

Scientometrie

Gerhard Reichmann* und Christian Schlögl

Möglichkeiten zur Steuerung der Ergebnisse einer Forschungsevaluation

Ein Vergleich der Forschungsleistung zweier informationswissenschaftlicher Institute

<https://doi.org/10.1515/iwp-2021-2148>

Zusammenfassung: Ein Leistungsvergleich zwischen den (ehemaligen) Instituten für Informationswissenschaft der Universitäten Düsseldorf und Graz auf Basis der Forschungsleistung für einen Zeitraum von zehn Jahren zeigt, wie sehr die Ergebnisse einer Forschungsevaluation gesteuert werden können. Durch die Wahl „geeigneter“ Indikatoren gelingt es – je nach Wunsch – entweder das eine oder das andere Institut an die erste Stelle zu bringen. Hält man sich dagegen an das wissenschaftliche Gebot der Unparteilichkeit, führt dies im hier gezeigten Anwendungsbeispiel zu gemischten Ergebnissen.

Deskriptoren: Bewertung, Bibliometrie, Forschungsevaluation, Universitätsinstitut, Vergleich, Informationswissenschaft, Neutralität

Possibilities to influence the results of a research evaluation

A comparison of the research performance of two information science institutes

Abstract: A comparison between the (former) Institutes for Information Science at the Universities of Düsseldorf and Graz, based on research performance for a period of ten years shows how much the evaluators can influence the results. When choosing “suitable” indicators, it is possible – depending on the desired result – to rank the favored institute first. If, on the other hand, the evaluators adhere to the scientific principle of impartiality, this leads to mixed results for the example shown here.

*Kontaktperson: ao. Univ.-Prof. DDr. Gerhard Reichmann,

Karl-Franzens-Universität Graz, Institut für Operations und Information Systems, Universitätsstraße 15/F3, 8010 Graz, Österreich,
E-Mail: gerhard.reichmann@uni-graz.at

ao. Univ.-Prof. Dr. Christian Schlögl, Karl-Franzens-Universität Graz, Institut für Operations und Information Systems, Universitätsstraße 15/F3, 8010 Graz, Österreich,
E-Mail: christian.schloegl@uni-graz.at

Descriptors: Assessment, Bibliometrics, Research evaluation, University institute, Comparison, Information Science, Neutrality

Possibilités de contrôle des résultats d'une évaluation de la recherche

Une comparaison des performances de recherche de deux instituts de sciences de l'information

Résumé: Une comparaison des performances des (anciens) instituts des sciences de l'information des universités de Düsseldorf et de Graz sur la base des résultats de la recherche pendant une période de dix ans montre à quel point les résultats d'une évaluation de la recherche peuvent être contrôlés. En choisissant des indicateurs „appropriés“, il est possible – selon ses souhaits – de mettre l'un ou l'autre institut en première place. Si, en revanche, on adhère à l'impératif scientifique d'impartialité, cela conduit à des résultats mitigés dans l'exemple d'application montré ici.

Descripteurs: Évaluation, Bibliométrie, Évaluation de la recherche, Institut universitaire, Comparaison, Science de l'information, Neutralité

1 Einleitung

Leistungsmessungen und Leistungsnachweise spielen im Universitätssektor eine große Rolle (vgl. z. B. Vernon et al. 2018; Morgan 2004). Auch im deutschsprachigen Raum wurden universitäre Leistungen bzw. Aktivitäten in den vergangenen Jahren häufig durch komplexe Evaluationsverfahren gemessen, analysiert und interpretiert (z. B. Reichmann/Sommersguter-Reichmann 2010; Reichmann 2007), insbesondere im Bereich der Forschung (vgl. Albers 2015; Fülbier/Weller 2011). Die Ergebnisse dienen nicht nur als Nachweis für eine erfolgreiche Forschungstätigkeit einzelner, sondern zunehmend auch als Basis für Leistungsvergleiche zwischen mehreren, oftmals konkurrierenden

renden Personen¹ bzw. Leistungseinheiten, wie Instituten, Fachbereichen oder ganzen Universitäten (vgl. Aguillo et al. 2010). Ein gutes Abschneiden bei derartigen Rankings führt oftmals zu deutlichen Wettbewerbsvorteilen, wie eine bevorzugte Behandlung bei der Vergabe von Budgetmitteln.

Vor diesem Hintergrund ist es naheliegend, dass evaluierte Einheiten Anstrengungen unternehmen, um bei vergleichenden Evaluationen möglichst gut abzuschneiden. Ohne Einschränkung positiv zu sehen sind Bemühungen, den Forschungsoutput durch vermehrte Publikationen bzw. stärkere Fokussierung auf möglichst hochwertige Publikationen zu steigern. Bedenklich erscheinen dagegen Versuche, die Ergebnisse durch Auswahl „maßgeschneiderter“ Verfahren zur Leistungsmessung bzw. „geeigneter“ Leistungsindikatoren zum eigenen Vorteil zu beeinflussen. Dies kann im Falle einer Selbstevaluation (hier wird die Evaluation von Angehörigen der evaluierten Einheit selbst durchgeführt) direkt oder im Falle einer Fremdevaluation indirekt durch Auswahl bzw. Beauftragung wohl gesonnenerer Evaluierender erfolgen.

Das Ausmaß solcher Beeinflussungen können Außenstehende oft schwer beurteilen, da sich diese in der Regel plausibel argumentieren und so gut tarnen lassen. Dies soll der vorliegende Beitrag durch einen Vergleich der Forschungsleistung zweier informationswissenschaftlicher Institute zeigen. Dieser wird anhand von Kennzahlen (Leistungsindikatoren) dreimal auf Basis derselben Datengrundlage, aber mit unterschiedlichen Zielsetzungen und Verfahren bzw. Kennzahlen durchgeführt.

