, the most important data sources available for bibliometric analyses are discussed. In Chapters 3 and 4, an overview of different types of bibliometric analyses is provided, focusing in particular on analyses of scientific impact. Responsible use of bibliometrics is discussed in Chapter 5.

2 Bibliometric data sources

Bibliometric analyses require a bibliometric data source. Simple analyses sometimes can be carried out using an internal data source of a research institution, for instance using data extracted from an internal information system of a university. In most cases, however, an external data source is needed. Web of Science, produced by Clarivate Analytics, and Scopus, produced by Elsevier, are the two most commonly used bibliometric data sources. Google Scholar is also used quite frequently. In addition, bibliometric analyses that are restricted to particular scientific fields sometimes rely on field-specific data sources, such as INSPIRE, MathSciNet, PsycINFO, and PubMed. Interesting new data sources that can potentially be of value for bibliometric analyses are Microsoft Academic and Crossref. However, at the moment, the use of these data sources is still in an early stage and therefore experience with these data source is still quite limited.

Web of Science and Scopus

Web of Science consists of information about a number of citation indices, for example *the Science Citation Index, Social Science Citation Index and Arts & Humanities Citation Index*. a number of citation indices. The most commonly used citation indices are part of the so-called Web of Science Core Collection. This includes the Science Citation Index Expanded, the Social Sciences Citation Index, the Arts & Humanities Citation Index, and the Emerging Sources Citation Index. These indices cover scientific journals in the sciences, social sciences, and arts and humanities. In addition, the Conference Proceedings Citation Index and the Book Citation Index, which cover conference proceedings and books, are also included in the Web of Science Core Collection. Like Web of Science, Scopus covers journals, conference proceedings, and books in the sciences, social sciences, and arts and humanities, although no distinction between different citation indices is made in Scopus.

Web of Science and Scopus both require a subscription. Subscribers have access to the data sources through a web interface. These web interfaces can be used to search for scientific literature and also to perform simple small-scale bibliometric analyses. However, large-scale bibliometric analyses cannot be carried out using these web interfaces. Such analyses can be performed using InCites and SciVal, commercial bibliometric analysis tools that make use of, respectively, Web of Science and Scopus data. Large-scale bibliometric analyses can also be performed based directly on Web of Science or Scopus data, but this requires a special license that provides direct access



to the data. CWTS has a license that provides direct access to Web of Science core collection data for carrying out large-scale bibliometric analyses.

CWTS uses its Citation Index system (CI-system) for bibliometric analyses. The core of this system comprises an enhanced version of Web of Science core collection. This means that CWTS runs its own algorithms for accurate citation matching and filtering for self-citations. In addition, CWTS builds its own unification system for name- and address variants of institutions.

The sources run back to the 1980s, are updated quarterly, and contain over 50 million publications. A combination of smart computer algorithms and manual data cleaning ensures a better unification of the names and addresses of universities and other organisations. CWTS is working continuously on expanding its data system with other major scientific databases, such as PATSTAT (patent analysis) and Altmetrics (social media, blogs, newsfeeds etc.). CWTS has more than twenty-five years of experience in working with these data.

Google Scholar

Google Scholar is an online search engine for scientific literature. It can also serve as a data source for bibliometric analyses. Compared with Web of Science and Scopus, Google Scholar has two key advantages. One advantage is that Google Scholar is freely available. No subscription is required. The other advantage is that Google Scholar offers a more comprehensive coverage of the scientific literature. This is important especially for bibliometric analyses in computer science, the social sciences, and the humanities and also for bibliometric analyses that focus not only on international scientific literature but also on national and regional literature. For such analyses, Web of Science and Scopus provide only a limited coverage, while the coverage of Google Scholar is more comprehensive.

However, Google Scholar also has major limitations. First of all, it lacks transparency. Very little information is available on what is covered by Google Scholar and what is not. Second, Google Scholar data is of lower quality than Web of Science and Scopus data. Web of Science and Scopus data is of significantly higher quality than data in Google Scholar (see Section 3: <https://doi.org/10.1016/j.joi.2016.02.007>). Google Scholar data can also be manipulated more easily than Web of Science and Scopus data. Finally, it is almost impossible to use Google Scholar for large-scale bibliometric analyses. Google Scholar does not provide large-scale access to its data.

3 Bibliometric analyses

Bibliometrics can be used for a broad range of problems in research management and research evaluation. It can for instance be used to support strategic decision making by the management of a research institution, but also to support the assessment of candidates for a scientific position. Another application of bibliometrics is to assist peer reviewers in the evaluation of a research unit (e.g., a researcher, research group, department, or research institution), resulting in so-called informed peer review. Bibliometrics is also sometimes used to define indicators that can be used for allocating research funding.

Types of bibliometric analyses

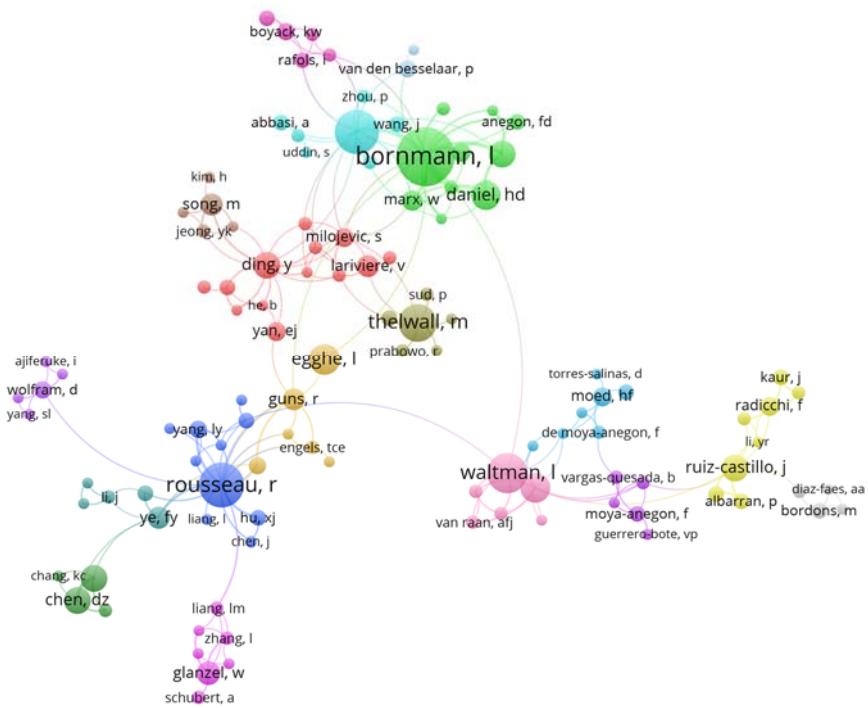
Depending on the context in which bibliometrics is applied, different types of bibliometric analyses can be used. Strategic decision making at the level of an entire research institution requires a high-level overview that may be best served by a limited set of numerical indicators. Decision making at lower levels within an organization may benefit from richer bibliometric analyses in which numerical indicators are complemented with other more in-depth types of bibliometric information, for instance by providing bibliometric visualizations. Likewise, when comparing candidates for a scientific position, it may be helpful to have a comprehensive bibliometric profile of each candidate, presenting not only numerical indicators but also linking these indicators to the underlying bibliometric information. For instance, when reporting the number of highly cited publications of a candidate, such a profile could also present a list of these publications, thereby enabling a more substantive evaluation of the publications.

