

Dimensions - Evaluierung der Inhalte und bibliometrischen Datenbasis

Impressum

Dieser Report wurde erstellt von:

Barbara Scheidt & Andreas Meier (Forschungszentrum Jülich, Zentralbibliothek)

Zeitraum der Erstellung: Januar 2020 – September 2020

Kontakt:

Forschungszentrum Jülich

Zentralbibliothek - AG Bibliometrie

E-Mail: zb-bibliometrie@fz-juelich.de

Telefon: 02461 / 61-6198

Weitere Informationen zur Bibliometrie in der Zentralbibliothek unter:

www.bibliometrie.de

Auftragsnummer Forschungszentrum Jülich

DZ.004198

Inhaltsverzeichnis

1. Einführung	6
2. Methodik	7
3. Ergebnisse Datenbanken	9
Gesamtbestand der Datenbanken	9
Identifizier (DOI)	9
Dokumenttypen	10
Quellen	13
Publisher	14
Sprachen	15
Kategorien	17
Durchschnittliche Anzahl Adressen je Publikation 2009-2018	19
Durchschnittliche Anzahl Autoren je Publikation 2009-2018	19
Publikationen mit Corresponding Author	20
Autorenadressen: Länder	20
Funding Information	21
Zusammenfassung statistischer Überblick	23
4. Ergebnisse Case Studies	24
Institutionenanalyse: Forschungszentrum Jülich	24
Publikationsjahre	25
Dokumenttypen	25
Quellen und Publisher	26
Sprachen	27
Kategorien	27
Anzahl Autoren	28
Corresponding Authors	28
Anzahl Adressen	29
Länder	29
Funding Information	30
Institutionenanalyse: Fraunhofer-Institut für Keramische Technologien und Systeme (IKTS)	31
Publikationsjahre	32
Dokumenttypen	32
Quellen und Publisher	32
Sprachen	33
Kategorien	33
Anzahl Autoren	33
Corresponding Authors	33
Anzahl Adressen	33
Länder	34
Funding Information	34
Autorenanalyse: Christian Drost	34
Publikationsjahre	36
Dokumenttypen	37
Quellen	37

Publisher.....	38
Sprachen.....	39
Kategorien	39
Anzahl Autoren.....	40
Corresponding Authors	40
Anzahl Adressen	41
Länder.....	41
Funding Information.....	42
5. Zusammenfassung der Ergebnisse und Fazit	42
6. Literaturverzeichnis.....	44

Tabellenverzeichnis

Tabelle 1: DIM19-Article und entsprechende Dokumenttypen in WOS. Stichprobe von 475 Publikationen.....	12
Tabelle 2: DIM19-Article und entsprechende Dokumenttypen in Scopus. Stichprobe von 490 Publikationen.....	13
Tabelle 3: DIM19, die häufigsten Quellen.....	14
Tabelle 4: DIM19, Top 10-Herausgeber	15
Tabelle 5: die 10 häufigsten Sprachen in DIM20 (1996 - 2018) und entsprechende Publikationszahlen aller drei Datenbanken im Vergleich.....	16
Tabelle 6: Durchschnittliche Anzahl Autoren je Publikation und je Datenquelle 2009-2018.....	19
Tabelle 7: Anteil der Publikationen mit Angaben zu einem Corresponding Author	20
Tabelle 8: Anzahl der Publikationen je Datenbank mit Angabe mindestens einer Förderorganisation 1996-2018.....	22
Tabelle 9: TopTen-Quellen aller drei Datenbankergebnisse für das FZJ.....	26
Tabelle 10: FZJ-Publikationen: Kategorien je Datenbank für einen Artikel im Vergleich.....	27
Tabelle 11: Autoren und Corresponding Authors in den Publikationen des FZJ jeweils in Dimensions, WOS und Scopus.....	28
Tabelle 12: Publikationen des IKTS nach Publikationsjahren je Datenbank	32
Tabelle 13: Top 8 Quellen aller Datenbanken der IKTS-Publikationen	32
Tabelle 14: Top 8 Länder der Publikationen des Fraunhofer IKTS in den Datenbanken.....	34
Tabelle 15: Anzahl der Publikationen von C. Drosten in den unterschiedlichen Datenquellen	35
Tabelle 16: Publikationsjahre der Publikationen von C. Drosten in allen drei Datenbanken	36
Tabelle 17: Drosten-Publikationen:Gegenüberstellung der Dokumenttypen aller drei Datenbanken	37
Tabelle 18: Publikationen von C.Drosten nach First-Level-Kategorien in DIM20	39
Tabelle 19: Publikationen von C. Drosten nach Second-Level-Kategorien in DIM20.....	39
Tabelle 20: Publikationen von C. Drosten nach Subject Category in WOS	40
Tabelle 21: Publikationen von C. Drosten nach Fields in Scopus	40

Abbildungsverzeichnis

Abbildung 1: Anzahl Publikationen insgesamt bei DIM19, Scopus und Web of Science mit Publikationsjahren 1996-2018	9
Abbildung 2: DIM19 Anteil der Dokumenttypen, 1996-2018	11
Abbildung 3: Web of Science Anteil Dokumenttypen, 1996-2018.....	11
Abbildung 4: Scopus Anteil Dokumenttypen, 1996-2018	12
Abbildung 5: Anzahl englischsprachiger Publikationen in DIM20, WOS und Scopus 1996 - 2018	16
Abbildung 6: Anzahl Publikationen von weiteren 9 Sprachen in DIM20, WOS und Scopus 1996 - 2018.....	16
Abbildung 7: Anzahl der Publikationen je 'Division' in Dimensions, unterschieden nach den Datenständen von April 2019 und März 2020	18
Abbildung 8: Entwicklung der durchschnittlichen Anzahl der Autoren je Publikation von 2009 bis 2018 in den Datenquellen.....	20
Abbildung 9: Anzahl Publikationen der acht stärksten Länder in DIM20, WOS und Scopus 1996-2018.....	21
Abbildung 10: Anzahl Publikationen des FZJ mit DOI je Datenbank und Schnittmengen.....	24
Abbildung 11: FZJ- Publikationen des Dokumenttyps Article in den Datenbanken.....	25
Abbildung 12: Die 5 Länder mit den höchsten Anteilen in der Analyse zum FZJ in allen 3 Datenbanken im Vergleich	30
Abbildung 13: Anzahl Publikationen des IKTS mit DOI je Datenbank und Überschneidungsmengen.....	31
Abbildung 14: Ergebnisse für C. Drost in den Datenbanken -- Schnittmengen	35
Abbildung 15: Die 5 Länder mit den höchsten Anteilen in der Analyse von C. Drost in allen 3 Datenbanken im Vergleich	41

1. Einführung

Als multidisziplinäre Zitationsdatenbank ist Dimensions von Digital Science eine interessante Quelle für bibliometrische Analysen, die nach Angaben des Unternehmens mit dem Anspruch entwickelt wurde, 'neue Wege zur Entdeckung, Erschließung und Analyse von Forschung herzustellen' (Williams, 2018). In Hinblick auf eine Verwendung in der KB-Infrastruktur prüft die vorliegende Studie diesen Anspruch und untersucht Dimensions im Vergleich mit den aktuellen KB-Quellen Web of Science (WoS) und Scopus.

Die kommerzielle Ausgabe von Dimensions umfasste beim Start im Januar 2018 bereits mehr als 133 Millionen Forschungsdokumente, die nach dem Anspruch des Datenbankbetreibers den gesamten Forschungszyklus abdecken sollen. Zu den Publikationen wissenschaftlicher Journals, die mit 95 Millionen Dokumenten den Hauptteil ausmachen, kamen damals schon mehr als 4 Millionen Grants, 400.000 Clinical Trials und 35 Millionen Patente hinzu (Orduna-Malea & Delgado Lopez-Cozar, 2018; Williams, 2018). Seitdem sind die Zahlen aller Bereiche weiter angewachsen auf insgesamt mehr als 160 Millionen Dokumente: so enthält die Datenbank mittlerweile z.B. über 112 Millionen Publikationen wissenschaftlicher Journals und wurde außerdem um über 540.000 Policy Documents erweitert (Stand 09.09.2020).

Im Hinblick auf eine mögliche Verwendung der Dimensions-Daten innerhalb der KB-Infrastruktur ist erforderlich, ein umfassendes Bild über die Inhalte, die Datenkonsistenz und ihre Qualität zu erhalten. Durch das bei Dimensions verwendete automatisierte Verfahren zur Generierung der Nachweise wissenschaftlicher Publikationen, welches das Vorhandensein eines DOI oder einer PubMed-ID voraussetzt, und dem ausdrücklichen Verzicht auf eine inhaltliche Auswahl der indexierten Quellen (Hook, Porter, & Herzog, 2018), ist zumindest in Teilen ein Unterschied der in Dimensions indexierten Nachweise zu den Zitationsindizes der aktuellen KB-Quellen zu erwarten. Diese Unterschiede werden durch die vorliegende Untersuchung benannt und die Frage nach einer sinnvollen Verwendung der Inhalte im Rahmen bibliometrischer Untersuchungen beantwortet. Die Untersuchung erfolgt ergänzend zu Studien des DZHW und des I²SOS an der Universität Bielefeld. Die Arbeitsgruppe Bibliometrie im Forschungszentrum Jülich bearbeitet im Rahmen dieser Kooperation die folgenden Untersuchungsthemen:

1. Quantitativer Überblick: statistische Erfassung und Darstellung der Inhalte von Dimensions und Überprüfung auf mögliche Lücken (z.B. fehlende Fächer, Verlage, Regionen oder Sprache) und die diesbezüglichen Differenzen zu WOS und Scopus
2. Bibliometrische Basis: Identifikation der für bibliometrische Analysen relevanten Informationen (bibliografische Angaben, zusätzliche Datenfelder) und Evaluierung der Datenkonsistenz und –Abdeckung hinsichtlich Vollständigkeit und möglicher Probleme als Gewähr für belastbare Datenanalysen

2. Methodik

Die für diesen Teil der Studie verwendeten Dimensions-Daten liegen im JSON-Dateiformat in der Open Source-Datenbank MongoDB ¹ in einer lokalen Installation des FZJ mit zwei unterschiedlichen Datenständen vor: April 2019 und März 2020. Zur Unterscheidung werden sie im Bericht als DIM19 und DIM20 bezeichnet.

Von Scopus und Web of Science wurden Daten der Versionen WOS_B_2019 und Scopus_B_2019 aus der SQL-Datenbank des Kompetenzzentrums Bibliometrie genutzt. Der Datenstand dieser Datensätze ist die Kalenderwoche 17 aus dem Jahr 2019 (April). Aus Web of Science beinhaltet die Datenbank die Indexe SCIE, SSCI, A&HCI, CPCI-S, CPCI-SS. Es wurden jeweils Publikationszahlen zu den folgenden Feldern erhoben:

- Gesamtbestand der Datenbanken (bei Dimensions auch in Publikationsarten unterteilt)
- DOI
- Dokumenttypen
- Quellen
- Herausgeber
- Sprachen
- Kategorien
- Autoren- und Adressangaben
- Länder
- Förderinformationen

Zusätzlich für die Publikationen der letzten 10 Jahre (2009-2018):

- Durchschnittliche Anzahl Autoren je Publikation
- Durchschnittliche Anzahl Adressen je Publikation
- Anzahl der Publikationen mit Benennung eines Corresponding Authors

Für die Gesamtübersicht wurden die Analysen jeweils auf die Publikationsjahre 1996 bis 2018 eingeschränkt, da von Scopus für die vorangehenden Jahre keine Daten vorliegen und der Datenstand eine Betrachtung über 2018 hinaus nicht möglich macht.