Die ersten beiden Vergleiche sind parteiliche Evaluationen, die zu einem möglichst guten Abschneiden einmal des einen und einmal des anderen Instituts führen sollen, jeweils mit plausiblen Argumenten für die gewählte Vorgangsweise. Erst der dritte Vergleich erfüllt eine wesentliche Anforderung an wissenschaftliche Evaluationen aus evaluationstheoretischer Sicht (vgl. Wottawa/Thierau 2003; Stockmann/Meyer 2014), nämlich das Kriterium der Unparteilichkeit. Sie zielt nicht darauf ab, das Ergebnis zugunsten der einen oder anderen Seite zu beeinflussen. Dabei wird auch auf die Schwächen der ersten beiden Evaluationen eingegangen.

In allen drei Fällen handelt es sich um summative, d. h. im Nachhinein durchgeführte Evaluationen, die sich auf denselben zehnjährigen Zeitraum beziehen. Objekt der Evaluation sind stets die beiden Institute als Einheit, wo-

bei die Datenerhebung auf den im Betrachtungszeitraum an diesen Instituten beschäftigten Forschenden basiert. Denn zentraler Anknüpfungspunkt von Forschungsevaluationen auf allen Ebenen ist die Forschungsleistung der einzelnen Wissenschaftlerinnen und Wissenschaftler, vorwiegend in Form von Publikationen.

Nachfolgend wird diskutiert, wie bei einer Forschungsevaluation auf Institutsebene vorgegangen werden kann. Dazu werden verschiedene Dimensionen vorgestellt, die in unterschiedlicher Form berücksichtigt werden können. Daraus ergeben sich zahlreiche Kennzahlen für die Leistungsvergleiche. Im dritten Abschnitt wird die durchgeführte empirische Studie beschrieben, indem die beiden Untersuchungsobjekte sowie die generierte Datenbasis erläutert werden. Es folgt die Präsentation der drei vorgenommenen Institutsvergleiche, wobei jeweils auf die gewählte Vorgangsweise und die sich daraus ergebenden Leistungskennzahlen sowie die dabei erzielten Ergebnisse eingegangen wird. Den Abschluss des Beitrages bildet ein knappes Resümee.

2 Dimensionen einer Forschungsevaluation

In Tabelle 1 sind unterschiedliche Dimensionen samt möglicher Ausprägungen aufgelistet, die bei einer Forschungsevaluation auf Institutsebene berücksichtigt werden können:

1.) Aus welcher (welchen) Datenquelle(n) wird der Forschungsoutput gewonnen? Untersuchungen zufolge hat die Wahl der Datenquelle großen Einfluss auf die Ergebnisse einer Forschungsevaluation (vgl. Meyer et al. 2012). Soll die Forschungsleistung möglichst umfassend erhoben werden, ist ein Rückgriff auf Publikationslisten die optimale Lösung (vgl. Dorsch et al. 2018). Die drei in Tabelle 1 genannten Datenbanken enthalten dagegen nur einen Bruchteil der Publikationen (vgl. Dorsch 2017), v. a. dann, wenn nicht überwiegend in englischsprachigen Zeitschriften publiziert wird.

2.) Weiters ist zu klären, welche Personen(gruppen) als Repräsentanten der Institutsleistung angesehen und deren Publikationen für den Vergleich verwendet werden sollen. Einschränkungen könnten entweder aufgrund inhaltlicher (z. B. ein Großteil der Publikationen stammt von der Professorenschaft) oder erhebungstechnischer (z. B. von jungen Forschenden existieren kaum Publikationslisten) Überlegungen vorgenommen werden.

3.) Hinsichtlich der Publikationen stellt sich die Frage, welche Arten für die Evaluation relevant sein sollen.

¹ Es wurde soweit möglich eine gendergerechte Ausdrucksweise gewählt und nur in Einzelfällen für die bessere Lesbarkeit des Textes das generische Maskulinum verwendet. Es sollen sich jedoch alle Geschlechter angesprochen fühlen.

Tabelle 1: Dimensionen einer Forschungsevaluation auf Institutsebene. (Erläuterung: > bedeutet höhere Wertigkeit)

Dimension	Ausprägungen
1.) Datenquelle	a) Publikationslisten, b) Web of Science, c) Scopus, d) Google Scholar
2.) Repräsentanten der Institutsleistung	a) Alle (wissenschaftliche) Mitarbeitende, b) Nur Post-Docs, c) Nur Professoren
3.) Relevante Publikationsarten	a) Alle, b) Nur ausgewählte (einzeln oder in Kombination): Bücher, Zeitschriftenbeiträge, Beiträge in Sammelwerken, Sonstige
4.) Wertigkeit der relevanten Publikationsarten	a) Keine Berücksichtigung, b) Differenzierungen: z. B. Bücher > Zeitschriftenbeiträge > Beiträge in Sammelwerken > Sonstige
5.) Koautorenschaft	a) Keine Berücksichtigung (full counting), b) Differenzierungen (fractional counting): z. B. $1/n$, $2/(n+1)$, tatsächliche Anteile
6.) Publikationssprache	a) Keine Berücksichtigung, b) Differenzierungen: z. B. englischsprachige und/oder deutschsprachige Publikationen
7.) Umfang der einzelnen Publikationen	a) Keine Berücksichtigung, b) Differenzierungen: z. B. Seitenanzahl und/oder Mindestumfang (in Seiten)
8.) Bei Zeitschriftenbeiträgen: Wertigkeit der Zeitschrift	a) Keine Berücksichtigung, b) Differenzierungen: z. B. Impact-Faktor der Zeitschrift
9.) Größe der evaluierten Einheit (Institutsgröße)	a) Keine Berücksichtigung, b) Differenzierungen: z. B. Anzahl der (wissenschaftlichen) Mitarbeitenden, Anzahl der Professoren

Die in Tabelle 1 genannte Kategorie „Sonstige“ umfasst u. a. Editorials, Rezensionen, Lexikoneinträge oder auch (umfangreiche) Leserbriefe; graue Literatur und Herausgeberschaften bleiben dagegen immer unberücksichtigt. Die häufigste Einschränkung bei den Publikationsarten ist jene auf Zeitschriftenbeiträge. Hier kommt es nicht selten zu einer Berücksichtigung des Impact-Faktors der Zeitschrift, in welcher publiziert wurde (vgl. 8.) in Tab. 1).