Bibliometric analyses provide quite some flexibility in the level of detail at which information is presented, making it possible to adjust the level of detail to the purpose of the analysis. In some cases, a bibliometric analysis presents just a limited number of numerical indicators. Typically, however, more detailed information is provided by also presenting time trends and breakdowns by scientific field. Such information is often presented visually in graphs. Visual approaches to presenting bibliometric information can also be used to offer a geographical perspective (e.g., bibliometric indicators at the level of countries or institutions presented in a geographical map) or a network perspective (e.g., co-authorship networks or citation networks, visualized for instance using the VOSviewer software developed at CWTS; see Box 2).



More detailed information can also be presented by showing the distribution underlying a bibliometric indicator, for instance the distribution of citations over publications (see for example <https://doi.org/10.1101/062109>). Even more detail can be provided by listing publications that are of special interest, for instance publications that are highly cited or publications that have appeared in reputable journals. Furthermore, when dealing with large numbers of publications, a high-level overview of the scientific content of the publications can be obtained by presenting a visualization of the most important terms occurring in the titles and abstracts of the publications. Using overlay visualizations, it is possible to include additional information in such a visualization, for instance to compare the focal research areas of different research units (for detailed examples, please visit <http://www.vosviewer.com/features/examples>).

Box 2. Visualization of bibliometric networks in the field of scientometrics using VOSviewer



VOSviewer is a popular software tool developed by CWTS for visualizing bibliometric networks, such as co-authorship networks or citation networks. VOSviewer is freely available at www.vosviewer.com. The above figure shows an example of a VOSviewer visualization of a co-authorship network. Each circle represents a researcher. The size of a circle indicates the number of publications of a researcher. Lines indicate co-authorship links between researchers. Finally, colors designate clusters of researchers that are strongly connected to each other by co-authorship links.

Information provided by bibliometric analyses

Important information that can be provided by a bibliometric analysis includes:

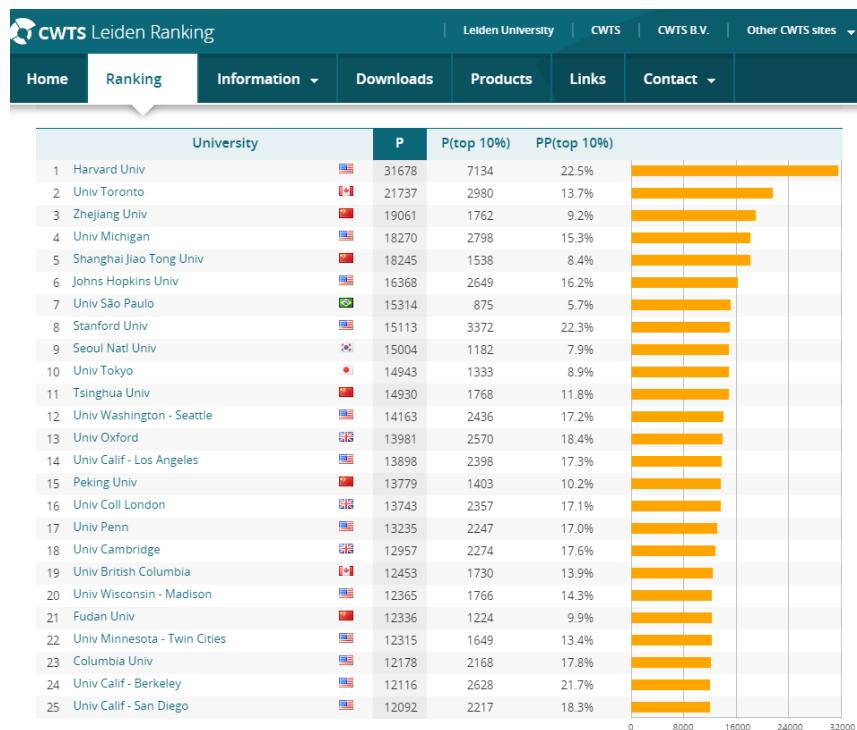
- *Scientific output.* Information about the number of publications produced by a research unit.
- *Scientific impact.* Information about the number of citations that publications of a research unit have received.
- *Scientific collaboration.* Information about co-authored publications, focusing for instance on national and international collaboration or on university-industry collaboration.
- *Mobility.* Information about researchers that change their affiliation.
- *Interdisciplinarity.* Information about the field classification of publications.

- *Gender.* Information about the gender of the authors of publications.
- *Open access publishing.* Information about the open access status of publications, distinguishing for instance between gold open access, green open access, and no open access.

In Box 3, we offer a brief discussion of the CWTS Leiden Ranking, a ranking of universities based on bibliometric indicators. The Leiden Ranking provides an illustration of the use of indicators of scientific output, impact, and collaboration.

Because bibliometric indicators of scientific impact play a prominent role in many research evaluations, we discuss these indicators in more detail in the next chapter.