Die Verwendung aktuellerer Dimensions-Daten (Datenstand März 2020, DIM20) wurde für einzelne Auswertungen als sinnvoll erachtet, weil erst für diesen Zeitpunkt für viele Teilbereiche vollständigere Informationen vorliegen und Vergleiche mit den anderen beiden Datenbanken überhaupt erst möglich machen. Außerdem ist die Analyse der neueren Daten im Hinblick auf eine zukünftige Nutzung notwendig. Bis auf wenige Ausnahmen gingen in die Analysen jedoch nur Publikationen aus der aktuelleren Version ein, die bereits 2019 (DIM19) indexiert waren.

Als Ergebnis werden die formalen und inhaltlichen Unterschiede zwischen Dimensions auf der einen und WOS und Scopus auf der anderen Seite beschrieben: welche Quellen beinhalten die Datenbanken und welche Datenfelder stehen zur Verfügung? Gibt es in der zeitlichen Entwicklung starke Unterschiede innerhalb und im Vergleich der Datenbanken?

Die Eignung der Dimensions-Daten als Grundlage für bibliometrische Analysen sollte anhand beispielhafter Fragestellungen zu verschiedenen Aggregationsebenen geprüft werden, jeweils im Vergleich zu den KB-Bibliometriedatenbanken. Dazu wurden Case-Studies zu zwei deutschen Einrichtungen (Forschungszentrum Jülich, Fraunhofer-Institut für Keramische Technologien und

¹ mongodb.com

Systeme) und zu einer Person (C. Drosten) durchgeführt. Die unterschiedlichen Ergebnisse wurden schrittweise betrachtet und analysiert: von den rein quantitativen Differenzen (Ergebnismengen) bis zu den konkreten Inhalten einzelner Felder und der Datenqualität.

Innerhalb der Case Studies-Vergleiche wurden folgende Felder besonders beachtet:

- Die Research Fields in Dimensions, die nicht, wie die Klassifikationssysteme von WoS und Scopus, den Zeitschriften zugeordnet sind, sondern den Artikeln (Bode, Herzog, Hook, & McGrath, 2018; Bornmann, 2018). Da sich sowohl Methoden als auch Systeme unterscheiden, kann im Rahmen dieser Untersuchung nur eine Abschätzung zu Überschneidungen, Abweichungen und Problemen der Systeme vorgenommen werden, aber keine abschließende Bewertung der Klassifikationsqualität.
- Zeitschriftentitel: Für keine der Datenbanken liegen normierte Daten zu den Zeitschriftentiteln vor. Dimensions gewinnt die Publikationsdaten i.d.R. aus CrossRef² und PubMed³ und verwendet Zeitschrifteninformationen z.B. aus dem Directory of Open Access Journals (DOAJ).⁴ Bisherige Erfahrungen haben jedoch gezeigt, dass diese Informationen z.B. bei Titelveränderungen nicht ausreichen.

Im Fall der Institutionenanalyse wurden die Publikationen über die jeweiligen Institutionen-Identifizier ausgewählt. Um vor allem die Inhalte der Datensätze vergleichen zu können, wurde eine Schnittmenge identischer Publikationen aus allen drei Datenbanken analysiert. Der Abgleich ist mit den DOI erfolgt. Das bedeutet aber auch, dass die Differenzen im Bestand zwar dargestellt, aber nicht näher untersucht werden konnten. Der Aufwand für damit verbundene Untersuchungen hätte den Rahmen der Studie gesprengt.

2 crossref.org

3 pubmed.gov

4 doaj.org

3. Ergebnisse Datenbanken

Gesamtbestand der Datenbanken

Eine vergleichende Betrachtung ist aufgrund des Scopus-Bestandes erst ab dem Publikationsjahr 1996 möglich. Der Datenstand der KB-Daten begrenzt den Zeitraum wiederum auf das Jahr 2018 als jüngstes Publikationsjahr. Der Datenstand der verwendeten KB-Bibliometriedatenbanken (17. KW 2019) ist jedoch für die Dimensions-Daten problematisch, da es seither inhaltlich einige Verbesserungen gab, die für die Vergleiche relevant sind. Das bezieht sich nicht nur auf neu indexierte Publikationen, sondern auch auf bereits im April 2019 enthaltene Items mit ergänzten Informationen. Am Beispiel der Fachgebiete lässt sich die deutliche Veränderung aufzeigen: lediglich bei 55 % der fast 63 Millionen Dokumente der Publikationsjahre 1996 bis 2018 ist im April 2019 ein 'Field of Research' verzeichnet. Die Daten aus dem März 2020 sind diesbezüglich deutlich verbessert: bei 72 % der gut 65,5 Millionen Publikationen dieses Zeitabschnitts existiert ein Divisionseintrag (FORS first level).

Dimensions (DIM19) umfasst im Untersuchungszeitraum die größte Menge an Veröffentlichungen (fast 63 Millionen), vor Scopus (rund 48 Millionen) und Web of Science (ca.43 Millionen).

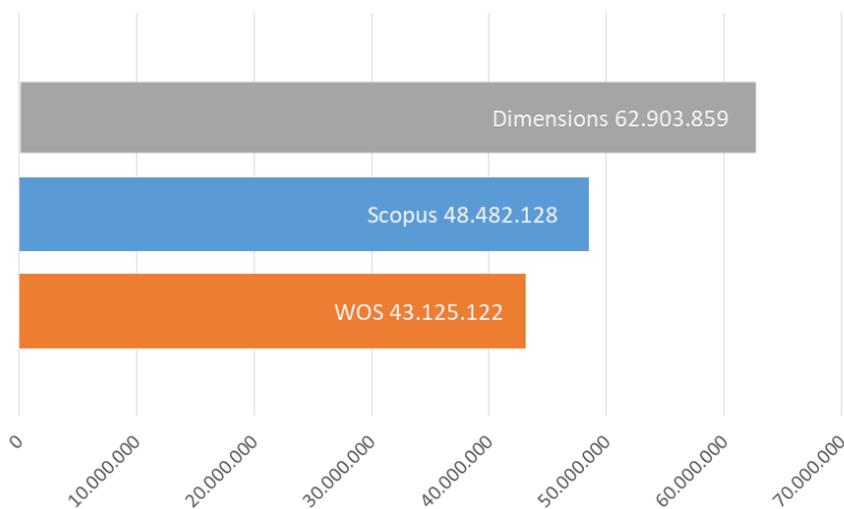


Abbildung 1: Anzahl Publikationen insgesamt bei DIM19, Scopus und Web of Science mit Publikationsjahren 1996-2018

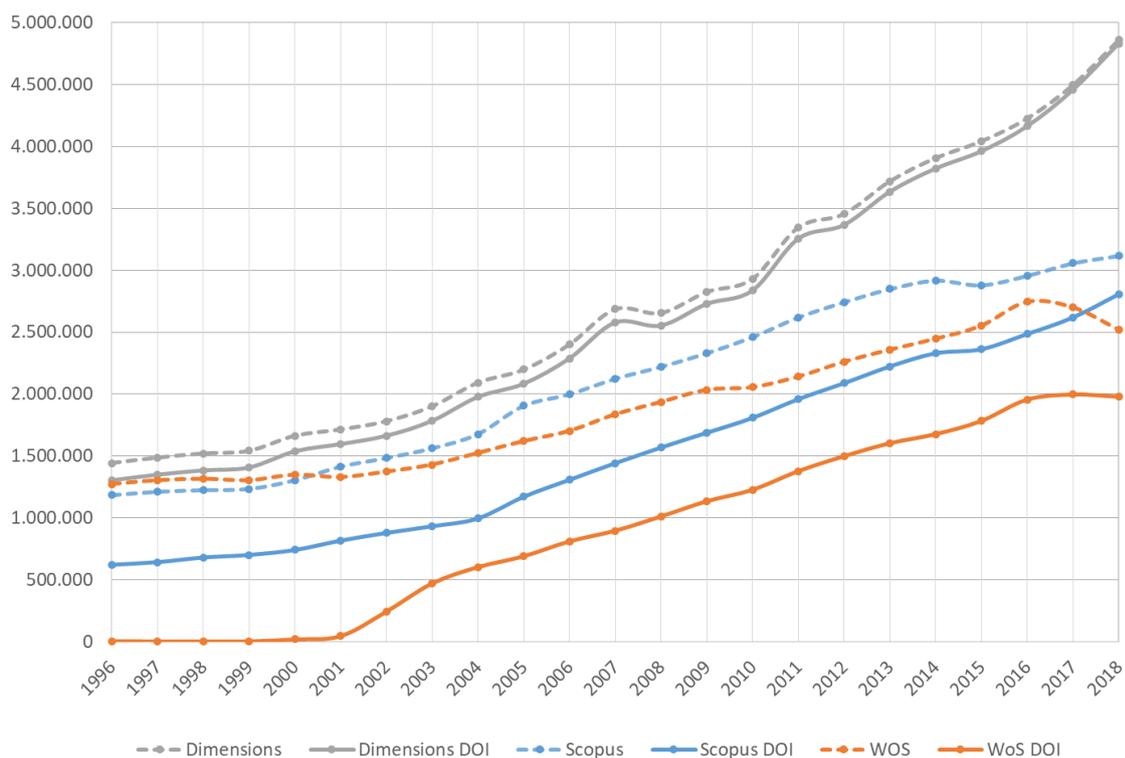
Identifizier (DOI)

Beim Vergleich der Datenbestände aller drei Datenbanken fällt auf, dass Dimensions nicht nur die meisten Publikationen indexiert, sondern auch den höchsten Anteil an Publikationen mit DOI aufweist. Das ist begründet durch die bevorzugte Methode der Inhaltsgenerierung von Digital Science, nämlich Publikationsdaten mit eindeutigen IDs wie DOI oder PubMed Identifier (Bode et al., 2018) zu sammeln. Dass dadurch auch Publikationen aus dem 15. Jahrhundert in den Bestand kommen, hängt mit deren Online-Veröffentlichungen in jüngerer Zeit zusammen, durch die sie einen DOI erhalten haben. Bei einem Beispiel, das im Original 1724 und online 1997 veröffentlicht wurde, handelt es sich um eine Abbildung aus den *Philosophical Transactions of the Royal Society of London*. Der Verlag vermerkt dazu: "This text was harvested from a scanned image of the original document using optical character recognition (OCR) software."⁵ Die eher allumfassende Auswahlmethode für den Dimensions-Bestand

⁵ DOI: 10.1098/rstl.1724.0080

führt im Vergleich zur intellektuellen Auswahl des WOS-Bestandes also zu einer größeren Vielfalt. Dass die genannte Abbildung als Publikationstyp 'Article' klassifiziert ist, wird im anschließenden Kapitel über die Dokumenttypen Thema.

Im WOS dagegen ist eine Etablierung des DOI erst ab dem Jahr 2000 erkennbar, zuvor beinhalten im Verhältnis nur wenige Publikationsdatensätze einen DOI. Während Scopus und WOS relativ ähnliche jährliche Steigerungen der Anzahl von Publikationen mit DOI aufweisen, ist der Anstieg bei Dimensions im Vergleich zur Konkurrenz deutlich größer: 2018 sind in Dimensions fast 5 Millionen DOIs erfasst, während WOS unter 2 Millionen und Scopus knapp unter 3 Millionen bleiben. Lediglich gut 28.000 Einträge (0,6 %) aus dem Jahr 2018 haben in Dimensions keinen DOI. Bei Scopus hat im selben Jahr 10 % (rund 311.000) des Bestandes keinen DOI, bei WOS sogar fast 27 % (723.000). Da die Anzahl der Publikationen, die einen DOI haben, aber generell steigt und sich dies auch im Bestand von Scopus und WOS zeigt, ist die Wahrscheinlichkeit groß, dass es sich bei den Publikationen jüngerer Datums ohne DOI um solche handelt, für die kein DOI existiert. Stichproben zeigen außerdem, dass diese Publikationen im Bestand von Dimensions fehlen. Auch wenn die Auswahlmethode von Dimensions zu einem deutlich größeren Bestand gegenüber WOS und Scopus führt, gibt es durch sie auch Literatur, die ausgeschlossen bleibt.