4.) Werden dagegen unterschiedliche Publikationsarten berücksichtigt, kann deren Wertigkeit zueinander differenziert werden. Dies ist aber nicht einfach. Die in Tabelle 1 beispielhaft genannte Variante (danach haben Bücher die höchste und Sonstige Beiträge die geringste Wertigkeit) orientiert sich in erster Linie am durchschnittlichen Umfang (deshalb stehen Bücher an erster Stelle) sowie Renommee (Beiträge in Zeitschriften haben in den meisten Disziplinen ein höheres Renommee als solche in Sammelwerken) der einzelnen Publikationsarten.

Zu jeder berücksichtigten Publikation lassen sich zusätzliche Informationen erfassen, die bei Forschungsevaluationen Beachtung finden können. Dazu gehören die in Tabelle 1 genannten Dimensionen: „Koautorenschaft“ (5.), „Publikationssprache“ (6.) und „Umfang der einzelnen Publikationen“ (7.). Für die Zählung von Koautorenschaften, sog. „fractional counting“, gibt es mehrere Möglichkeiten (vgl. Stock 2000, 258). Am verbreitetsten ist die gleiche Berücksichtigung aller Autoren nach der Formel $1/n$ (n = Anzahl der Autoren). Die alternative Formel $2/(n+1)$ (vgl. Tab. 1) führt zu einer über eins liegenden Punktesumme pro Publikation, die mit steigender Anzahl der Autoren zu-

nimmt. Hier wird davon ausgegangen, dass eine Koautorenschaft infolge der notwendigen Koordinierung einen Mehraufwand gegenüber einer Alleinautorenschaft darstellt. Optimal wäre bei Koautorenschaften eine Berücksichtigung der prozentualen Anteile der einzelnen Autoren; diese Information liegt allerdings kaum vor.

9.) Eine Dimension, die in der Praxis wesentlich seltener Beachtung findet als erwartet, ist die „Größe“ der evaluierten Einheit. Bei einem Vergleich auf Institutsebene ist zu diskutieren, ob es fair und sinnvoll ist, kleinere und größere Institute undifferenziert miteinander zu vergleichen. Möchte man die „Größe“ berücksichtigen, so bietet sich eine Division der Gesamtleistung durch die Anzahl der wissenschaftlichen Institutsmitarbeitenden an.

Wählt man nun für jede der neun in Tabelle 1 genannten Dimensionen eine der dort angeführten Ausprägungen und kombiniert man diese miteinander, so ergibt sich daraus eine mögliche Leistungskennzahl bei einer Forschungsevaluation auf Institutsebene. Beispielsweise könnte man als 1.) Datenquelle a) Publikationslisten, als 2.) Repräsentanten der Institutsleistung c) nur Professoren, als 3.) relevante Publikationsarten nur b) Bücher und Zeitschriftenbeiträge, als 6.) Publikationssprache nur b) Englisch und für die übrigen fünf Dimensionen jeweils die Ausprägung a) Keine Berücksichtigung heranziehen. Die so generierte Leistungskennzahl könnte man beispielsweise als „Anzahl der englischsprachigen Bücher und Zeitschriftenbeiträge der Professoren auf Basis der Publikationslisten“ bezeichnen.

Bei Kombination aller in Tabelle 1 genannten Ausprägungen ließen sich mehrere tausend Leistungskennzahlen

generieren. Im empirischen Teil des Beitrages werden einzelne dieser Kombinationsmöglichkeiten für die durchgeführten Vergleiche herausgegriffen.

3 Institutsvergleich auf Basis einer Forschungsevaluation

3.1 Vergleichsobjekte und Datenbasis

Für den nun folgenden beispielhaften Institutsvergleich der Forschungsleistung für einen Zeitraum von zehn Jahren (2009–2018) wurden die beiden ehemaligen Institute für Informationswissenschaft an den Universitäten Düsseldorf und Graz² ausgewählt, da a) diese im gleichen Forschungsbereich angesiedelt sind, b) die beiden Autoren dieses Beitrages, die Angehörige des zweitgenannten Instituts sind, auch das Partnerinstitut sehr gut kennen, und c) die Forschungsleistung beider Institute in Form von Publikationslisten der Institutsangehörigen dokumentiert ist.

Um verschiedene Varianten von Institutsvergleichen (vgl. Tab. 1) durchführen zu können, wurde zunächst der Publikationsoutput der beiden Institute für den Untersuchungszeitraum auf Basis von Publikationslisten erhoben und in eine eigens dafür erstellte Access-Datenbank eingetragen. Dazu wurden sämtliche Publikationen aller während des Untersuchungszeitraumes an diesen Instituten beschäftigten Forschenden gezählt. Dazu gehören alle Professoren und alle sonstigen wissenschaftlichen Mitarbeitenden.

Für jede Publikation wurden folgende Merkmale dokumentiert:

- 1.) die Art der Publikation, wobei zwischen Büchern, Beiträgen in (wissenschaftlichen) Zeitschriften, Beiträgen in Sammelwerken und Sonstigen Beiträgen (für Details: vgl. Abschnitt 2) unterschieden wurde,
- 2.) die Anzahl der Autoren,
- 3.) die Sprache: hier gab es nur deutsch- oder englischsprachige Publikationen,
- 4.) der Umfang der Publikation in Seiten sowie
- 5.) im Falle von Zeitschriftenbeiträgen der Name der Zeitschrift.