Box 3. CWTS Leiden Ranking



The CWTS Leiden Ranking is a bibliometric ranking of about 900 major universities worldwide. The Leiden Ranking is produced annually by CWTS based on data from Web of Science. It offers bibliometric indicators of scientific output, impact, and collaboration. A breakdown of the bibliometric statistics is provided at the level of five broad fields of science, and the statistics are presented in three different ways, referred to as the list view, the chart view, and the map view. The Leiden Ranking is available at www.leidenranking.com.

4 Bibliometric indicators of scientific impact

Scientific impact is typically analyzed by counting the number of citations received by publications. There are many different impact indicators, with the journal impact factor and the h-index being the best-known examples. Citations are given for a variety of reasons. Some citations indicate that the citing publication builds on the cited publication. These citations may be seen as an acknowledgment of the importance of the cited publication for the citing one. Negative citations are of an opposite nature. They reflect a critical perspective of the citing publication on the cited publication. However, many citations are neither positive nor negative. These citations often reflect a more superficial connection between the citing and the cited publication. They are sometimes referred to as superficial citations. Given the diversity of citations, citation counts provide only an approximate indication of scientific impact. In some cases, citation counts are also interpreted as indicators of scientific quality, but this interpretation is of an even more approximate nature.

Types of impact indicators

Impact indicators can be classified in many different ways. An essential distinction is between size-dependent and size-independent impact indicators. Size-dependent impact indicators reflect the total scientific impact of the publications of a research unit, while size-independent impact indicators reflect the average scientific impact per publication. A further distinction that can be made is between impact indicators that directly count citations and impact indicators that first identify highly cited publications and then count these publications.

Table 1 summarizes the above classification of impact indicators. The different types of impact indicators are labeled based on the names that are used at CWTS. The simplest impact indicators are based directly on counting citations. These are the size-dependent total citation score and the size-independent mean citation score. The total citation score equals the total number of citations received by the publications of a research unit, while the mean citation score equals the average number of citations received per publication. Impact indicators based on counting highly cited publications first require the choice of the threshold n that determines whether a publication is classified as highly cited or not. A publication is classified as highly cited if it has received at least n citations. The value of n for instance could be 10, 20, 50, or 100,

depending on how strict one would like to be in classifying publications as highly cited. After the threshold n has been chosen, the number and the proportion of highly cited publications of a research unit can be calculated. These indicators provide respectively a size-dependent and a size-independent perspective on the scientific impact of the publications of a research unit.

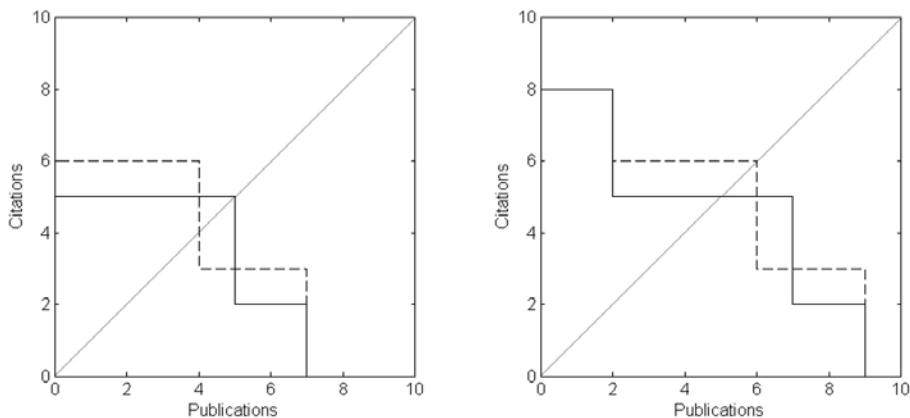
Table 1. Classification of impact indicators.

	Size-dependent	Size-independent
Citations	Total citation score TCS	Mean citation score MCS
Highly cited publications	Number of highly cited pub. $P(\geq n \text{ cit.})$	Prop. of highly cited pub. PP($\geq n \text{ cit.}$)

Compared with impact indicators based on counting citations, impact indicators based on counting highly cited publications are less sensitive to publications with a very large number of citations. Impact indicators based on counting highly cited publications are therefore more robust, which is often seen as an advantage of these indicators.

The classification of impact indicators presented in Table 1 includes some of the most commonly used indicators, but many other indicators are not included. An example is the h-index. The h-index of a research unit equals the largest number h such that the research unit has h publications that have received at least h citations each. The h-index is a size-dependent impact indicator. Like the number of highly cited publications, it is relatively insensitive to publications with a very large number of citations. The h-index is normally not used by CWTS. This is because of its inconsistency. When two research units make the same improvement in terms of publications and citations, their ranking relative to each other according to the h-index may reverse. The inconsistency of the h-index is illustrated in Box 4.

Box 4. Inconsistency of the h-index



The above figures illustrate the inconsistency of the h-index. The solid and the dashed line indicate the number of citations received by the publications of, respectively, research unit A and research unit B. Publications are presented in decreasing order of their number of citations. The left figure shows the initial situation, in which the h-index of research units A and B equals, respectively, 5 and 4. The right figure shows the situation after research units A and B have both published 2 new publications, each with 8 citations. In the new situation, the h-index of research units A and B equals, respectively, 5 and 6. Hence, compared with the initial situation, the two research units have made the same improvement in terms of publications and citations, but their ranking relative to each other according to the h-index has reversed. This is an illustration of the inconsistency of the h-index.

Normalization for scientific field and publication age

Different scientific fields have different citation practices. Because of this, there are large differences between fields in citation density, that is, in the average number of citations received per publication. For instance, the average number of citations received by publications in mathematics is about an order of magnitude smaller than the average number of citations received by publications in some fields in the life sciences. When a bibliometric analysis of scientific impact covers multiple fields, one often wants to correct for differences between fields in citation density. Performing such a correction is called field normalization. Field normalization is usually carried out by comparing the number of citations of a publication with the number of citations of other publications in the same field.

Older publications have had more time to receive citations than more recent publications, and on average older publications therefore tend to have been cited more often than more recent publications. One may want to perform a normalization that corrects for this. Such a normalization can be carried out by comparing the number of

citations of a publication with the number of citations of other publications from the same year.