Dokumenttypen

Die Bezeichnungen der Dokumenttypen unterscheiden sich zwischen den drei Datenbanken erheblich, worauf bereits Visser et al. in ihrer Studie hinweisen (Visser, van Eck, & Waltman, 2020). In Dimensions existieren überhaupt nur sechs Dokumenttypen, während die Publikationen im WOS in 41 und in Scopus in 16 Dokumenttypen eingeteilt sind. Aus diesem Grund wird für die Gesamtübersicht 1996-2018 vorrangig die jeweilige Verteilung der Dokumenttypen dargestellt, ohne datenbankübergreifende Vergleiche.

In DIM19 liegen überwiegend Dokumente des Typs Article vor (~80 %), gefolgt von Chapter (~11 %) und Proceedings (~8 %). Mit jeweils unter 1 % Anteil sind Monograph, Preprint und Book vertreten.

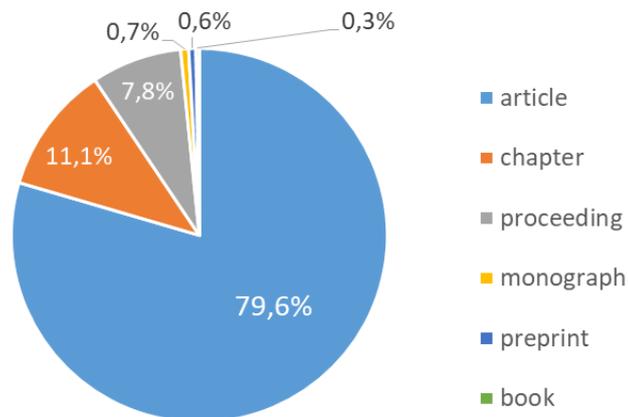


Abbildung 2: DIM19 Anteil der Dokumententypen, 1996-2018

Web of Science verzeichnet für denselben Zeitraum 40 Dokumententypen und der Anteil der Article beträgt nur knapp 60 % am Gesamtoutput des Zeitraums. Dargestellt sind die Dokumententypen, die mindestens 1 % Anteil am Gesamtbestand aufweisen (Abbildung 3).

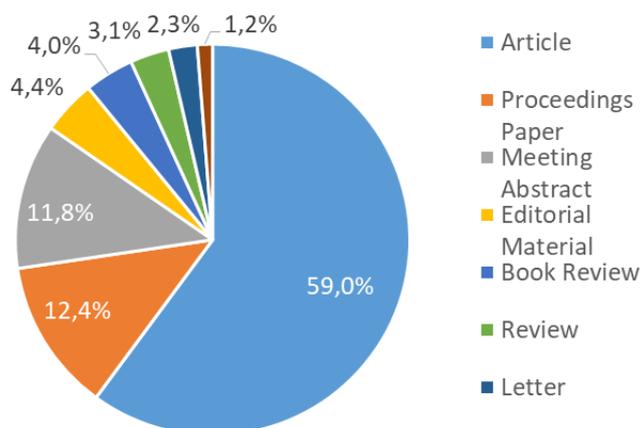


Abbildung 3: Web of Science Anteil Dokumententypen, 1996-2018

Scopus unterteilt in nur 16 Dokumententypen mit rund 66 % Article-Anteil. In Abbildung 4 sind die Dokumententypen mit mindestens 1 % Anteil an der Gesamtzahl der Publikationen des Zeitraums dargestellt.

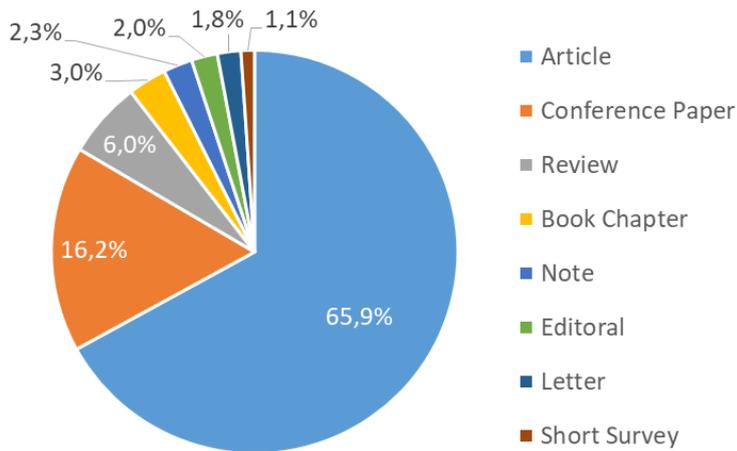


Abbildung 4: Scopus Anteil Dokumenttypen, 1996-2018

Da Article aus WOS und Scopus fester Bestandteil bibliometrischer Analysen sind, wurde aus DIM19 eine zufällige Stichprobe von 2.000 Dokumenten dieses Typs mit dem Publikationsjahr 2015 in den beiden anderen Datenquellen gesucht und die Dokumenttypen verglichen. Lediglich 475 Publikationen konnten über einen DOI-Abgleich auch in WOS identifiziert werden, sie weisen dabei 11 verschiedene Dokumenttypen auf.

Tabelle 1: DIM19-Article und entsprechende Dokumenttypen in WOS. Stichprobe von 475 Publikationen

WOS Dokumenttypen	Anzahl Publikationen
Article	199
Book Review	72
Editorial Material	69
Meeting Abstract	67
Correction	20
News Item	16
Letter	12
Proceedings Paper	9
Biographical-Item	6
Review	4
Retraction	1

Die Dimensions-Article werden in Web of Science demnach zu einem nicht unbedeutenden Teil gar nicht als Forschungsergebnisse klassifiziert. Insbesondere die große Zahl der Book Reviews fällt auf. Dieser Dokumenttyp ist in Analysen wissenschaftlicher Arbeiten in der Regel nicht enthalten.

Die Abfrage derselben 2.000 DOI ergab in Scopus 490 Treffer mit einer anderen Dokumenttypverteilung: immerhin waren auch hier in der Mehrzahl Article dabei (61 %). Meeting Abstracts und Book Reviews verzeichnet Scopus generell nicht (*Scopus Content Coverage Guide, 2020*), diese Vergleichsgruppen sind daher nicht im Bestand.

Tabelle 2: DIM19-Article und entsprechende Dokumenttypen in Scopus. Stichprobe von 490 Publikationen

Scopus Dokumenttyp	Anzahl Publikationen
Article	300
Note	54
Editorial	38
Review	36
Erratum	29
Conference Paper	14
Letter	12
Short Survey	7

Beispiele, wie die im vorherigen Kapitel genannte Abbildung aus den *Philosophical Transactions*⁶, für die Dimensions den Publikationstyp Article vergeben hat, zeigen eine Schwachstelle des Publikationstypensystems von Dimensions im Hinblick auf Differenzierung und Abgrenzung der Inhalte auf. Auch aktuellere Veröffentlichungen sind, abweichend von der Definition dieses Dokumenttyps bei den anderen Datenbanken oder nach den Angaben des Verlags, als Article indexiert, obwohl sich darunter beispielsweise auch Titelblätter befinden, wie *Cover and Front matter* der Zeitschrift *National Institute Economic Review*.⁷

Quellen

Die Ergebnisse der Zeitschriftentitel und anderen Quellenangaben aus Dimensions stellen sich übersichtlicher dar als die der anderen beiden Datenbanken. Grund sind Listen mit vereinheitlichten Titelnamen, auf denen die Einträge basieren (Hook et al., 2018, p. 9; Wang & Waltman, 2016).

Die Auswertungen der Dimensionsdaten hat gezeigt, dass der Anteil von Dokumenten ohne einen Quellentitel sowohl in DIM19 als auch DIM20 bei über 20 % liegt. Unter den Quellen mit der höchsten Anzahl an Publikationen befinden sich eine Plattform wie SSRN Electronic Journal (für Open Access-Preprints) oder ein Abstractservice wie ChemInform. Noch aktuellere Daten (online, vom 23.06.2020) führen den Preprint-Server arXiv als häufigste Quelle. In der Indexierung dieser Inhalte unterscheidet sich Dimensions maßgeblich von Scopus und Web of Science. Vier von den zehn Hauptquellen in DIM19 sind weder in WOS noch in Scopus auf den Quellenlisten: neben den beiden erstgenannten auch Choice, eine Rezensionszeitschrift für englischsprachige wissenschaftliche Bücher und digitale Medien, und Reactions Weekly.

6 DOI: 10.1098/rstl.1724.0080

7 DOI: 10.1017/s0027950100007894

Tabelle 3: DIM19, die häufigsten Quellen

	DIM19 – Quelle	Anzahl Publikationen	Anteil an Gesamt
1	<i>SSRN Electronic Journal</i>	333.092	0,53 %
2	<i>ChemInform</i>	322.179	0,51 %
3	PLoS ONE	211.510	0,34 %
4	<i>Choice</i>	147.533	0,23 %
5	<i>Reactions Weekly</i>	134.127	0,21 %
6	Physical Review B	122.741	0,20 %
7	Applied Mechanics and Materials	108.983	0,17 %
8	Science	102.485	0,16 %
9	Journal of Biological Chemistry	101.093	0,16 %
10	The BMJ	99.038	0,16 %
11	American Journal of Cancer	94.006	0,15 %
12	Nature	89.504	0,14 %
13	Applied Physics Letters	89.359	0,14 %
14	Gastroenterology	88.769	0,14 %
15	Journal of Clinical Oncology	84.795	0,13 %
16	The Journal of Urology	83.330	0,13 %
17	Scientific Reports	83.129	0,13 %
18	Proceedings of the National Academy of Sciences	83.126	0,13 %
19	The Journal of the Acoustical Society of America	79.680	0,13 %
20	Journal of Applied Physics	76.653	0,12 %

Die Titel in kursiver Schrift kennzeichnen Preprints-, Abstracts- oder Rezensionensammlungen

An der Vereinheitlichung von Titeln ist problematisch, dass historische Veränderungen teilweise nicht mehr erkennbar sind und, zumindest in Einzelfällen, Zeitschriften zusammengeführt wurden, die völlig unterschiedlich sind. Zum letztgenannten Fall gehört das Beispiel der Zeitschrift *Robotics and Autonomous Systems*. Die ISSN 0921-8890 gehört zur Zeitschrift bei Elsevier, bis 1987 lautete der Titel *Robotics*. Die zweite ISSN 2218-6581 gehört zu einem OA-Journal namens *Robotics* des Verlags MDPI.

Publisher

Die Anzahl von Dokumenten ohne Angabe zum Herausgeber ist von April 2019 bis März 2020 von 5 auf 9,2 % gestiegen. Aufgrund der besseren Datenlage können also die Daten aus 2019 für diesen Teil der Untersuchung verwendet werden.

Es werden 9.781 unterschiedliche Namen von Herausgebern für die Publikationsjahre 1996-2018 verzeichnet. In Scopus und WOS gibt es deutlich mehr unterschiedliche Herausgebernamen (15.051 / 10.731). Gleichzeitig scheint noch nicht einmal die Hälfte aller Publikationen dieses Zeitraums in Scopus (Bibliometriedatenbank Kompetenzzentrum) einem Herausgeber zugeordnet zu sein. Die Problematik nicht bereinigter Namen auch für die Herausgeber besteht schon lange. Die jeweiligen Ergebnisse können demnach nicht ohne vorherige Bearbeitung der Namen durchgeführt werden.

Tabelle 4: DIM19, Top 10-Herausgeber

DIM19 – Publisher	Anzahl Publikationen
Elsevier	11.283.728
Springer Nature	7.860.898
Wiley	4.874.250
Institute of Electrical and Electronics Engineers (IEEE)	3.552.729
Taylor & Francis	3.026.859
Oxford University Press (OUP)	1.664.839
SAGE Publications	1.440.379
Wolters Kluwer	1.427.144
Cambridge University Press (CUP)	856.262
American Chemical Society (ACS)	837.567

Als Ergebnis kann an dieser Stelle die relativ gute Datenlage bezüglich der Vollständigkeit in Dimensions festgehalten werden. Innerhalb der Case Study-Analysen werden die Ergebnisse einzelner Publikationen genauer betrachtet werden.