² Das Düsseldorfer „Institut“, bei dem es sich genau genommen „nur“ um eine Abteilung des Instituts für Sprache und Information handelt, befindet sich gerade in Auflösung; das Grazer Institut wurde mit Oktober 2020 – zusammen mit zwei anderen Instituten – in das neu gegründete Institut für Operations und Information Systems übergeführt.

Darüber hinaus wurde ermittelt, welche der auf Basis der Publikationslisten erhobenen Veröffentlichungen auch im Web of Science vorhanden waren. Im positiven Fall wurden auch noch der Impact-Faktor der Zeitschrift im Jahr der Publikation sowie die erhaltenen Zitate festgehalten.

3.2 Institutsvergleich zugunsten von Düsseldorf

Ausgangspunkt ist, dass sich das Düsseldorfer Institut möglichst gut darstellen will und daher eine entsprechend parteiliche Evaluation in Auftrag gibt.

Bei dieser Vorgangsweise wird argumentiert, dass Publikationslisten als Datenquelle verwendet werden, da diese den Forschungsoutput am besten repräsentieren. Ebenso wird auch die Heranziehung aller wissenschaftlichen Mitarbeitenden als Repräsentanten der Institutsleistung begründet. Was die relevanten Publikationsarten betrifft, so sollen hier zunächst einmal alle zusammen betrachtet werden, um einen Gesamtüberblick über die Forschungsleistung zu erhalten. Dies ergibt – sofern alle übrigen Dimensionen laut Tabelle 1 keine Berücksichtigung finden – die Kennzahl „**Anzahl der Publikationen**“. Diese und alle folgenden Kennzahlen werden für den Institutsvergleich laut Tabelle 2 eingesetzt.

Der Ausschluss der Sonstigen Publikationen führt zur Kennzahl „**Anzahl der Publikationen ohne Sonstige**“. Begründet wird dieser Ausschluss mit der Heterogenität der Sonstigen Publikationen, die, wie beispielsweise Rezensionen, nicht allzu viel mit Forschung zu tun hätten. In Folge wird anhand der Kennzahlen „**Anzahl der Bücher**“, „**Anzahl der Zeitschriftenbeiträge**“ und „**Anzahl der Beiträge in Sammelwerken**“ dargelegt, dass Düsseldorf nicht nur insgesamt, sondern auch bei den einzelnen Publikationsarten führend ist.

Um zu zeigen, dass Düsseldorf nicht nur mehr, sondern auch „wertiger“ publiziert hat als Graz, wird die Wertigkeit der relevanten Publikationsarten berücksichtigt. Und zwar auf zwei unterschiedliche Varianten, die demonstrieren, dass Düsseldorf – verfahrensunabhängig – immer vorne liegt. Bei Variante 1 werden Bücher/Zeitschriftenbeiträge/Beiträge in Sammelwerken/Sonstige Beiträge mit dem Faktor 10/5/3/1 multipliziert, bei Variante 2 werden Zeitschriftenbeiträge und Beiträge in Sammelwerken zu einer Kategorie zusammengefasst, sodass es nur mehr drei relevante Publikationsarten gibt, die mit den Faktoren 3 (Bücher), 2 (neue Kategorie) und 1 (Sonstige Beiträge) multipliziert werden. Durch die geringeren Abstände zwischen den vergebenen Punkten sowie die Bildung der neuen Kategorie erfahren bei Variante 2 Beiträge in Sammelwerken gegen-

über Zeitschriftenbeiträgen und Sonstige Beiträge generell eine Aufwertung gegenüber Variante 1. Die entsprechenden Kennzahlen werden als „**Wert der Publikationen-Variante 1**“ bzw. „**Wert der Publikationen-Variante 2**“ bezeichnet.

Die aktuellen Publikationslisten deutschsprachiger Forschender lassen erkennen, dass diese mittlerweile vielfach in englischer Sprache publizieren, um international „konkurrenzfähig“ zu sein und z. B. in internationalen, zumeist englischsprachigen Spitzenzeitschriften vertreten zu sein. Zur Darlegung entsprechender englischsprachiger Publikationstätigkeit wird die Kennzahl „**Anzahl der englischsprachigen Publikationen**“ eingesetzt.

Es gibt zwar keine Standards für die optimale Länge der einzelnen Publikationen, die natürlich sehr stark von Publikationsart und -medium abhängt, doch vor allem Zeitschriftenbeiträge und Beiträge in Sammelwerken sollten einen gewissen Umfang aufweisen, um als vollwertige Fachbeiträge angesehen zu werden. Für die Messung dieses Umfanges wäre die Anzahl der Zeichen oder Wörter optimal, in Folge mangelnder Verfügbarkeit entsprechender Daten wird meist – wie auch hier – auf die Seitenanzahl zurückgegriffen, obwohl diese stark vom Layout beeinflusst wird. Zum Publikationsumfang sollen zwei Kennzahlen berechnet werden, einerseits der „**Umfang der Publikationen ohne Sonstige in Seiten**“, und andererseits der „**Umfang der Zeitschriftenbeiträge in Seiten**“.