Table 2 presents the normalized counterparts of the unnormalized impact indicators listed in Table 1. The indicators are again labeled based on the names that are used at CWTS. The total and the mean normalized citation score equal, respectively, the total and the average normalized number of citations of the publications of a research unit. The normalized number of citations of a publication is calculated by dividing the number of citations of the publication by the average number of citations of all publications in the same field and from the same year. In the case of impact indicators based on counting highly cited publications, a publication is classified as highly cited if it belongs to the top $x\%$ most highly cited publications of its field and its year. One often focuses on the top 10% most highly cited publications, but it is for instance also possible to consider the top 1%, top 5%, or top 50% most highly cited publications.

Table 2. Classification of normalized impact indicators.

	Size-dependent	Size-independent
Citations	Total normalized citation score TNCS	Mean normalized citation score MNCS
Highly cited publications	Number of highly cited pub. P(top $x\%$)	Prop. of highly cited pub. PP(top $x\%$)

The use of normalized impact indicators involves some choices. Normalization for scientific field requires the choice of a field classification system. In practice, the journal categories in Web of Science and Scopus are often used as a field classification system. An alternative is to define fields algorithmically at the level of individual publications instead of journals. This can for instance be done by grouping publications into fields based on citation relations. Normalization based on an algorithmically constructed publication-level classification system can be expected to yield more accurate results, while normalization based on a journal-level classification system may be more transparent and easier to understand.

CWTS works preferably field its proprietary publication-level classification system. This is particularly relevant in our work for the Austrian Science Board, because earlier studies have shown that citation analysis may underestimate the impact of, for

instance, clinical research as compared to basic research (Van Eck [2013] PLoS ONE, 8(4), e62395).

Another choice that needs to be made relates to the minimum age of publications that are included in the calculation of normalized impact indicators. Very recent publications usually have received no or almost no citations. Normalization for publication age does not give meaningful results for these publications. Very recent publications are therefore often excluded from the calculation of normalized impact indicators. Typically at least publications that are less than one year old are excluded.

Credit allocation

In most scientific fields, a large majority of the publications are co-authored by multiple researchers and often also by multiple research institutions or even by multiple countries. This leads to the problem of credit allocation. When a publication is co-authored by multiple research units, how should the credits of the publication be allocated to the different research units?

The two most commonly used approaches for dealing with the credit allocation problem are referred to as the full and the fractional counting approach. In the full counting approach, the credits of a publication are fully allocated to each of the co-authoring research units. In the fractional counting approach, the credits of a publication are fractionally allocated to each of the co-authoring research unit. Hence, in the case of a publication co-authored by three research units, each unit receives one-third of the credits of the publication.

Table 3 shows a simple example illustrating the full and the fractional counting approach in the calculation of the total and the mean citation score of a research unit. We are interested in research unit A. This research unit has authored three publications. It is the only author of publication 1, while it has co-authored publications 2 and 3 with other research units. For each publication, Table 1 reports the number of citations received by the publication. In the full counting approach, the three publications and their citations are fully assigned to research unit A (and to the other research units B and C). This results in a total and a mean citation score of, respectively, 17 and $17 / 3 = 5.67$ for research unit A. In the fractional counting approach, publications and citations are allocated fractionally to research unit A. Consider for instance publication 2. As can be seen in Table 3, this publication is co-authored by two research units, A and B, and therefore the publication is allocated to research unit A with a weight of $1 / 2 = 0.50$. The publication has received 3 citations,

which are allocated to research unit A with a weight of 0.50, yielding $0.50 \times 3 = 1.50$ citations for research unit A. By performing these calculations for all three publications, the fractional counting approach results in a total citation score of 10.17 and a mean citation score of $10.17 / 1.83 = 5.55$.

Table 3. Example illustrating the full and the fractional counting approach.

	Co-authoring research units	No. of citations	Fractional publication allocation	Fractional citation allocation
Publication 1	A	6	1.00	6.00
Publication 2	A, B	3	0.50	1.50
Publication 3	A, B, C	8	0.33	2.67
Total		17	1.83	10.17

The choice between the full and the fractional counting approach can be made based on which approach one considers more appropriate for a particular analysis. However, when working with normalized impact indicators, the fractional counting approach has an important advantage over the full counting approach. Using the fractional counting approach, normalized impact indicators correct not only for differences between fields in citation density but also for differences between fields in collaboration practices.

In addition to the full and the fractional counting approach, there are also other approaches that can be taken to deal with the credit allocation problem. Most of these approaches rely on the order of the authors in the author list of a publication. They for instance allocate the credits of a publication mostly, or even exclusively, to the first or the last author of a publication, or sometimes to the corresponding author. However, there are no universal norms that determine the order of the authors in the author list of a publication, and relying on this order therefore always involves some uncertainty. Most importantly, different fields have different practices for determining the order of authors. In some fields, in particular in economics, high energy physics, and mathematics, it is common practice to order authors alphabetically. Credit allocation clearly should not be based on the order of authors in these fields.

Author self-citations

Author self-citations are often excluded from the calculation of impact indicators. Many self-citations are given for perfectly valid reasons. Nevertheless, if self-citations

are not excluded from the calculation of impact indicators, they can be used to manipulate the indicators in a relatively easy way. To prevent such manipulation, CWTS excludes self-citations by default in its analyses.

Impact indicators for journals

Indicators of scientific impact can also be calculated for journals. The best-known example of an impact indicator for journals is the journal impact factor. The impact factor of a journal essentially equals the average number of citations received in a certain year by publications that appeared in the journal in the two preceding years. The journal impact factor, calculated by Clarivate Analytics based on Web of Science data, is published in the Journal Citation Reports. Other examples of impact indicators for journals are the 5-year impact factor and the article influence score, which are also published in the Journal Citation Reports, and CiteScore, SJR, and SNIP, which are made available by Elsevier based on Scopus data.

Journal impact indicators are often used not only for evaluating journals but also for evaluating individual publications in a journal, or the research units by which individual publications have been authored. This is a controversial way of using journal impact indicators. Within a journal, there are typically large differences between publications in the number of citations they receive, and therefore it is often considered inappropriate to use a journal impact indicator for evaluating individual publications in a journal. The San Francisco Declaration on Research Assessment (www.ascb.org/dora/), an influential statement that has been signed by a large number of individuals and organizations, for instance rejects the use of journal impact indicators at the level of individual publications.