Sprachen

Angaben über die Sprache der Dokumente gehören in Dimensions zu den Informationen, die im Datenstand April 2019 nur geringfügig vorhanden waren: 157.082 Datensätze hatten einen Eintrag zur Sprache, das entspricht 0,25 %. Im März 2020 gab es hingegen nur noch 0,73 % der Daten *ohne* Angabe zur Sprache. Rund 86 % der Dokumente sind englischsprachig, weitere 135 Sprachen sind für die verbleibenden 14 % verzeichnet. Es gibt keine Mehrfachzuordnung von Sprachen für eine Dokumenten-ID. Von den Sprachen erreichen 116 gemeinsam lediglich einen Anteil von knapp 1 % aller verzeichneten Publikationen. 15 weitere erreichen jeweils noch nicht mal ein halbes Prozent Anteil an der Gesamtpublikationsmenge des untersuchten Zeitraums. Nur Deutsch (1.817.375 Dokumente/ 2,8 %) und Japanisch (1.544.228/ 2,4 %) erreichen über 2 % Anteil. Französisch, Portugiesisch und Spanisch liegen zwischen 1,9 - 1,0 Prozent Anteil.

In Web of Science sind 95,4 % der Dokumente englischsprachig bei einer Vielfalt von 52 Sprachen.

In Scopus werden 60 verschiedene Sprachen ausgewiesen, 90 % der Publikationen sind den Angaben zufolge in englischer Sprache verfasst worden und in fast 400.000 Fällen sind wohl mehrere Sprachen zu einer Veröffentlichung erfasst. Deutschsprachige Dokumente sind in DIM20 und WOS am zweithäufigsten vertreten. In Scopus werden chinesische Artikel häufiger als deutsche indexiert.

Vergleicht man nur die absoluten Zahlen der drei Datenbanken, so gibt es in WOS sieben Sprachen, für die jeweils mehr als 100.000 Dokumente verzeichnet sind, in Scopus 10 und in Dimensions 15 Sprachen.

Da für die nachfolgenden Tabellen und Grafiken die Daten aus DIM20 mit einem aktuelleren Datenbestand auch im Vergleich zu WOS und Scopus ausgewertet wurden, gilt es, den Zuwachs der absoluten Publikationszahlen gegenüber DIM19 zu beachten. Dieser betrug in den 11 Monaten rund 4,3 %, so dass z.B. die vergleichbare Zahl englischsprachiger Dokumente im April 2019 auf ca. 54 Millionen geschätzt werden kann.

Tabelle 5: die 10 häufigsten Sprachen in DIM20 (1996 - 2018) und entsprechende Publikationszahlen aller drei Datenbanken im Vergleich

Sprache	DIM	WOS	Scopus
English	56.547.830	41.145.952	43.681.348
German	1.817.375	544.105	764.074
Japanese	1.544.228	46.319	360.371
French	1.267.001	485.610	608.387
Portuguese	799.986	90.953	222.677
Spanish	680.715	249.010	497.747
Chinese	301.766	162.371	1.630.641
Russian	263.130	121.815	330.081
Romanian	209.925	5.831	14.565
Italian	208.770	104.303	148.995

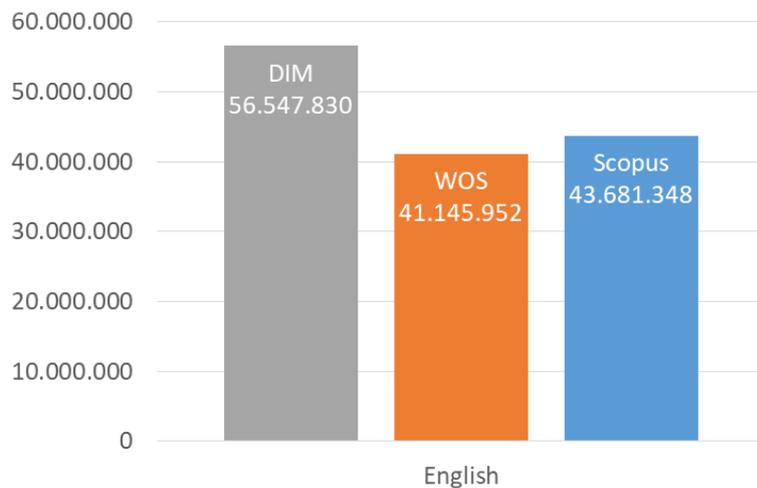


Abbildung 5: Anzahl englischsprachiger Publikationen in DIM20, WOS und Scopus 1996 - 2018

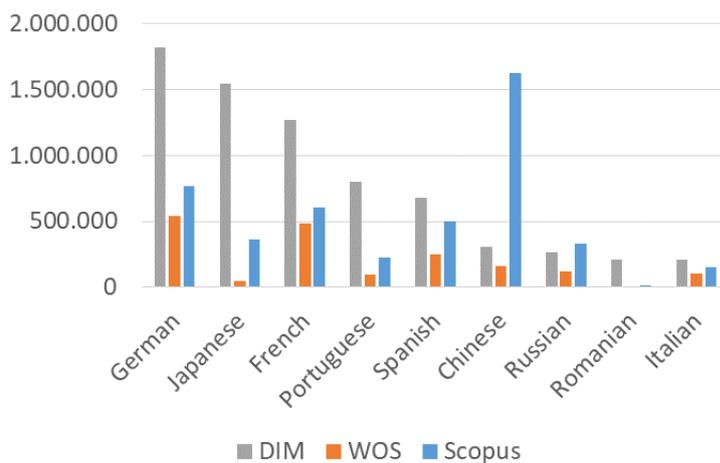


Abbildung 6: Anzahl Publikationen von weiteren 9 Sprachen in DIM20, WOS und Scopus 1996 - 2018

Aus dem Vergleich der Bestände nach Sprachen kann festgehalten werden, dass Dimensions den geringsten Anteil englischsprachiger Dokumente führt und gleichzeitig die größte Sprachenvielfalt

indexiert. Allerdings weisen die meisten Sprachen keine nennenswerten Anteile am Gesamtbestand der Datenbank auf. Und im Vergleich der absoluten Zahlen, die durch die unterschiedlichen Datenstände nur bedingt gültig ist, hat Dimensions klar die meisten englischsprachigen Dokumente aufzuweisen.

Kategorien

Jede Datenquelle verwendet ein anderes Klassifikationssystem. Außerdem werden in Web of Science und Scopus jeweils die Quellen einer oder mehreren Kategorien zugeteilt, während Dimensions die einzelnen Items klassifiziert.

Das Web of Science besitzt mit den Web of Science Subject Classifications sein eigenes Klassifikationssystem. Die Classifications sind in drei Oberkategorien (Arts & Humanities, Science & Technology und Social Sciences), sowie insgesamt 153 Unterkategorien (Research Areas) unterteilt. Alle Journals und Bücher der Web of Science Core Collection sind mindestens einer der 256 Subject Categories (Web of Science Categories) zugeordnet. Die Subject Categories sind auf die Research Areas abgebildet. Das Klassifikationssystem in Scopus ist ebenfalls ein datenbankeigenes System und besteht aus vier Subject Areas (Physical Sciences, Health Sciences, Social Sciences und Life Sciences) mit 334 Fields. Dahinter liegt die All Science Journal Classification (ASJC), die den Serientiteln von in-house experts zugewiesen wird. Eine Quelle kann mehrere Felder erhalten.

Dimensions beinhaltet gleich mehrere Klassifikationssysteme und erschließt auf Artikelbasis. In dieser Studie wurden die Fields of research (FOR) der 'Australian and New Zealand Standard Research Classification' (ANZSRC) untersucht: dieses System besteht aus 22 Oberkategorien (Divisions) und 157 Unterkategorien (Groups). Die Klassifizierung erfolgt maschinell und soll nach Aussage des Datenbankbetreibers alle Dokumentarten berücksichtigen können, nicht nur die Beiträge klassischer Zeitschriften (Bode et al., 2018).

Die Verschiedenheit der Systeme der drei Datenquellen, die sich u.a. durch die unterschiedliche Anzahl der Kategorien zeigt, erlaubt keinen direkten Vergleich der jeweiligen Bestände aufgrund der zugewiesenen Inhaltsklassen. Die Qualität der Sacherschließung kann auch nur mit guter Fachkenntnis der jeweiligen Beiträge beurteilt werden, wie dies beispielsweise Bornmann bereits 2018 auf Basis seiner eigenen Publikationen durchgeführt hat (Bornmann, 2018). Die Klassifikationssysteme von WOS und Scopus haben u.a. Wang und Waltman (Wang & Waltman, 2016) beleuchtet. Dabei galten jeweils die Kategorien zitierender Publikationen als Kriterium für die korrekte Zuordnung von Kategorien zu den Zeitschriften.⁸ Aufgrund der großen Unterschiedlichkeit der Systeme und des enormen Aufwands, den Vergleiche dieser Systeme bedeuten würden, ist das vorrangige Ziel dieser Studie, jeweils die Abdeckung der Klassifizierungen im Gesamtbestand zu ermitteln.

In der Dimensions-Datenbank gehören die Kategorien zu dem Bereich mit deutlichem Informationszuwachs zwischen den beiden Datenständen: gab es im April 2019 bei rund 55 % der indexierten Publikationen die Zuordnung zu einer der 22 Divisions, so waren es im März 2020 mit 72 % deutlich mehr. Auch die Groups wurden in DIM20 für mehr Publikationen vergeben (Anstieg von 55 % auf 67 %).

Dennoch sind 18.397.268 Publikationen ohne Zuordnung zu einer Disziplin ein großer Anteil (28 %) vom Gesamtbestand. Fast ¼ der Publikationen mit fehlender Zuordnung entfallen auf den

⁸ Die Autoren weisen darauf hin, dass der zitationsbasierte Analyseansatz nur eine Möglichkeit ist, die Zuordnung der Journals zu den jeweiligen Kategorien zu überprüfen ist, und empfehlen außerdem Analysen auf Text- oder Expertenbasis.

Publikationstyp Chapter und 9 % auf den Typ Proceedings. Über 60 % der Publikationen ohne Disziplin sind jedoch vom Typ Article, der, wie bereits gezeigt, nicht analog zu WOS und Scopus ausschließlich auf Forschungsbeiträge begrenzt ist.