Allen bisherigen Kennzahlen lag das „full counting“ zugrunde, bei dem Koautorenschaften unberücksichtigt bleiben, d. h. bei einer Betrachtung auf Institutsebene zählt jede Publikation mit einem Institutsangehörigen als Koautor für das Institut voll. Dies führt bei einem Institutsvergleich zu einer Bevorzugung jenes Instituts, das häufiger Koautorenschaften mit institutsfremden Personen aufweist, da institutsfremder Input dem Institut zugerechnet wird. Um mögliche Kritik in diese Richtung zu entkräften, soll die Anzahl der Publikationen auf Basis eines „fractional counting“ nach der Formel $1/n$ berechnet werden. Die entsprechende Kennzahl wird „**Anzahl der Publikationen-Fractional counting**“ genannt.

Um zu demonstrieren, wie robust die positiven Ergebnisse für Düsseldorf sind, wird, wieder auf Basis des „full counting“, gezeigt, was passieren würde, wenn man z. B. aus ressourcenökonomischen Gründen nicht auf die Publikationslisten als Datenquelle zurückgreifen könnte, sondern das (unvollständige) Web of Science verwenden müsste, in dem praktisch nur englischsprachige Zeitschriftenbeiträge enthalten sind. Die entsprechende Kennzahl lautet „**Anzahl der Publikationen-Web of Science**“. Ein Rückgriff auf das Web of Science bietet den Vorteil, dass auf

Basis dieser Daten zwei weitere Leistungskennzahlen berechnet werden können: der „**Impact der Publikationen**“ – dieser wird ermittelt, indem die einzelnen Impact-Faktoren der Zeitschriften, in denen die Publikationen veröffentlicht wurden, aufsummiert werden – sowie die „**Anzahl der Zitate**“, welche die Publikationen eines Instituts im Web of Science erhalten haben. Dabei handelt es sich um eine international gebräuchliche und dementsprechend anerkannte Kennzahl.

Tabelle 2: Institutsvergleich zugunsten von Düsseldorf.

Leistungskennzahl	Düsseldorf	Graz	Differenz
Anzahl der Publikationen	345	228	51 %
Anzahl der Publikationen ohne Sonstige	286	145	97 %
Anzahl der Bücher	10	2	400 %
Anzahl der Zeitschriftenbeiträge	94	62	52 %
Anzahl der Beiträge in Sammelwerken	182	81	125 %
Wert der Publikationen-Variante 1	1170	652	79 %
Wert der Publikationen-Variante 2	636	371	71 %
Anzahl der englischsprachigen Publikationen	219	101	117 %
Umfang der Publikationen ohne Sonstige in Seiten	8348	1812	361 %
Umfang der Zeitschriftenbeiträge in Seiten	1187	630	88 %
Anzahl der Publikationen-Fractional counting	272	173	57 %
Anzahl der Publikationen-Web of Science	43	23	87 %
Impact der Publikationen	35	17	106 %
Anzahl der Zitate	705	204	246 %

Anhand von Tabelle 2 lässt sich erkennen, dass Düsseldorf in allen Fällen mehr oder weniger (vgl. die Prozentsätze in der Spalte „Differenz“ der Tabelle) deutlich vor Graz liegt. Besonders groß sind die Differenzen im Hinblick auf die Anzahl der Bücher, den Umfang in Seiten sowie die Anzahl der Zitate. Für diesen Vergleich hat es sich demnach für Düsseldorf gelohnt, so viele (und so umfangreiche) Bücher verfasst zu haben. Und die im Web of Science enthaltenen Düsseldorfer Publikationen werden wesentlich häufiger zitiert als jene von Graz. Im Durchschnitt ist die Leistung des Düsseldorfer Instituts um mehr als das Doppelte (138 %) besser als jene des Grazer Instituts. Eine isolierte Publikation dieses Institutsvergleiches würde also Düsseldorf über Graz geradezu triumphieren lassen.

3.3 Institutsvergleich zugunsten von Graz

Dass es auch umgekehrt gehen kann, wird nun in diesem Abschnitt gezeigt. Eine zweite Evaluation soll die Ergebnisse des obigen Vergleiches entkräften und Graz möglichst gut dastehen lassen. Dabei könnte man sich – entsprechende Parteilichkeit für Graz vorausgesetzt – relativ ressourcenschonend und simpel weitgehend auf die Kennzahlen und Argumente des ersten Institutsvergleiches stützen, diese aber in zweifacher Hinsicht „adaptieren“.

Die erste Adaption bezieht sich auf die Repräsentanten der Institutsleistung. Beim ersten Vergleich waren dies alle wissenschaftlichen Mitarbeitenden der Institute. Die zweite Evaluation könnte nun argumentieren, dass die Schlüsselpersonen für die Forschungsleistung eines Instituts die dort tätigen Professorinnen und Professoren seien, von denen vermutlich auch ein Großteil der Publikationen stammen wird. Aus diesem Grund sollten auch nur diese die Institutsleistung mit ihren Publikationen repräsentieren. Dieses Argument kommt dem Grazer Institut entgegen, da dort während des gesamten Betrachtungszeitraumes stets vier Professoren beschäftigt waren. In Düsseldorf gab es dagegen immer nur einen Professor, der allerdings – im Gegensatz zu Graz – permanent auf eine Vielzahl von wissenschaftlichen Mitarbeitenden zurückgreifen konnte. Und da im vorliegenden Fall in der Regel ein „full counting“ auf Institutebene durchgeführt wird, kommen Koautorenschaften mit wissenschaftlichen Mitarbeitenden dem Institut voll zugute. Die erste Adaption führt dazu, dass alle für den ersten Vergleich herangezogenen Kennzahlen unter alleiniger Berücksichtigung der Professoren als Repräsentanten der Institutsleistung neu berechnet und bezeichnet werden. Die neuen Bezeichnungen entsprechen den alten mit dem Zusatz „Professoren“ (vgl. Tab. 3).