CWTS does not reject the use of journal impact indicators for evaluating individual publications. Journals have different quality standards, and this is partly reflected by journal impact indicators. Hence, journal impact indicators can be used for evaluating individual publications because they provide information about the quality standards of the journals in which publications have appeared. In addition to the journal impact indicators mentioned above, other journal impact indicators can be used as well. For instance, the mean normalized citation score, discussed earlier in this chapter, can be calculated for journals, and the publications of a research unit can then be evaluated using the mean normalized citation scores of the journals in which they have appeared. This results in an indicator that is referred to as the mean normalized journal score at CWTS.

5 Responsible use of bibliometrics

Bibliometrics offers powerful tools to support research management and research evaluation. However, these tools should be used in a responsible way. The following eight principles can be used as guidelines for the responsible use of bibliometrics.

1 Be aware of the limited coverage of bibliometric data sources

Bibliometric data sources such as Web of Science and Scopus offer only a limited coverage of the scientific literature. International journals in the sciences are typically well covered, but coverage of national journals, journals in the social sciences and humanities, and conference proceedings and books is much more limited. One should be aware of the consequences of this for bibliometric analyses.

2 Acknowledge the importance of accurate data collection

Accurate data collection is essential for high-quality bibliometric analyses. Poor data collection may result in bibliometric analyses that provide incorrect or misleading conclusions. The efforts needed for accurate data collection should not be underestimated. In a research evaluation context, the units under evaluation should have the opportunity to verify the data collection.

3 Recognize that bibliometric analyses capture research performance only in a partial manner

It is essential to recognize that bibliometric analyses reflect only specific aspects of the performance of research units. For instance, citation statistics provide insight into the scientific impact of research units, but they do not capture the broader societal impact. Also, bibliometric analyses focus on the outputs of the research process and typically do not consider the inputs. This for instance means that bibliometric analyses provide no insight into the productivity of research units.

4 Account for differences between scientific fields in publication, authorship, and citation practices

Different scientific fields have different publication, authorship, and citation practices. Publications in the life sciences for instance tend to have more authors and tend to receive more citations than publications in the social sciences. In bibliometric analyses that extend over multiple fields, differences between fields should be carefully accounted for. This can be done by normalizing bibliometric indicators for field

differences or by comparing research units with relevant benchmarks active in the same field.

5 Find an appropriate balance between transparency and analytical sophistication

Transparency of a bibliometric analysis helps to ensure that the analysis is interpreted correctly and facilitates a well-informed discussion about the outcomes of the analysis. Analytical sophistication, for instance the use of advanced field-normalized bibliometric indicators, has the potential to provide insights that are hard to obtain using more straightforward bibliometric approaches. Unfortunately, an increase in analytical sophistication often causes a decrease in transparency. Bibliometric analyses therefore require a careful trade-off between transparency and analytical sophistication.

6 Embrace the value of multidimensional and contextualized bibliometrics

Bibliometrics is sometimes used as a tool for making one-dimensional performance measurements. When bibliometrics is used in this way, its value is limited. To take full advantage of bibliometric information, a multidimensional and contextualized approach to bibliometrics needs to be adopted. Such an approach recognizes that research management and research evaluation benefit from being supported by diverse types of information. It also recognizes that bibliometric information needs to be contextualized (e.g., by explicitly linking the information to the underlying data) to enable in-depth interpretation of the information.

7 Use bibliometrics as part of a broader range of information sources

Bibliometrics offers just one source of information to support research management and research evaluation. There are other sources of information as well. In addition to quantitative sources, for instance data on research funding and research staff and altmetric data (e.g., data on Mendeley, Twitter, blogs, etc.), this also includes peer review by scientific experts. The use of bibliometrics should be considered within this broader context. The best way to support research management and research evaluation typically is to combine bibliometric information with other information sources.

8 Anticipate the effects of bibliometric analyses on the science system

The use of bibliometrics for research management and research evaluation purposes is likely to influence the behavior of researchers and other actors in the science system. These actors may change their behavior both in intended and in unintended ways. It



is important to anticipate these effects of the use of bibliometrics and to be aware that a strong reliance on bibliometrics may have undesirable consequences (e.g., researchers trying to improve their citation statistics in questionable ways).

Leiden Manifesto

The above principles for the responsible use of bibliometrics have partly been derived from the Leiden Manifesto, an influential statement presenting best practice guidelines for the proper use of numerical indicators in research evaluations. The Leiden Manifesto is available at www.leidenmanifesto.org.

6 Examples of bibliometric analyses

CWTS has ample experience in providing bibliometric analyses in the field of (bio)medical health sciences. Two examples and the links to the accompanying reports are provided below.

Every year, CWTS conducts a bibliometric benchmark analysis of the eight University Medical Centers (UMCs) in the Netherlands for the Netherlands Federation of University Medical Centers (NFU). The NFU represents the eight cooperating UMCs in the Netherlands. The objective of the NFU is to ensure that agencies that decide on healthcare issues in the Netherlands take into account the special role of the UMCs. Furthermore, the UMCs have their own governance code that takes into account the specific responsibilities of the UMCs with respect to higher education and scientific research, training and the associated care.

The most recent report can be found under the following link:
http://www.nfu.nl/img/pdf/CWTS-analyse_2015.pdf

In 2016, the Department of Health (DH) in the U.K. announced a new, open competition to designate and fund NIHR Biomedical Research Centers (BRCs). BRC (and Biomedical Research Unit (BRU)) designation and funding was awarded to single NHS-university partnerships for the first time in 2007/2008 and a second time in 2012. For both these previous rounds of the competition, an accompanying bibliometric analysis of biomedical and health research in England was produced.

The most recent report presents the findings of a bibliometric analysis of biomedical and health research in England for the period 2004–2013:

https://www.rand.org/pubs/research_reports/RR1363.html

The purpose of the analysis is to support the third NIHR competition for Biomedical Research Centres (BRCs) in England. The report is intended to assist potential applicants in deciding whether to submit a proposal, as well as to inform one of the shortlisting criteria in the deliberations of the International Selection Panel for the BRCs.