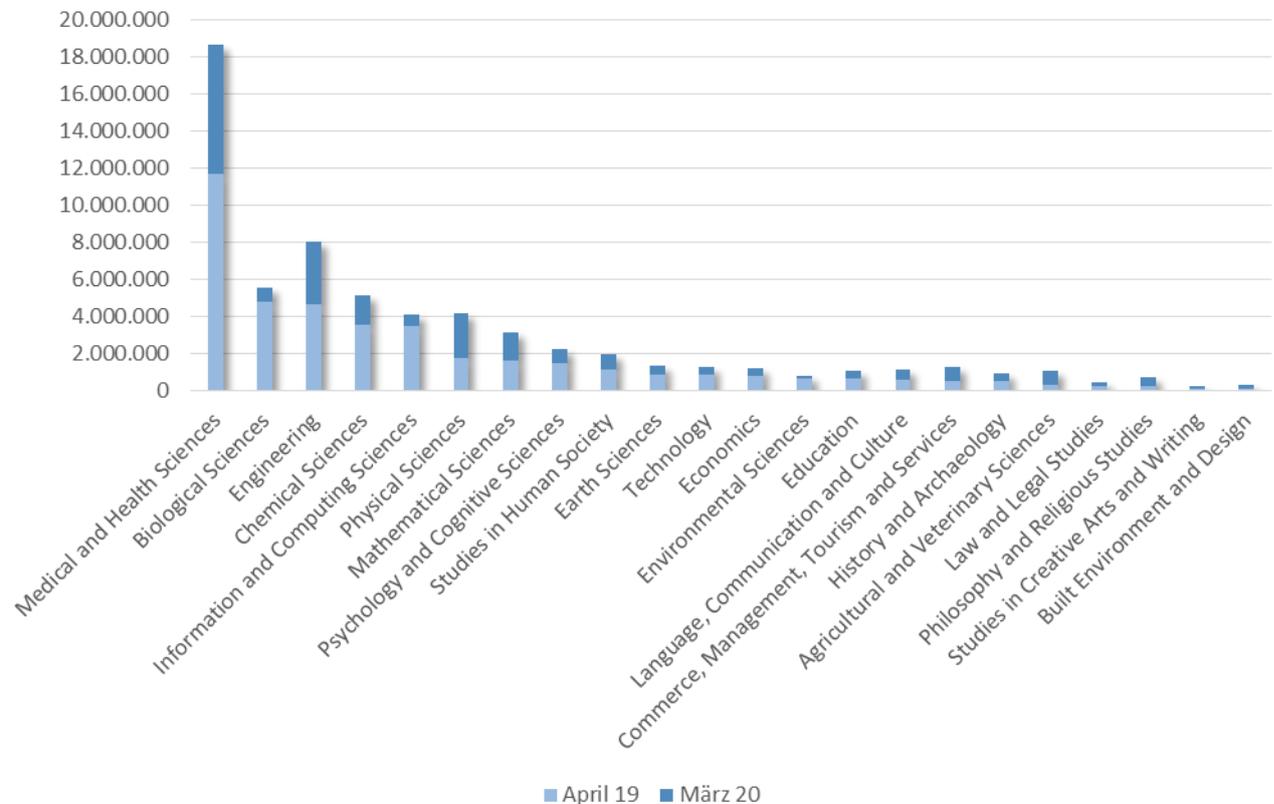


Abbildung 7: Anzahl der Publikationen je 'Division' in Dimensions, unterschieden nach den Datenständen von April 2019 und März 2020

Der Anteil der Publikationen mit einer Division-Zuordnung ist allerdings bei den älteren Beiträgen des untersuchten Zeitraums höher als bei den jüngeren und fällt von 77 % (1996) auf 69 % (2018). Eine Erklärung könnte sein, dass mehr Publikationen indiziert werden, für die zu wenige Informationen für eine fachliche Einordnung vorliegen. Als Beispiel kann eine Publikation mit dem Titel "Felsefe Medjmū 'asī"⁹ aus dem Jahr 2017 dienen. Die Sprache für dieses Buchkapitel wird mit Lettisch angegeben, was aber eine zweifelhafte Angabe ist.¹⁰ Entweder sind die selteneren Sprachen grundsätzlich ein Problem für die maschinelle Kategorisierung oder in diesem Fall die möglicherweise falsche Sprachzuordnung.

Von den mehr als 43 Millionen Publikationen aus Web of Science liegt für knapp 99 % eine Kategorienguteilung vor. Dazu wurden die Items mit einem Eintrag in der Tabelle *items_classifications* ermittelt.¹¹ Für Scopus hat diese Abfrage ergeben, dass 99,6% der Items klassifiziert sind und davon nahezu alle mit einer Kategorie der All Science Journal Classification.

9 DOI 10.1515/9783110695298-013

10 Der Autor lebte und lehrte in Istanbul (https://referenceworks.brillonline.com/entries/encyclopaedia-of-islam-3/baha-tevfik-COM_24274, 7.8.2020). Google Translate erkennt in 'Felsefe' ein türkisches Wort.

11 Erst seit 2017 ist der Anteil der Items ohne Kategorie auf über 1 % gestiegen, wobei 2017 ein ungewöhnlich hoher Wert von 6,6 % erreicht wurde.

In den Case Studies wird untersucht, ob und welche Auswirkungen sich für Dimensions durch den geringeren Anteil an Publikationen mit einer fachlichen Zuordnung ergeben. Für Analysen des Gesamtbestandes von Dimensions auf Basis von Forschungskategorien (z.B. online über die Analytical Views), fehlen nach jetziger Erkenntnis mehr als ein Viertel der verzeichneten Publikationen. Zu ähnlichen Ergebnissen kommen auch weitere Studien (Visser et al., 2020).

Durchschnittliche Anzahl Adressen je Publikation 2009-2018

Zur Untersuchung der Autorenadressen standen für diesen Studienteil nur die einer GRID ID zugeordneten Adressen zur Verfügung. Leider fehlen auch in DIM20 bei rund 48 % der Publikationen des Zeitraums 1996-2018 die GRID IDs.¹² Von der Berechnung eines Durchschnittswerts für die Adressen wird daher abgesehen.

Durchschnittliche Anzahl Autoren je Publikation 2009-2018

Sowohl Scopus als auch Dimensions beinhalten Publikationen ohne Angabe eines Autorennamens. In WOS scheint das für den Zeitraum 1996–2018 nicht der Fall zu sein, zu jeder indextierten Publikation ist auch der Name mindestens eines Autors oder einer Autorin erfasst. Bei Dimensions ist die Datenbasis für die Autorennamen deutlich besser als für die GRID IDs. Für die DIM20-Daten fehlen bei rund 8 % der Publikationen die Angaben zu den Autoren, bei DIM19 rund 10 %. Geprüft wurde, ob es Einträge zu einem Nachnamen gibt oder nicht. Ob die Autorennamen vollständig erfasst sind, oder nur teilweise, ist damit nicht geprüft worden. Bei Scopus fehlten im Schnitt jährlich für 2 % der Publikationen Autorenangaben für die untersuchten Jahre.

In allen drei Datenbanken ist die durchschnittliche Anzahl der Autoren je Publikation im Untersuchungszeitraum angestiegen. Am deutlichsten ist der Anstieg in den Daten des Web of Science um das 1,7-fache von 3 auf rund 5,2 Autoren je Publikation, am geringsten bei den Dimensions-Daten, die mit Datenstand April 2019 einen Anstieg um das 1,5-fache von 2,7 auf 3,9 Autoren verzeichnen.¹³

Tabelle 6: Durchschnittliche Anzahl Autoren je Publikation und je Datenquelle 2009-2018

Durchschnittliche Anzahl Autoren	
2009 - 2018	
Dim20	3,8
Web of Science	4,6
Scopus	4,2

¹² In Dimensions werden zur Darstellung von Organisation die GRID-ID aus der Global Research Identifier Database (<https://www.grid.ac/>) genutzt. Ihr kuratierter Datensatz enthält Informationen wie den Typ der Institution, die Geokoordinaten und Länderzuordnung, die offizielle Website und die Wikipedia-Seite.

¹³ Der aktueller Datenstand von März 2020 weist ähnliche Ergebnisse auf: Anstieg der Durchschnittswerte von 2,8 auf 3,9 um das 1,4-fache, für den Gesamtzeitraum beträgt die durchschnittliche Anzahl von Autoren je Publikation 3,5.

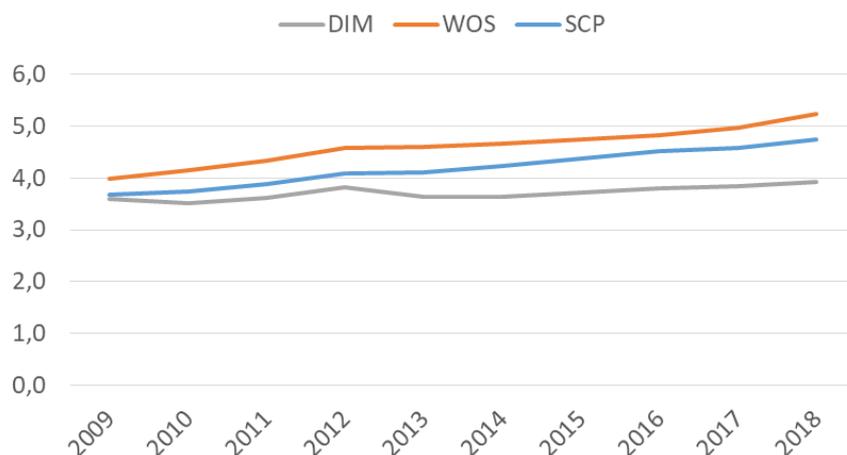


Abbildung 8: Entwicklung der durchschnittlichen Anzahl der Autoren je Publikation von 2009 bis 2018 in den Datenquellen

Beachtet werden muss jedoch die Heterogenität der verglichenen Datenbestände und dass eine hohe Anzahl an Publikationen ohne Autorenangabe auch die genannten Durchschnittswerte beeinflusst. Eine bessere Erkenntnis ist vom Vergleich dieser Aspekte bei den Case Studies zu erwarten, da hier die Publikationen für alle Quellen gleich definiert sind.

Publikationen mit Corresponding Author

Tabelle 7: Anteil der Publikationen mit Angaben zu einem Corresponding Author

Anteil Publikationen mit Corresponding Author 2009 - 2018	
DIM20	20 %
WOS	85 %
Scopus	85 %

In DIM20 werden bei 20 % der Publikationen ein Autor oder eine Autorin als 'Corresponding Author' ausgewiesen. Im älteren Datenstand gibt es keine Angaben zu 'Corresponding Authors'. Im Vergleich zu WOS und Scopus ist das nur ein sehr geringer Anteil: dort gibt es im selben Zeitraum bei jeweils 85 % der Publikationen entsprechende Angaben.

Bei den Zahlen für WOS muss berücksichtigt werden, dass es in der KB-Datenbank WOS_B_2019, auf der die Analyse der WOS-Daten beruht, Fehler in der Tabelle gegeben hat. Betroffen sind Publikationen mit mehr als einem 'Reprint Author'. Die Zahlen fallen daher vermutlich etwas zu gering aus.

Fehlt die Angabe eines 'Corresponding Authors' auf der Originalpublikation, hat WOS möglicherweise den Erstautor/ die Erstautorin mit dieser Rolle indexiert, so dass auch aus diesem Grund die Angaben nicht fehlerfrei sind.

Autorenadressen: Länder

Die Länderangaben stammen aus den Autorenadressen aller Publikationen, sofern diese im Einzelnen dokumentiert sind. Jedes Land wird dabei nur einmal je Publikation gezählt, auch wenn mehrere unterschiedliche Adressen eines Landes für eine Publikation existieren.

Wie schon in der Auswertung der Autorenadressen deutlich wurde, sind bei fast der Hälfte der Publikationen auch in DIM20 keine Adressen in Form von GRID IDs verzeichnet, so dass auch die

Länderangaben fehlen. Eine Auswertung der Adressen über den Gesamtbestand im Vergleich zu Scopus und WOS kann daher nur mit großer Vorsicht erfolgen.

Dennoch ähneln die Ergebnisse hinsichtlich der Länderanteile denen von Scopus und Web of Science: die Nationen USA, China, Großbritannien, Deutschland, Japan, Frankreich, Kanada und Italien gehören in allen drei Datenbanken zu den am stärksten publizierenden Ländern.