Die zweite Adaption besteht darin, von den im ersten Schritt adaptierten Kennzahlen nur jene heranzuziehen, bei denen Graz besser abschneidet. Die übrigen werden nicht genannt, was einem Außenstehenden bei isolierter Betrachtung des zweiten Vergleiches nicht weiter auffallen sollte (vgl. Tab. 2 mit Tab. 3: demnach scheinen in Tab. 3 sechs Kennzahlen laut Tab. 2 nicht auf).

Tabelle 3: Institutsvergleich zugunsten von Graz.

Leistungskennzahl	Düsseldorf	Graz	Differenz
Anzahl der Publikationen-Professoren	107	192	79 %
Anzahl der Publikationen ohne Sonstige-Professoren	103	117	14 %

Tabelle 3: fortgesetzt.

Leistungskennzahl	Düsseldorf	Graz	Differenz
Anzahl der Zeitschriftenbeiträge-Professoren	39	50	28 %
Anzahl der Beiträge in Sammelwerken-Professoren	59	65	10 %
Wert der Publikationen-Variante 1-Professoren	425	536	26 %
Wert der Publikationen-Variante 2-Professoren	214	307	43 %
Anzahl der Publikationen-Fractional counting-Professoren	48	139	190 %
Anzahl der Publikationen-Web of Science-Professoren	17	20	18 %

Gemäß Tabelle 3 ist Graz bei allen aufgelisteten Kennzahlen besser als Düsseldorf. Besonders groß ist die Differenz hinsichtlich der Anzahl der Publikationen bei Anwendung des „fractional counting“ nach der Formel $1/n$. Dies war zu erwarten, da es in Düsseldorf viele Koautorenschaften zwischen dem Lehrstuhlinhaber (dem einzigen Professor) und den wissenschaftlichen Mitarbeitenden gab, die bei dieser Zählweise eben nur anteilmäßig (für den Professor) berücksichtigt werden. Im Durchschnitt liegt die Differenz zwischen Graz und Düsseldorf bei immerhin 51 Prozent zugunsten von Graz. Eine isolierte Betrachtung dieses Vergleiches durch Außenstehende würde diesen also den Eindruck vermitteln, dass die Forschungsleistung des Grazer Instituts wesentlich besser ist als jene des Düsseldorfer Instituts.

3.4 Unparteilicher Institutsvergleich

Im Falle eines unparteilichen Institutsvergleiches kann man die Kennzahlen und Argumente der bisherigen parteilichen Vergleiche als Basis nehmen und den Anforderungen an eine objektive Evaluation anpassen.

Dies bedeutet folgende Adaptionen hinsichtlich des ersten Vergleiches: Beim Institutsvergleich zugunsten von Düsseldorf wird eine wesentliche Komponente außer Acht gelassen, und zwar die Institutgröße (vgl. Tab. 1: Dimension „Größe der evaluierten Einheit“). Im Betrachtungszeitraum verfügte das Düsseldorfer Institut im Durchschnitt über 11,3 wissenschaftliche Mitarbeitende (wMA: diese Zahl inkludiert die Professorenstelle und ist in Vollzeitäquivalenten (VZÄ) gemessen), während der entsprechende Wert für das Grazer Institut bei 5,1 lag. Demnach hatte das Düsseldorfer Institut doppelt so viele personelle Ressourcen, was für einen Leistungsvergleich von großer

Bedeutung ist. Daher bietet sich eine Neuberechnung und -bezeichnung der in Tabelle 2 aufgelisteten Kennzahlen an, indem die Werte durch die Anzahl der wMA dividiert werden, und die so generierten Kennzahlen wie in Tabelle 2, aber mit dem Zusatz „/wMA“ bezeichnet werden. Diese Anpassung ist in der oberen Hälfte von Tabelle 4 abgebildet. Dagegen sind alle im Zuge der Kennzahlenbildung vorgebrachten Argumente der ersten Evaluation nicht nur schlüssig, sondern halten auch einer Prüfung im Hinblick auf Unparteilichkeit stand, weshalb diese für den hier vorgenommenen unparteilichen Institutsvergleich vollinhaltlich gültig sind.

Beim Vergleich zugunsten von Graz wurde eine eher unübliche Beschränkung auf die Forschungsleistung der Professoren vorgenommen. Diese würde auch von den Autoren dieses Beitrages im Falle einer unparteilichen Forschungsevaluation eher nicht eingesetzt, lässt sich aber durchaus objektiv begründen (vgl. Abschnitt 3.3). Auch hier wurde die Größe der evaluierten Einheit außer Acht gelassen. In diesem Fall zum Vorteil von Graz, wo während des Betrachtungszeitraumes vier Professoren beschäftigt waren – im Gegensatz zu einem Professor in Düsseldorf. Insofern ist der Größenvorteil von Graz hier noch viel eklatanter als jener von Düsseldorf beim ersten Vergleich. Zur Bereinigung dieses Effekts bietet sich eine Neuberechnung und -bezeichnung bereits vorhandener Kennzahlen, diesmal jener laut Tabelle 3, an, indem durch die Anzahl der Professoren dividiert und die neue Kennzahl durch den Zusatz „/Prof“ charakterisiert wird. Das Ergebnis dieser Vorgangsweise spiegelt die untere Hälfte von Tabelle 4 wider. Zudem sind hier nicht nur die adaptierten Kennzahlen auf Basis von Tabelle 3 angeführt, sondern auch die sechs dort verschwiegenen (vgl. Tab. 2 und 3), bei denen Graz schlechter abgeschnitten hätte als Düsseldorf.

Eine Gesamtbetrachtung von Tabelle 4 legt offen, dass bei den 28 durchgeführten Vergleichen 18-mal Düsseldorf und 10-mal Graz besser abschneidet. Allerdings sind die prozentuellen Differenzen zugunsten von Düsseldorf meist deutlich höher als jene zugunsten von Graz, so dass es in Summe eine durchschnittliche Differenz zugunsten von Düsseldorf in Höhe von 179 Prozent gibt.