Centre for Science and Technology Studies – Leiden University

Postal address

CWTS B.V.
P.O. Box 905
2300 AX Leiden
The Netherlands

Office address

Willem Einthoven Building
Kolffpad 1
2333 BN Leiden
The Netherlands

Contact info

Tel. +31 71 527 3909
Fax +31 71 527 3911
info@cwts.leidenuniv.nl

www.cwts.nl and www.cwtsbv.nl

Anhang 2: DORA: Matthias, Lisa und Holger Breithaupt, *San Francisco Vereinbarung über die Forschungsbewertung* (o.D)



San Francisco Vereinbarung über die Forschungsbewertung

Es besteht dringender Bedarf, die Art und Weise zu verbessern wie die Ergebnisse der wissenschaftlichen Forschung von Geldgebern, akademischen Einrichtungen und anderen Beteiligten bewertet werden. Am 16. Dezember 2012 traf sich eine Gruppe von Redakteuren und Herausgebern wissenschaftlicher Zeitschriften während der Jahrestagung der American Society for Cell Biology (ASCB) in San Francisco, Kalifornien, um Lösungen und Alternativen zu diskutieren. Die Teilnehmer erstellten eine Reihe von Empfehlungen: die San Francisco Declaration on Research Assessment. Wir laden alle Interessierten aus allen wissenschaftlichen Disziplinen ein, diese Erklärung namentlich zu unterstützen.

Die Ergebnisse wissenschaftlicher Forschung sind vielfältig und umfassen Fachpublikationen über neues Wissen, Daten, Reagenzien und Software; geistiges Eigentum; und gut ausgebildete junge Wissenschaftler. Geldbeger, wissenschaftliche Institutionen, und Wissenschaftler selber wollen und müssen ihre Qualität und Auswirkungen beurteilen können. Es ist daher unerlässlich dass diese Ergebnisse präzise und mit Bedacht gemessen und bewertet werden.

Der „Journal Impact Factor“ wird häufig als wesentlicher Parameter verwendet, um die wissenschaftliche Produktivität von Individuen und Institutionen zu vergleichen. Der von Thomson Reuters* berechnete Journal Impact Factor wurde ursprünglich als Instrument für Bibliothekare entwickelt, um Entscheidungen über den Ankauf wissenschaftlicher Fachzeitschriften zu treffen, nicht als Maß für die Qualität der Forschung in einem Artikel. Aus diesem und weiteren gut dokumentierten Gründen ist der Journal Impact Factor kein geeignetes Instrument für die Bewertung von Forschungsergebnissen. Zu den Einschränkungen gehören: A) die Verteilung der Häufigkeit mit der einzelne Artikel in einem Journalen zitiert werden ist stark verzerrt [1-3]; B) der Journal Impact Factor ist feldspezifisch: Er basiert auf unterschiedlichen Artikeltypen, einschließlich primären Forschungsartikeln und Reviews [1, 4]; C) der Journal Impact Factor kann durch redaktionelle Richtlinien manipuliert werden [5]; und D) die Daten, die zur Berechnung der Journal Impact Factors verwendet werden, sind weder transparent noch öffentlich zugänglich [4, 6, 7]. Im Folgenden geben wir eine Reihe von Empfehlungen um die Qualität der Bewertung von Forschungsergebnissen zu verbessern. Zukünftig werden auch andere Leistungen als primäre Forschungsartikel bei der Bewertung der Forschungseffektivität an Bedeutung gewinnen, aber der Peer-Review-Artikel wird auch weiterhin eine zentrale Rolle für die Bewertung haben. Unsere Empfehlungen konzentrieren sich daher in erster Linie auf Praktiken in Bezug auf Forschungsartikel, die in Fachzeitschriften veröffentlicht werden. Sie können und sollten jedoch erweitert werden, um auch andere Ergebnisse wie Datensätze als wichtige Forschungsleistungen anzuerkennen. Diese Empfehlungen richten sich an Geldgeber, akademische



Einrichtungen, Fachzeitschriften, Organisationen, die Metriken erstellen, und Wissenschaftler.

Diese Empfehlungen betreffen eine Reihe von Aufgaben:

- die Notwendigkeit journalbasierte Metriken wie Journal Impact Factors in Bezug auf Finanzierung, Berufung und Beförderung zu vermeiden;
- die Notwendigkeit, die Qualität der Forschung selber zu bewerten, und nicht auf Grundlage der Zeitschrift, in der die Forschung veröffentlicht wird; und
- die Möglichkeiten der Online-Veröffentlichung zu nutzen (z. B. keine unnötige Begrenzung der Anzahl der Wörter, Abbildungen und Zitate in einem Artikel, sowie die Analyse neuer Indikatoren zu Bedeutung und Wirkung).

Wir sind uns bewusst, dass viele Geldgeber, Forschungsinstitute, Verlage und Wissenschaftler bereits verbesserte Praktiken der Forschungsbewertung empfehlen. Diese Ansätze erhöhen die Motivation differenziertere und sinnvollere Methoden für die Bewertung von Forschungsergebnissen zu benutzen, die von allen entscheidenden Beteiligten übernommen und weiter verbessert werden können.

Die Unterzeichner der San Francisco Declaration on Research Assessment unterstützen die folgenden Praktiken in der Forschungsbewertung:

Allgemeine Empfehlung

1. Verwenden Sie keine journalbasierten Metriken, wie z. B. Journal Impact Factor, als Ersatz für die Bewertung der Qualität einzelner Forschungsartikel, um die Beiträge einzelner Wissenschaftler zu bewerten, oder um Entscheidungen über Einstellung, Beförderung oder Finanzierung zu treffen.

Für Geldgeber

2. Stellen Sie klar welche Kriterien Sie bei der Bewertung der wissenschaftlichen Arbeit von Antragstellern benutzen, und betonen Sie, besonders gegenüber Antragstellern am Beginn ihrer wissenschaftlichen Karriere, dass der wissenschaftliche Inhalt eines Artikels wichtiger ist als die Publikationsmetrik oder das Journal in dem er veröffentlicht wurde.
3. Berücksichtigen Sie für die Forschungsbewertung neben Publikationen auch alle anderen Forschungsleistungen (einschließlich Datensätze und Software) sowie eine Reihe anderer Faktoren, so wie qualitativer Indikatoren für die Auswirkungen auf der Forschung, z. B. Einfluss auf die Politik.