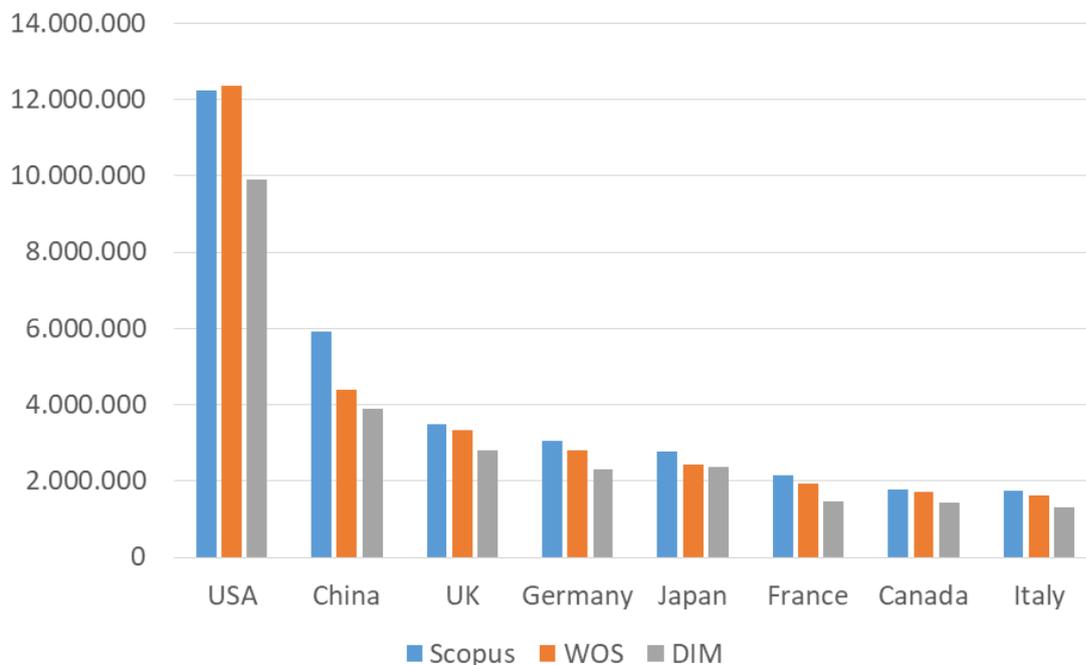


Abbildung 9: Anzahl Publikationen der acht stärksten Länder in DIM20, WOS und Scopus 1996-2018

Beim Vergleich der absoluten Zahlen der Top 8 – Länder aller drei Datenbanken wird jedoch deutlich, dass die Zahlen aus Dimensions teils deutlich geringer sind, obwohl die Datenbank die höchste Anzahl an Publikationen enthält und diese Länderauswertung auf dem aktuelleren Datenbestand beruht (65.589.398 statt 62.903.859 Publikationen). Dies ist sicherlich auf das Fehlen der Länderangaben bei knapp der Hälfte der Publikationen im untersuchten Zeitraum zurückzuführen. Die Reihenfolge dieser acht Länder nach Anzahl der Publikationen ist bei den Datenbanken nahezu identisch: lediglich in Dimensions sind die Plätze von Deutschland und Japan gegenüber der Reihenfolge bei WOS und Scopus vertauscht.

Funding Information

Der Anteil der Publikationen in Dimensions mit der Angabe mindestens einer Grant-Nummer und mit der Angabe von Förderorganisationen unterscheidet sich nach Datenständen nicht wesentlich, so dass für den Vergleich der Funding Informationen die Daten von DIM19 verwendet werden.

Im Vergleich zu Dimensions, wo nur 15,2 % der Publikationen über eine GRID ID Angaben zu einer Förderorganisation und 8,1 % Grant-Informationen beinhalten, ist der durchschnittliche Anteil von Publikationen mit Förderinformationen in WOS für den gewählten Zeitraum höher: 23,8 % der Publikationen haben Angaben zu mindestens einer Förderorganisation, 17,1 % zu Förderprogrammen (Grants). Im Verlauf der Jahre sind allerdings deutliche Steigerungen der Anteile zu verzeichnen: bis einschließlich 2005 ist der Anteil beider Informationen je Jahr geringer als 1 %, danach steigen die Anteile jährlich deutlich an auf zuletzt 53,3 % für die Organisationen und 41,5 % für die Grants im Jahr 2018.

Tabelle 8: Anzahl der Publikationen je Datenbank mit Angabe mindestens einer Förderorganisation 1996-2018

Jahr	Anzahl Publikationen mit Förderorganisationen		
	DIM19	WOS	Scopus
1996	57.812	821	1.248
1997	71.692	637	1.394
1998	98.121	843	2.332
1999	114.512	821	2.024
2000	132.587	854	3.033
2001	158.465	1.633	3.178
2002	175.234	1.707	4.829
2003	204.958	2.119	5.575
2004	235.294	3.157	2.064
2005	266.326	3.544	3.410
2006	298.087	56.410	3.064
2007	333.950	64.427	3.826
2008	377.187	257.344	4.248
2009	422.234	620.110	4.954
2010	463.447	719.416	6.567
2011	530.777	803.676	8.407
2012	591.247	870.122	17.141
2013	664.144	931.802	197.201
2014	740.319	972.346	379.693
2015	841.197	1.040.234	392.184
2016	904.759	1.171.872	502.473
2017	908.431	1.378.393	709.761
2018	999.984	1.345.033	771.741
	9.590.764	10.247.321	3.030.347

Scopus stellt bis 2012 kaum Informationen über Förderer und Programme bereit und erreicht bis zum Jahr 2018 6,3 % Anteil bei den Förderorganisationen und 19,4% bei den Grants. Im Durchschnitt werden für den Zeitraum aber nur 6,3 % bzw. 4,8 % erreicht.

Bei Betrachtung der absoluten Zahlen, z.B. der Förderorganisationen, zeigt sich eine stärkere Ähnlichkeit von DIM19 und WOS als dieser beiden Quellen mit Scopus. Bis 2008 liegt DIM19 gemessen an Publikationen mit Angabe von Förderorganisationen allerdings deutlich vor WOS (und Scopus) und wird erst 2009, nachdem WOS einen Anstieg von fast 400.000 Publikationen mit solchen Informationen aufweisen kann, von dieser Datenbank überholt. Scopus bleibt in absoluten Zahlen diesbezüglich immer weit unter den Größenordnungen von DIM19 und WOS.

Da unbekannt ist, wie viele Publikationen überhaupt Förderinformationen beinhalten könnten, ist eine Bewertung dieser Zahlen nicht möglich. Auch die Richtigkeit und Vollständigkeit der Daten war nicht Gegenstand der Untersuchung. Deutlich wird nur, dass WOS und Scopus erst in den jüngeren Jahren des Untersuchungszeitraums umfangreicher Informationen zu einer Förderung bereitstellen, wogegen Dimensions über den Gesamtzeitraum, wenn auch auf niedrigem Niveau, für einen beständigeren Anteil von Publikationen diese Informationen erfasst.

Zusammenfassung statistischer Überblick

Die quantitativen Auswertungen der vorangegangenen Kapitel geben einen Überblick über den Bestand der Datenbank Dimensions insgesamt und inwiefern dieser im Vergleich zu den bekannten Zitationsdatenbanken Web of Science und Scopus ähnliche Inhalte und Datenvollständigkeit aufweist. Vor allem die bibliografischen Kernfelder ergänzenden Informationen wie Autorenadressen, Kategorien und Funding Information, aber auch Dokumenttypen und Sprachen sind im Rahmen bibliometrischer Analysen von Bedeutung.

Die Ergebnisse insgesamt lassen den Schluss zu, dass Dimensions diese relevanten Daten zur Verfügung stellt und sich durch einen größeren Bestand an indexierten Publikationen hervorhebt. Zudem gab es in manchen Bereichen (z.B. der Anzahl von Publikationen mit Kategorien) im vergangenen Jahr positive Entwicklungen der noch jungen Datenbank, die auf weitere Verbesserungen der Inhalte hindeuten. Auffällig sind – vor allem im Vergleich zu den Ergebnissen des WOS – die hohen Anteile von Publikationen ohne Kategorisierung und ohne Angabe von Autorenadressen. Die stichprobenartigen Untersuchungen legen nahe, dass davon vor allem Publikationen betroffen sind, die im automatisch generierten Dimensions-Bestand, nicht aber in den ausgewählten Beständen von WOS und Scopus erfasst sind. Zu diesen Fragen wird es im Rahmen der Case Studies weitere Erkenntnisse geben.

4. Ergebnisse Case Studies

Untersucht wurden die Informationen derselben Publikationen einer Einrichtung aus allen drei Datenquellen. In den Vergleich sind nur Publikationen mit einem DOI eingegangen. Die Differenzen bezüglich der Anzahl an Publikationen für die Einrichtung wurden daher nur zahlenmäßig erfasst. Eine andere Vorgehensweise zum Vergleich und zur Abgrenzung wäre im Rahmen dieses Projekts zu aufwändig gewesen. Die Ergebnisse können jedoch die Unterschiede von Analysen bei identischen Publikationen aufzeigen.

Institutionenanalyse: Forschungszentrum Jülich

Die Auswahl der Publikationen erfolgte ausschließlich über die bereinigten und zugeordneten Adressen der jeweiligen Datenbanken: in Dimensions mit der GRID ID, in Scopus und Web of Science mit der Institutionen-Kodierung des Kompetenzzentrums Bibliometrie.

Beim Vergleich der Ergebnisse fielen Publikationen auf, die nur in einer oder zwei der Datenbanken nachgewiesen sind. Einige wurden stichprobenartig überprüft. Dabei wurden eine Reihe problematischer Fälle aufgedeckt: z.B. fehlende oder falsche Zuordnungen zur gesuchten Institution oder fehlende Adressen im Datensatz. Eine manuelle Korrektur der Daten wurde als zu aufwändig abgelehnt und so blieb es in dieser Untersuchung bei der (unvollständigen) Auswahl.

Die Untersuchungsmenge umfasst die Publikationen der Jahre 2009 bis 2018. Insgesamt wurden mit Hilfe des DOI 12.606 Publikationen für das Forschungszentrum Jülich (FZJ) ermittelt, die sowohl in Dimensions als auch in Web of Science und Scopus mit demselben Datenstand (KW 17 2019) erfasst sind.



Abbildung 10: Anzahl Publikationen des FZJ mit DOI je Datenbank und Schnittmengen

Für Dimensions wurden die Publikationen, die in DIM19 ermittelt werden konnten ausgewählt, die Informationen stammen jedoch aus DIM20.¹⁴

¹⁴ Einige Publikationen der DIM19 sind in DIM20 nicht mehr nachgewiesen. Der Abgleich erfolgte über die Publikations-ID der Datenbank (DIM19: 15.972, davon noch 15.771 Publikationen in DIM20).

Publikationsjahre

In 96 Fällen weicht das Publikationsjahr der Dimensions-Daten von den Angaben in Scopus oder WOS ab: häufiger von Scopus (82 mal) als von WOS (59 mal), von beiden zugleich in 45 Fällen. Zu diesen wurden zwei Stichproben durchgeführt. Dabei stellte sich heraus, dass in Dimensions das 'published first'-Jahr erfasst ist, während WOS und Scopus das Jahr des Erscheinens des Zeitschriften-Bandes angegeben haben. Das Datum in Dimensions stammt von CrossRef, wie online geprüft werden konnte.

Hinsichtlich der Publikationsjahre ist der Anteil an Abweichungen der drei DB voneinander verschwindend gering (0,76 %). Welche Auswirkungen das jeweils auf Analysen hat, hängt von der Größe der Untersuchungsmengen und z.B. den relevanten Zitationszahlen ab.

Dokumenttypen

Die generelle Verwendung des Dokumenttyps Article in Dimensions für alle Arten von Journal-Beiträgen erschwert auch in der Institutionenanalyse den Vergleich der drei Datenbanken: knapp 99 % (12.445) der Publikationen des FZJ werden in Dimensions als Article eingeordnet (Scopus 87 %, WOS 90 %). 10.831 Publikationen (86 %) erhalten in allen drei Datenquellen parallel den Typ Article, wobei ca. 91 % der Dimensions-Articles in WOS ebenfalls mit diesem Dokumenttyp erfasst sind, in Scopus sind es etwa 88 %. Einige wenige Publikationen vom Typ Chapter in Dimensions sind in WOS und Scopus als Article (Scopus 11/ WOS 4) erfasst. Aber alle Dimensions-Publikationen vom Typ Proceedings (119) sind in Scopus und WOS ebenfalls Conference Paper bzw. Proceedings Paper.