Besonders günstig fallen die Ergebnisse für Düsseldorf aus, wenn man ausschließlich die Professoren als Repräsentanten der Institutsleistung heranzieht (vgl. Tab. 4 unten). Hier gewinnt Düsseldorf sämtliche Vergleiche deutlich und weist im Durchschnitt um 349 Prozent höhere Kennzahlenwerte auf als Graz. Diese massiven Differenzen haben vor allem drei Ursachen: Erstens die überdurchschnittlich hohe Forschungsproduktivität des Düsseldorfer Professors. Zweitens die hohe Zahl von wMA in Düsseldorf, die für Koautorenschaften oder überhaupt nur

für Recherche- und sonstige Vorarbeiten im Zuge von Publikationsprojekten eingesetzt werden konnten. Während Düsseldorf im Betrachtungszeitraum über durchschnittlich mehr als zehn derartige wMA verfügte, gab es in Graz nur einen entsprechenden Mitarbeitenden. Und drittens profitiert Düsseldorf, vor allem aufgrund der hohen Anzahl an Koautorenschaften mit den eben genannten wMA, vom hier primär verwendeten „full counting“.

Tabelle 4: Unparteilicher Institutsvergleich. (Erläuterung: wMA steht für wissenschaftliche Mitarbeitende, Prof steht für Professoren).

Leistungskennzahl	Düsseldorf	Graz	Differenz
Anzahl der Publikationen/wMA	30,5	44,5	-31 %
Anzahl der Publikationen ohne Sonstige/wMA	25,3	28,3	-11 %
Anzahl der Bücher/wMA	0,9	0,4	126 %
Anzahl der Zeitschriftenbeiträge/wMA	8,3	12,1	-31 %
Anzahl der Beiträge in Sammelwerken/wMA	16,1	15,8	2 %
Wert der Publikationen-Variante 1/wMA	103,4	127,2	-19 %
Wert der Publikationen-Variante 2/wMA	56,2	72,4	-22 %
Anzahl der englischsprachigen Publikationen/wMA	19,3	19,7	-2 %
Umfang der Publikationen ohne Sonstige in Seiten/wMA	737,5	353,6	109 %
Umfang der Zeitschriftenbeiträge in Seiten/wMA	104,9	122,9	-15 %
Anzahl der Publikationen-Fractional counting/wMA	24	33,8	-29 %
Anzahl der Publikationen-Web of Science/wMA	3,8	4,5	-15 %
Impact der Publikationen/wMA	3,1	3,3	-5 %
Anzahl der Zitate/wMA	62,3	39,8	56 %
Anzahl der Publikationen-Professoren/Prof	107	48	123 %
Anzahl der Publikationen ohne Sonstige-Professoren/Prof	103	29,3	252 %
Anzahl der Bücher-Professoren/Prof	5	0,5	900 %
Anzahl der Zeitschriftenbeiträge-Professoren/Prof	39	12,5	212 %
Anzahl der Beiträge in Sammelwerken-Professoren/Prof	59	16,3	263 %
Wert der Publikationen-Variante 1-Professoren/Prof	425	134	217 %
Wert der Publikationen-Variante 2-Professoren/Prof	214	76,8	179 %

Tabelle 4: fortgesetzt.

Leistungskennzahl	Düsseldorf	Graz	Differenz
Anzahl der englischsprachigen Publikationen-Professoren/Prof	82	18,3	349 %
Umfang der Publikationen ohne Sonstige in Seiten-Professoren/Prof	4536	394,3	1051 %
Umfang der Zeitschriftenbeiträge in Seiten-Professoren/Prof	557	129	332 %
Anzahl der Publikationen-Fractional counting-Professoren/Prof	47,8	34,7	38 %
Anzahl der Publikationen-Web of Science-Professoren/Prof	17	5	240 %
Impact der Publikationen-Professoren/Prof	17,1	4,1	315 %
Anzahl der Zitate-Professoren/Prof	264	50,8	420 %

Legt man dem Institutsvergleich dagegen die Forschungsleistung aller wMA zugrunde (vgl. Tab. 4 oben), dann befindet sich Graz in zehn von 14 Fällen an erster Stelle, u. a. bei den zentralen Kennzahlen „Anzahl der Publikationen/wMA“ (sowohl nach dem „full“ als auch nach dem „fractional counting“), „Anzahl der Zeitschriftenbeiträge/wMA“ und „Anzahl der englischsprachigen Publikationen/wMA“. Allerdings sind die Differenzen zugunsten von Graz meist gering. Eine Stärke von Düsseldorf sind dagegen stets Bücher, Beiträge in Sammelwerken (v. a. in Proceedings) und erhaltene Zitate.

4 Resümee

Es konnte anhand der drei durchgeführten Institutsvergleiche gezeigt werden, wie sehr Evaluierende die Ergebnisse durch Auswahl entsprechender Verfahren und Leistungskennzahlen beeinflussen können. Minimalvoraussetzung, um ein Institut an die Spitze eines Vergleiches bringen zu können, ist, dass dieses Institut zumindest irgendwelche Forschungsleistungen in Form von Publikationen erbracht und in einzelnen Nischen mehr geleistet hat als das bzw. die Vergleichsinstitut(e).