Für Forschungsinstitutionen

4. Stellen Sie klar welche Kriterien Sie für Entscheidungen über Einstellungen und Beförderungen benutzen, und betonen Sie, besonders gegenüber Wissenschaftlern am Beginn ihrer Karriere, dass der wissenschaftliche Inhalt eines Artikels wichtiger ist als die Publikationsmetrik oder das Journal, in dem er veröffentlicht wurde.
5. Berücksichtigen Sie für die Forschungsbewertung neben Publikationen auch alle anderen Forschungsleistungen (einschließlich Datensätze und Software) sowie eine Reihe anderer Faktoren, so wie qualitativer Indikatoren für die Auswirkungen auf der Forschung, z. B. Einfluss auf die Politik.

Für Verlage

6. Verringern Sie die Bedeutung des Impact Factors als Marketinginstrument, idealerweise indem Sie den Impact-Faktor nicht mehr zur Werbung benutzen oder durch die Verwendung alternativer journalbasierter Metriken (z. B. 5-Jahres-Impact-Faktor, EigenFaktor 8], SCImago [9], h-Index, Entscheidungs- und Publikationszeiten usw.), die einen besseren Überblick über die Leistungen des Journals bieten.
7. Stellen Sie alternative Metriken auf der Artikelebene zur Verfügung, um die Bewertung des wissenschaftlichen Inhalt eines Artikels, anstelle der Veröffentlichungs metriken des Journals in dem er veröffentlicht wurde, zu bestärken.
8. Ermutigen Sie gewissenhafte Autorenpraktiken und veröffentlichen Sie weitere Informationen über die spezifischen Beiträge jedes Autors.
9. Unabhängig davon, ob ein Journal frei zugänglich oder abonnementbasiert ist, entfernen Sie alle Beschränkungen für Referenzlisten in Forschungsartikeln und stellen Sie sie as Creative Commons Public Domain zur freien Verfügung [10].
10. Entfernen oder reduzieren Sie Beschränkungen der Anzahl der Referenzen in Forschungsartikeln und, wenn möglich, setzen Sie durch dass die Primärliteratur zitiert wird anstelle von Reviews, um der oder den Arbeitsgruppe(n) die zuerst veröffentlicht haben, Anerkennung zu verschaffen.

Für Organisationen, die Metriken erstellen

11. Seien Sie offen und transparent, indem Sie die Daten und Methoden zur Berechnung aller Metriken bereitstellen.
12. Veröffentlichen Sie die Daten unter einer Lizenz, die eine uneingeschränkte Wiederverwendung ermöglicht, und bieten Sie nach Möglichkeit rechnergestützten Zugriff.



13. Stellen Sie klar dass eine unangemessene Manipulation von Metriken nicht toleriert wird; stellen Sie klar was unangemessene Manipulationen sind und welche Maßnahmen dagegen ergriffen werden.
14. Berücksichtigen Sie unterschiedliche Arten von Artikeln (z. B. Reviews oder Primärliteratur) und die verschiedenen Themenbereichen, wenn Metriken verwendet, vereint oder verglichen werden.

Für Wissenschaftler

15. Wenn Sie an Gremien teilnehmen, die Entscheidungen in Bezug auf Finanzierung, Einstellung oder Beförderung treffen, sollten Sie Ihre Bewertungen auf wissenschaftlichen Inhalten statt auf Publikationskennzahlen basieren.
16. Wann immer möglich, zitieren Sie die Primärliteratur in der die Resultate zuerst veröffentlicht wurden, anstelle von Reviews, um die originale Forschungsleistung zu würdigen.
17. Verwenden Sie verschiedene Artikelmetriken und -indikatoren in persönlichen oder bekräftigenden Aussagen die die Bedeutung einzelner Artikel und anderer Forschungsbeiträge bewerten [11].
18. Bezweifeln Sie Bewertungspraktiken, die sich unangemessenerweise auf Journal Impact Factors stützen, und empfehlen Sie „Best Practices“, die sich auf den Wert und die Bedeutung spezifischer Forschungsleistungen konzentrieren.

Verweise

1. [Adler, R., Ewing, J., and Taylor, P. \(2008\) Citation statistics. A report from the International Mathematical Union.](#)
2. [Seglen, P.O. \(1997\) Why the impact factor of journals should not be used for evaluating research. BMJ 314, 498–502.](#)
3. [Editorial \(2005\). Not so deep impact. Nature 435, 1003–1004.](#)
4. [Vanclay, J.K. \(2012\) Impact Factor: Outdated artefact or stepping-stone to journal certification. Scientometric 92, 211–238.](#)
5. [The PLoS Medicine Editors \(2006\). The impact factor game. PLoS Med 3\(6\): e291 doi:10.1371/journal.pmed.0030291.](#)
6. [Rossner, M., Van Epps, H., Hill, E. \(2007\). Show me the data. J. Cell Biol. 179, 1091–1092.](#)
7. [Rossner M., Van Epps H., and Hill E. \(2008\). Irreproducible results: A response to Thomson Scientific. J. Cell Biol. 180, 254–255.](#)
8. <http://www.eigenfactor.org/>
9. <http://www.scimagojr.com/>
10. <http://opencitations.wordpress.com/2013/01/03/open-letter-to-publishers>
11. <http://altmetrics.org/tools/>

* Der Journal Impact Factor wird jetzt von Clarivate Analytics veröffentlicht.

Anhang 3: “Responsible Metrics”, in Wilsdon, James u. a., *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management* (2015), S. 134-135

10. Responsible metrics

“We need new metrics...To move forward, we need to challenge the current norms; define merit much more broadly; measure qualities we value in people but which are hard to measure...This way we can bequeath new models of success and leadership to the next generation to help fix the problems we have inherited from the past.”

Jenny Martin⁴¹⁰

Over the past five years, there has been an increasingly sophisticated debate in the UK, across Europe and internationally about the governance of research and innovation – and the need for better evidence and intelligence to underpin policies in this area. The concept of ‘responsible research and innovation’ (RRI) has gained currency as a framework for research governance within the European Commission, UK Research Councils,⁴¹¹ and research funders in the Netherlands, Germany, the US and Japan. At its simplest, RRI can be defined as “taking care of the future through collective stewardship of science and innovation in the present”. It brings to research policy a focus on questions of anticipation, reflexivity, deliberation, inclusiveness and responsiveness.⁴¹² But we also need robust evidence to underpin this shift and to validate expert opinion.