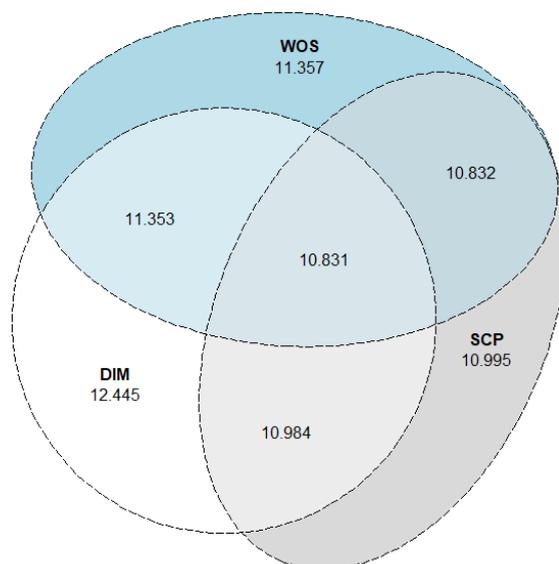


Abbildung 11: FZJ- Publikationen des Dokumenttyps Article in den Datenbanken

Articles aus Dimensions mit anderen Dokumenttypen in den anderen Datenquellen sind in Scopus größtenteils als Conference Paper (765/ 6 %) und Reviews (436/ 4 %) erfasst, ähnlich wie in WOS (427/ 3 %; 399/ 3 %), wo außerdem der Dokumenttyp Editorial Material eine größere Anzahl (145/ 1 %) der Dimensions-Articles auf sich vereint.

Im Vergleich zur Dokumenttypenverteilung der Untersuchungsmenge aller Publikationen von 1996 bis 2018 ist in der institutionellen Analyse aber eine größere Überschneidung der Dokumenttypen

festzustellen. Grund ist vermutlich die Definition der Publikationen als Forschungsoutput, die durch die Eingrenzung auf eine wissenschaftliche Einrichtung erzielt wird und damit bestimmte Dokumentarten, wie z.B. Book Reviews, unberücksichtigt lässt.

Quellen und Publisher

Bei Dimensions fehlen zwar bei einem sehr geringen Anteil der Publikationen (172/ 1,4 %) die Quellentitel, aber stichprobenartige Prüfungen der online verfügbaren Daten Mitte Mai haben ergeben, dass viele davon mittlerweile erfasst sind. Angaben zum Publisher fehlen nur in 3 Fällen. Die WOS-Daten sind in den Angaben zu den Quellen und Herausgebern vollständig, bei Scopus fallen die fehlenden Angaben zum Publisher bei 4.701 Publikationen auf. Dabei handelt es sich überwiegend um Dokumente vom Typ Article und es befinden sich viele große Zeitschriften darunter. Durch eine Online-Recherche Mitte Mai ist das Fehlen der Angaben bei einigen Datensätzen auch im aktuellen Datenbestand verifiziert worden, so dass nicht nur die fehlerhafte Datenlieferung an das Kompetenzzentrum eine Ursache sein kann.

Tabelle 9: TopTen-Quellen aller drei Datenbankergebnisse für das FZJ

Quelle	Anzahl Publikationen
PHYSICAL REVIEW B	469
JOURNAL OF NUCLEAR MATERIALS	267
PHYSICAL REVIEW LETTERS	250
ATMOSPHERIC CHEMISTRY AND PHYSICS	245
PLOS ONE	159
FUSION ENGINEERING AND DESIGN	156
NEUROIMAGE	144
SCIENTIFIC REPORTS	141
PHYSICAL REVIEW C	134
JOURNAL OF POWER SOURCES	131

Die zehn Quellen mit den höchsten Publikationszahlen haben jeweils identische Ergebnisse – wenn man die Titel *Physical Review B* und *Physical Review B - Condensed Matter and Materials Physics* einerseits sowie *Physical Review C* und *Physical Review C - Nuclear Physics* andererseits in der Scopus-Auswertung zusammenführt.

Die unterschiedlichen Längen der Ergebnislisten – Dimensions 1.502, WOS 1.806 und Scopus 1.581 Titel – zeigen die voneinander abweichenden Indexierungsmethoden dieses Informationsbereichs an, der sich für die Publisher ähnlich darstellt. Grundsätzlich aber ist diese Art von Information in allen Datenbanken verfügbar, wenn man von den größeren Problemen bei Scopus bezüglich der Publisher-Daten absieht.

Publisher werden in Dimensions fast vollständig genannt, nur in drei Fällen fehlt ein Publisher. Die Änderungen im Verlagswesen sind enorm und spiegeln sich nicht nur in wechselnden Verlagsnamen wider. Da Dimension durch Normierungen in diesem Feld einiges vereinfacht, steht der Elsevier-Verlag als Herausgeber zwar bei allen drei Ergebnisauswertungen an erster Stelle, jedoch mit sehr unterschiedlichen Zahlen. Dimensions verzeichnet unter Elsevier 3.979 Publikationen, WOS unter ELSEVIER SCIENCE BV 1.866 Publikationen – aber auch mit den weiteren 18 Verlagen, die Elsevier im Namen führen, kommt man bei WOS insgesamt nur auf 3.886 Publikationen. Welche Verlage bei Dimensions unter der einen Bezeichnung zusammengeführt wurden und ob diese Normierung korrekt

ist, ist nicht Aufgabe dieser Untersuchung. Die Beispiele sollen lediglich einen Einblick in die vorhandene Datenstruktur geben.

Sprachen

Dimensions zählt 29 Publikationen weniger in englischer Sprache als WOS und Scopus (je 12.563), darunter 22 ohne Angabe zu einer Sprache. Für die weiteren sind die Sprachen Deutsch (1), Afrikaans (2) und Rumänisch (7) angegeben. Scopus listet 2 Publikationen doppelt, die jeweils in zwei Sprachen erfasst sind. Fünf Publikationen, die sowohl WOS als auch Scopus als deutschsprachig indexieren, werden in DIM20 als englischsprachig erfasst. Insgesamt liegt der Anteil fehlerhafter Sprachbezeichnungen von Publikationen in dieser Auswertung bei nur 0,3 %.

Kategorien

Auch der Bereich der Kategorien wurde hinsichtlich der Abdeckung der Publikationen in der Untersuchungsmenge und beispielhaft zu erkennbaren Unterschieden anhand einzelner Publikationen geprüft.

In Dimensions fehlen lediglich bei 116 Publikationen Angaben zu Kategorien („Divisions“) und damit bei 0,9 % der ausgewählten Datensätze. Bei WOS sind es noch weniger, nämlich 62 Veröffentlichungen ohne Kategorie, hier handelt es sich in der Regel um Konferenzveröffentlichungen. Die Scopus-Daten sind dagegen vollständig: allen Publikationen in der Auswahl sind auch Kategorien („Fields“) zugewiesen. Während im Gesamtbestand von Dimensions also noch ein verhältnismäßig großer Anteil (28 %) von Publikationen ohne Kategorie auffiel (vgl. S. 17), ist in dieser konkreten Studie die Zahl der Publikationen, die keinem Forschungsfeld zugewiesen sind, sehr klein.

Tabelle 10: FZJ-Publikationen: Kategorien je Datenbank für einen Artikel im Vergleich

DIM – Division - Group	WOS – Subject Category	SCP – Field (ASJC)
Agricultural and Veterinary Sciences	Environmental Sciences	General Chemistry
Biological Sciences	--	Environmental Chemistry
<i>Crop and Pasture Production</i>	--	Environmental Engineering
<i>Plant Biology</i>	--	Health, Toxicology and Mutagenesis

Untersuchte Publikation: Mehmood, K. et al. (2017). No effect of digestate amendment on Cs-137 and Sr-90 translocation in lysimeter experiments. Chemosphere, 172, 310–315. doi:10.1016/j.chemosphere.2016.12.134.

In Dimensions sind 110 Publikationen des FZJ der Division 'Agricultural and Veterinary Sciences' zugeordnet. Die Publikationen haben noch bis zu zwei weitere Divisions erhalten. Ähnlich ist es bei den WOS-Daten, die in der Summe aller DOI-Subject Category-Kombinationen kleiner als Dimensions bleiben (194 statt 239) und damit je Publikation eine geringere Anzahl an Kategorienzuordnungen haben. Die Subject Categories sind eigentlich kleinteiliger, so dass den 22 Divisions in Dimensions 250 Subject Categories gegenüberstehen.

In Scopus haben dieselben 110 Publikationen bis zu fünf Kategorien (Fields), es gibt 257 Kombinationen der DOIs mit den Kategorien.

Die unterschiedlichen Methoden der Kategorisierung und Systeme führen zwangsläufig zu Ergebnisunterschieden, wie am Beispiel in Tabelle 10 aufgezeigt werden kann.

Anzahl Autoren

Bei Web of Science und Scopus kommen abweichende Angaben von Autoren und Autorenadressen im Vergleich zur Originalpublikation vereinzelt vor. Der Vergleich der jeweils erfassten Autoren bei allen drei Datenbanken soll aber zeigen, ob zumindest ähnliche Werte erreicht werden, oder die Differenz sehr groß ist. Durch stichprobenartige Überprüfung bei unterschiedlicher Zählung können ggf. Erkenntnisse über Fehlerquellen gewonnen werden. Insgesamt sollten vereinzelt oder kleinere Abweichungen weniger stark ins Gewicht fallen, da die Datenbasis recht groß ist.

Den Berechnungen der WOS- und Scopus-Daten liegen die Werte aus dem Feld "D_AUTHOR_CNT" der KB Items-Tabellen zugrunde. Für die Berechnungen der Anzahl der Autoren in Dimensions ist mit DIM20 der neuere Datenstand genutzt worden. Gezählt wurden jeweils die *Fullnames* der Autoren, damit Autoren desselben Nachnamens nicht als Dublette behandelt werden. Identische Namen je Publikation sind zwar auch bei der Fullname-Zählung möglich, sind aber nicht so häufig. Die Summe aller Autoren ergab für DIM20 116.398, für WOS 116.380 und für Scopus 117.007.

Im Schnitt indexieren DIM20 und WOS demnach durchschnittlich 9,2 Autoren je Publikation, Scopus 9,3.

Trotz unterschiedlicher Zählmethoden der Autorennamen bei DIM20 einerseits und WOS und Scopus deuten diese Ergebnisse auf eine hohe Übereinstimmung bei der Erfassung von Autorennamen. Mit diesem Ergebnis von drei unterschiedlichen Quellen kann gefolgert werden, dass die Erfassung der Autoren i.d.R. bei allen drei Quellen nah an den Originalen und damit von hoher Qualität ist.

Die Überprüfung einzelner Publikationen mit stark abweichenden Autorenzahlen offenbart, dass z.B. von WOS oder Scopus auch aus den Acknowledgements oder kooperierenden Arbeitsgruppen Namen als Autorennamen indexiert wurden.¹⁵

Tabelle 11: Autoren und Corresponding Authors in den Publikationen des FZJ jeweils in Dimensions, WOS und Scopus

FZJ 2009-2018	DIM20	WOS	SCP
Anzahl Autoren absolut	116.398	116.380	117.007
Durchschnittliche Anzahl Autoren je Publikation	9,2	9,2	9,3
Anzahl der Publikationen mit Corresponding Authors	5.559	12.596	11.468
Anteil der Publikationen mit Corresponding Authors	44,1%	99,9%	91,0%

Corresponding Authors

Publikationen, die Corresponding Authors bzw. Reprint Authors benennen, weichen in den drei Datenbanken in der Anzahl stark voneinander ab: Dimensions hat aus der Untersuchungsmenge nur 5.559 Publikationen mit Angabe eines Corresponding Authors, WOS 12.596 und Scopus 11.468. Damit ist für weniger als die Hälfte (44,1 %) der 12.606 Publikationen in Dimensions ein Autor als Corresponding Author benannt. Bei WOS sind die Informationen diesbezüglich fast vollständig (zu 99,9 % der Publikationen ist mindestens einen Corresponding Author angegeben), bei Scopus liegen sie bei 91 %.

Verglichen mit den jeweiligen Anteilen an den Gesamtpublikationen für den Zeitraum 2009 - 2018 hat sich der Anteil bei Dimensions mehr als verdoppelt und ist auch bei WOS und Scopus um fast 15 bzw. 6 % angestiegen.