Die Frage, welches der beiden verglichenen Institute nun tatsächlich besser war, lässt sich objektiv kaum beantworten. Beide waren aktiv und produktiv und hatten ihre Stärken und Schwächen. Stärken von Düsseldorf waren ein überdurchschnittlich aktiver Lehrstuhlinhaber, der selbst viel publizierte, die Publikation von zahlreichen Büchern und Proceedings-Beiträgen (dies lässt wiederum auf intensive Konferenzaktivitäten und damit auf eine gute Verankerung in der scientific community schließen)

sowie die Behandlung von Themen, die auf Resonanz stießen, wie die zahlreichen erhaltenen Zitate belegen. Stärken von Graz waren die homogenere Verteilung der erbrachten Forschungsleistung auf die vorhandenen Mitarbeitenden, die Konzentration auf Zeitschriftenbeiträge sowie das breitere Themenspektrum (vgl. Dorsch et al. 2017), das wiederum eine Folge der deutlich höheren Anzahl an Professoren in Graz sein dürfte.

Gegenstand weiterer Forschungsaktivitäten könnte einerseits eine detailliertere Diskussion der vorgestellten Verfahren und Kennzahlen und – damit verbunden – eine umfassendere Auswertung des vorhandenen Datenmaterials sein. Zu untersuchen wäre, welche der hier vorgestellten neun Dimensionen eines Leistungsvergleichs besonders relevant für die Ergebnisse sind. Zudem wäre es interessant, die hier untersuchten Institute wieder einmal (vgl. Friedländer 2014) mit anderen informationswissenschaftlichen Instituten oder auch mit anderen Instituten der jeweiligen Universität bzw. Fakultät zu vergleichen.

Literatur

- Albers, S. (2015): What drives publication productivity in German business faculties?, in: *sbr* 67 (Jan.), 6–33.
- Aguillo, I./Bar-Ilan, J./Levene, M./Ortega, J. (2010): Comparing university rankings, in: *Scientometrics* 85 (1), 243–256.
- Dorsch, I. (2017): Relative visibility of authors' publications in different information services, in: *Scientometrics*, 112 (2), 917–925.
- Dorsch, I./Askeridis, J./Stock, W.G. (2018): Truebounded, overbounded, or under-bounded? Scientists' personal publication lists versus lists generated through bibliographic information services, in: *Publications*, 6 (1), 1–9.
- Dorsch, I./Rauch, W./Schlögl, C./Stock, W.G. (2017): Forschungsthemen der Düsseldorfer und Grazer Informationswissenschaft (2010 bis 2016), in: *Information. Wissenschaft & Praxis* 68 (5–6), 1–9.
- Friedländer, M. (2014): Informationswissenschaft an deutschsprachigen Universitäten – eine komparative informetrische Analyse, in: *Information. Wissenschaft & Praxis* 65 (2), 109–119.
- Fülbier, R./Weller, M. (2011): A glance at German financial accounting research between 1950 and 2005: A publication and citation analysis, in: *sbr* 63 (Jan.), 2–33.
- Meyer, M./Waldkirch, R./Zaggl, M. (2012): Relative performance measurement of researchers: the impact of data source selection, in: *sbr* 64 (Oct.), 308–330.
- Morgan, K. (2004): The research assessment exercise in English universities, 2001, in: *Higher Education* 48 (4), 461–482.
- Reichmann, G. (2007): Evaluierung von Studienplänen. Ermittlung von Möglichkeiten zur Verbesserung von Studienbedingungen, in: *Zeitschrift für Hochschulrecht, Hochschulmanagement und Hochschulpolitik (zfhr)* 6 (2), 37–45.
- Reichmann, G./Sommersguter-Reichmann, M. (2010): Efficiency measures and productivity indexes in the context of university library benchmarking, in: *Applied Economics* 42 (1–3), 311–324.

Stock, W.G. (2000): Was ist eine Publikation? Zum Problem der Einheitenbildung in der Wissenschaftsforschung, in: Fuchs-Kittowski, K./Laitko, H./Parthey, H./Umstätter, W. (Hrsg.): Wissenschaft und Digitale Bibliothek. Wissenschaftsforschung Jahrbuch 1998, Berlin: Gesellschaft für Wissenschaftsforschung, 239–282.

Stockmann, R./Meyer, W. (2014): Evaluation – Eine Einführung, 2. Aufl., Stuttgart: UTB.

Vernon, M./Balas, E./Momani, S. (2018): Are university rankings useful to improve research? A systematic review, in: PLoS ONE 13 (3), 1–15.

Wottawa, H./Thierau, H. (2003): Lehrbuch Evaluation, 3. Aufl., Bern: Hans Huber.



ao. Univ.-Prof. Dr. Gerhard Reichmann

Karl-Franzens-Universität Graz
Institut für Operations und Information
Systems

Universitätsstraße 15/F3

8010 Graz

gerhard.reichmann@uni-graz.at

Ao. Prof. Dr. Gerhard Reichmann ist seit 1997 an der Sozial- und Wirtschaftswissenschaftlichen Fakultät der Karl-Franzens-Universität Graz beschäftigt. Seine aktuellen Forschungsschwerpunkte sind Evaluierung (Evaluierung von universitärer Forschung, Lehre und Verwaltung, Evaluierung im Gesundheitsbereich), Informationskompetenz, Informationsrecht (Datenschutzrecht, Urheberrecht und Plagiatsproblematik).



ao. Univ.-Prof. Dr. Christian Schlögl

Karl-Franzens-Universität Graz
Institut für Operations und Information
Systems

Universitätsstraße 15/F3

8010 Graz

christian.schloegl@uni-graz.at

Ao. Prof. Dr. Christian Schlögl ist seit 1990 an der Sozial- und Wirtschaftswissenschaftlichen Fakultät der Karl-Franzens-Universität Graz beschäftigt, seit 2005 ist er stellvertretender wissenschaftlicher Leiter des an den Universitäten Wien, Innsbruck und Graz eingerichteten interuniversitären Universitätslehrgangs „Master of Science (MSc) in Library and Information Studies“. Seine aktuellen Forschungsgebiete sind Bibliometrie und Szientometrie, Informationskompetenz sowie Informations- und Wissensmanagement