Drawing on discussions over RRI, we propose the notion of **responsible metrics** as a way of framing appropriate uses of quantitative indicators in the governance, management and assessment of research. The notion of responsible metrics distils the essence of other important contributions to these debates, including the Leiden Manifesto and DORA. Responsible metrics can be understood in terms of a number of dimensions:

- **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;

⁴¹⁰ Martin, J. (2015). Merit and demerit. Blogpost, 6 June 2015. <https://cubistcrystal.wordpress.com/2015/01/12/imagine-theres-new-metrics-its-easy-if-you-try/>. Retrieved 20 June 2015.

⁴¹¹ Such as EPSRC; see <https://www.epsrc.ac.uk/research/framework/>. Retrieved 1 June 2015.

⁴¹² Stilgoe, J., Owen, R. and Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42, 1568-1580; Valdivia, W. and Guston, D. (2015). *Responsible innovation: a primer for policymakers*, Centre for Technology Innovation, Brookings Institution, May 2015. www.brookings.edu/research/papers/2015/05/05-responsible-innovation-valdivia-guston. Retrieved 1 June 2015.

- **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.

As stated in the Leiden Manifesto,⁴¹³ “research metrics can provide crucial information that would be difficult to gather or understand by means of individual expertise. But this quantitative information must not be allowed to morph from an instrument into the goal.”

Everyone can agree that one of the primary aims of research policy at all levels should be to recognise, promote and nurture ‘excellence’. The promise of metrics is that they can make the identification of research excellence more straightforward and more objective – and cheaper and easier to administer. When one designs a metric for research excellence, one also implicitly constructs a definition of excellence, and locks that definition in place. But excellence in research is multidimensional and, crucially, it changes over time. Within this multidimensionality, there are elements that relate to the advancement of knowledge, and elements that relate to benefits beyond the academy (‘impact’). The relationship between these is complex. Research that delivers impact doesn’t necessarily always advance fundamental knowledge, but evidence from REF2014 does suggest that, by-and-large, they are inextricably intertwined.

Because of this, we need to think very clearly about what we are assessing, and select methods and metrics appropriate to our needs. As new research fields develop, definitions of what constitutes excellent research evolve. Interdisciplinary and multidisciplinary research creates new kinds of excellence that don’t necessarily correspond to the definitions appropriate for the component disciplines from which new fields emerge.

Scientometric data are fairly well developed, and can be used to indicate knowledge progression and use and re-use among academic audiences (particularly in the sciences/biosciences but increasingly across different disciplines). However the danger of thoughtless over-reliance on metrics to evaluate research quality is that they can lock in place backward-looking definitions of what constitutes research excellence. A focus on responsible metrics is designed to avoid these pitfalls and ensure that quantitative indicators are used in appropriate, ethical ways.

⁴¹³ www.leidenmanifesto.org

Anhang 4: Fragebogen des ÖWR an österreichische Rektorate

1. Welche Organisationseinheit beschäftigt sich mit dem Thema Bibliometrie bzw. metrik-basierte Evaluationsmethoden in Bezug auf Forschungsleistungen?
2. Gibt es wissenschaftliche Projekte bzw. Arbeitsgruppen zum Thema Bibliometrie an Ihrer Universität?
3. Welche forschungsrelevanten Daten werden erhoben und welche Indikatoren kommen zur Anwendung? Welche metrischen Indikatoren spielen neben der Bibliometrie in der forschungsbezogenen Leistungsbewertung eine Rolle? Kommen hier z.B. Altmetrics¹ zur Anwendung?
4. Welche Kennzahlen bzw. Indikatoren(sets) verwenden Sie, um Fachbereiche, welche nicht traditionell bibliometrisch erfass- und bewertbar sind, international vergleichbar zu messen? Welche Möglichkeiten zur Entwicklung von (internationalen) Indikatoren gibt es?
5. Ist in Zukunft eine Weiterentwicklung metrik-basierter Kennzahlen geplant?
6. Besteht ein Austausch (ggf. auch intern/informell) mit anderen Institutionen/Universitäten zum Thema Bibliometrie bzw. ist eine Zusammenarbeit geplant?
7. Welche Rolle spielen metrische Leistungsindikatoren an Ihrer Universität im Zusammenhang mit
 - Berufungen/Berufungsmanagement;
 - der Vergabe von Qualifizierungsstellen;
 - dem internen Qualitätsmanagement (z.B. Zielvereinbarungen mit Fakultäten, Mittelverteilung innerhalb der Fachbereiche);
 - anderen Bereichen?
8. Gibt es Ansätze für ein Benchmarking mit ähnlich ausgerichteten nationalen und internationalen Fachbereichen?
9. Werden in Ihrer Institution auch interdisziplinäre Zusammenarbeit bzw. deren Potentiale durch bibliometrische Daten erfasst?
10. Was halten Sie von einer einheitlichen Datenerhebung auf nationaler Ebene mittels einer gemeinsamen (nationalen) Forschungsdokumentation, z.B. nach norwegischem Vorbild CRISTIN², zur Vereinfachung der Datenerhebung?
11. Was halten Sie von einer intensiveren österreichweiten Abstimmung bzw. Koordination hinsichtlich der Datenerhebung und Aufbereitung?

¹ Siehe: <https://www.fwf.ac.at/de/ueber-den-fwf/foerderungsstatistiken/altmetrics/> (27. Februar 2019).

² Siehe: <https://www.cristin.no/english/> (27. Februar 2019).

Anhang 5: Internationaler Fragebogen des ÖWR

Individual level

- What evaluation criterions are to be applied when assessing scholars and researches?
- How are they prioritised?
- What is the significance of bibliometric impact indicators for assessing scholars and researchers (e.g. publication figures, conferences, high- and low-ranked journals, h-index)?
- Are there significant differences with respect to academic disciplines?

Departmental level

- What criteria are to be applied when assessing departments?
- How is research output benchmarked on the departmental level?
- What is the significance of acquisition of external funding and publication figures?

University level

- Is there any kind of evaluation of the university?
- Are performance agreements concluded?
- If this is the case, what evaluation criterions are to be applied when assessing the institution?
- What is the significance of bibliometric impact indicators for assessing the institution?