Anzahl Adressen

Die Anzahl der Adressen insgesamt oder im Durchschnitt je Publikation kann ein Hinweis auf Unterschiede bezüglich der Vollständigkeit in den Datenbanken sein. In der vorliegenden Untersuchung werden die Adressen aus Dimensions ausschließlich als normierte Einträge über die GRID IDs betrachtet. Web of Science und Scopus haben durch Schreibvarianten eine höhere Anzahl von unterschiedlichen Adressen und ein direkter Vergleich ist daher nicht möglich.

Zu allen Publikationen existieren in Dimensions und WOS Adresseinträge.¹⁶ Zählt man in DIM20 jede GRID ID je Publikation nur einmal, so erhält man insgesamt 49.048 Adressen. Für das WOS ist eine vergleichbare Zählung ohne normierte Adressen nicht möglich, aber die Zählung für den Eintrag Orga1 (Items-Tabelle) ergibt insgesamt 55.500 Adressen. Für Scopus liegt in der KB-Datenbank nur die Zählung der vollen Adresseinträge vor.

Im Vergleich zwischen DIM20 und WOS gibt es bei 4.430 Datensätzen abweichende Zahlen, also bei gut einem Drittel der Publikationen. Die häufigste Abweichung ist die größere Anzahl von erfassten Institutionen bei WOS, was größtenteils an dessen nicht-normierten Bezeichnungen liegt. Vereinzelt sind aber auch Namen von Forschungseinrichtungen nicht durch GRID normalisiert,¹⁷ so dass diese mit dem angewendeten Verfahren auch nicht als Autorenadressen gezählt wurden, grundsätzlich aber erfasst sind und mit einer anderen Methode zählbar wären. Das Ergebnis des Vergleichs zwischen WOS- und DIM20-Daten legt also nahe, dass die Erfassung der Autorenadressen in beiden Datenbanken von ähnlicher Qualität und Abdeckung ist.

Länder

Die Länderauswertungen der drei Datenbanken fallen nicht identisch aus, abgesehen von den Ergebnissen für 'Germany'. Um die Betrachtung zu vereinfachen, wurde jedes Land je Publikation nur einmal gezählt.

Beim Aufsummieren aller Länderzählungen wird klar, dass die Datenbanken aus denselben Publikationen insgesamt unterschiedlich viele Länderzahlen durch die Adressen erfasst haben: Dimensions am wenigsten mit 27.032, Scopus am meisten mit 28.768 und WOS liegt dazwischen mit 28.520. Auch die Liste der erfassten Länder ist jeweils unterschiedlich lang: WOS dokumentiert nur 106 verschiedene Länder, Scopus 110 und Dimensions 102. Letzteres ist wohl hauptsächlich darauf zurückzuführen, dass die Datenbanken mit einigen Spezialfällen unterschiedlich umgehen: Scopus benennt Hongkong als Land, wo WOS die Publikationen China zuordnet. Puerto Rico ist ein ähnlicher Fall und wird in Scopus als eigenes Land, in WOS als USA geführt. Fehlerhafte Datensätze oder Zuordnungen sind in allen drei Datenbanken Realität. Die abweichenden Publikationszahlen für die Länder können aber – anders als bei den Institutionen – nicht so stark mit den unterschiedlichen Voraussetzungen für die Zählungen zusammenhängen, da jedes Land, auch wenn es durch verschiedene Adressschreibweisen bei WOS oder Scopus mehrfach erfasst ist, je Publikation nur einmal gezählt wird. Die Überprüfung einzelner Publikationen weist aber wieder auf das Problem hin,

¹⁶ Adresseinträge in Scopus sind teilweise fehlerhaft oder leer. Das hängt möglicherweise mit den fehlerhaften Datenlieferungen durch Elsevier an das KB zusammen.
¹⁷ z.B. DOI: 10.1002/2013jd021435.

dass bei DIM20 nur die GRID IDs gezählt wurden und einige Adressen zwar erfasst sind, aber in keine GRID ID auflösen. Auf diese Weise lassen sich die teilweise geringeren Publikationszahlen für manche Länder bei Dimensions im Vergleich zu WOS oder Scopus erklären.

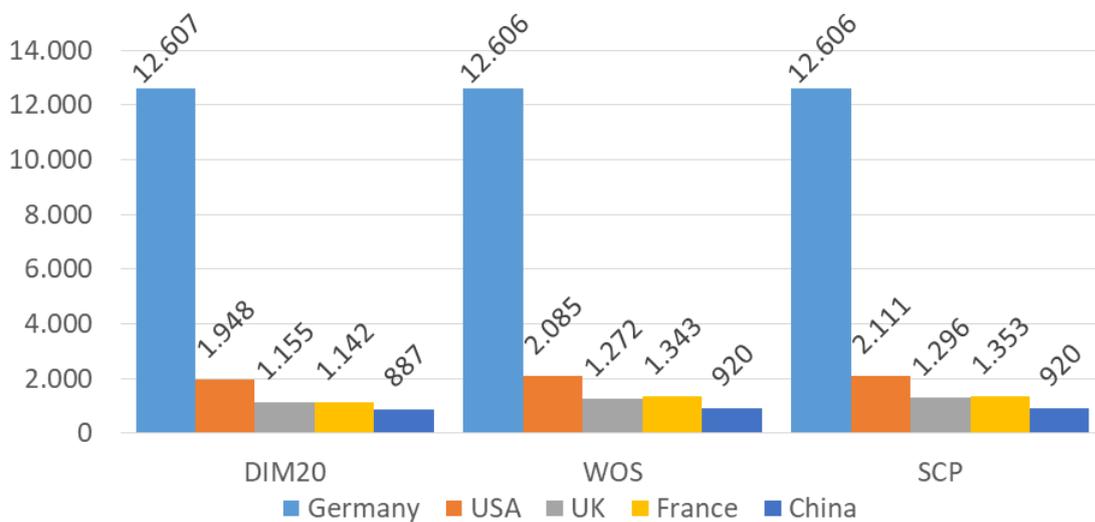


Abbildung 12: Die 5 Länder mit den höchsten Anteilen in der Analyse zum FZJ in allen 3 Datenbanken im Vergleich

Funding Information

In Dimensions sind 7.446 Publikationen (59 %) mit Informationen zu einem Förderer versehen. Es handelt sich jeweils um normierte Adressen, die im Datensatz als GRID ID eingetragen sind. Informationen zu einem Förderprogramm sind für 3.928 Fälle vorhanden, das sind 31 % der Publikationen. Es gibt 620 unterschiedliche Förderer in der Untersuchungsmenge. Die Anzahl der unterschiedlichen Grant-Nummern liegt allerdings bei 4.526 – viele Veröffentlichungen basieren auf mehr als einem Forschungsprogramm.

Web of Science nennt zu 9.084 (71 %) Publikationen mindestens einen Förderer und führt zusätzlich zu 7.109 (56 %) Publikationen Förderprogramme auf. Die Anzahl unterschiedlicher 'Grantnumbers' beträgt 9.854, die Anzahl der 'Fundingorganizations' 11.815. Letztere liegt tatsächlich niedriger, da die Namen im Gegensatz zu Dimensions nicht einheitlich sind. Unterschiedliche Schreibweisen desselben Organisationsnamens führen in WOS zu einer erhöhten Zahl von Förderern.

Nur 1.758 (knapp 14 %) Publikationen in Scopus enthalten eine Funding Information. Für diese Publikationen sind jeweils Angaben zu den Förderorganisationen und zu den Förderprogrammen angegeben.

Bei genauerer Untersuchung der Daten zu einer Publikation, zu der es in Dimensions, aber nicht in WOS Förderinformationen gab, stellte sich heraus, dass die genannten Förderorganisationen den Acknowledgments entnommen waren. Allerdings handelt es sich dabei um Einrichtungen, die Daten für die Untersuchung bereitgestellt hatten und nicht um Geldgeber. Zudem ist unklar, warum weitere, in den Acknowledgments genannte Einrichtungen nicht aufgeführt sind. Die Untersuchung einer anderen Publikation zeigt jedoch korrekte Daten über Förderer und Förderprogramm in Dimensions, die in WOS komplett fehlen. Die genannten Beispiele weisen darauf hin, dass nicht nur die Anzahl der Publikationen mit Funding Informationen Einfluss auf die Ergebnisse der unterschiedlichen Datenquellen haben dürfte. Bei der Frage nach der Anzahl von Publikationen des FZJ mit einer Förderung durch die Deutsche Forschungsgemeinschaft (DFG) ergeben sich erwartungsgemäß

unterschiedliche Resultate: Dimensions nennt die DFG bei 2.318 Publikationen als Förderer, WOS bei 2.860 Publikationen.¹⁸

Im Vergleich zu den Ergebnissen über die Funding Informationen der Gesamtdatenbankbestände (1996 - 2018) fallen bei allen drei Datenbanken die Anteile von Publikationen mit den jeweiligen Informationen an der Untersuchungsmenge (FZJ, 2009 – 2018) deutlich höher aus.

Institutionenanalyse: Fraunhofer-Institut für Keramische Technologien und Systeme (IKTS)

Auch für das Fraunhofer-Institut wurden die Publikationen der Jahre 2009 bis 2018 betrachtet. Insgesamt wurden mit Hilfe der DOI 473 Publikationen für das IKTS ermittelt, die sowohl in Dimensions als auch in Web of Science und Scopus zum selben Datenstand (KW 17 2019) erfasst sind.

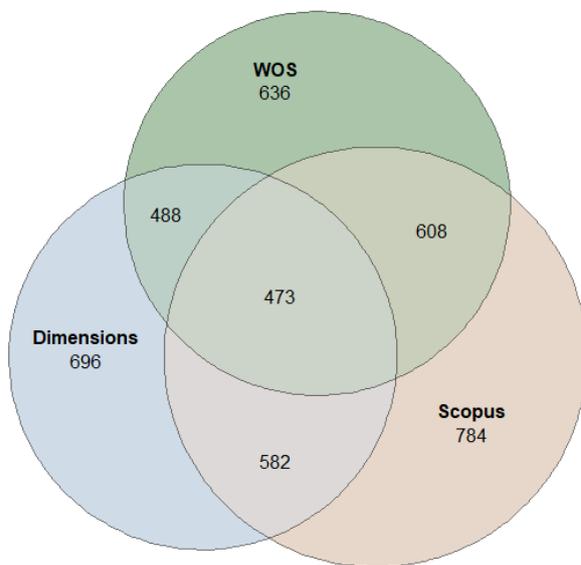


Abbildung 13: Anzahl Publikationen des IKTS mit DOI je Datenbank und Überschneidungsmengen

Mit Hilfe der jeweils gültigen Institutionen-Normierungen wurden in Scopus die meisten Publikationen (1.024, 784 mit DOI) und in Dimensions die wenigsten (696, alle mit DOI) ermittelt. WOS enthielt für das Institut 760 Publikationen (636 mit DOI). Die größte Überschneidung von Publikationen mit DOI gibt es zwischen Scopus und WOS mit 608 Publikationen.

¹⁸ In WOS sind die Förderernamen nicht bereinigt worden, sondern mit den folgenden Schreibweisen in SQL grob abgefragt worden: '%DFG%', '%German Research Foundation%', '%Deutsche Forschungs%', '%German Science Found%'

Publikationsjahre

Bei vier Publikationen gibt es abweichende Publikationsjahre, entweder von Dimensions oder von Scopus.

Tabelle 12: Publikationen des IKTS nach Publikationsjahren je Datenbank

	DIM20	WOS	Scopus
2009	27	26	26
2010	25	26	26
2011	39	38	38
2012	31	32	32
2013	42	42	42
2014	72	72	73
2015	45	45	43
2016	58	58	59
2017	64	64	64
2018	70	70	70

Dokumenttypen

Dimensions verzeichnet 453 Article und 20 Proceedings. Letztere haben in WOS und Scopus entsprechende Dokumenttypen, aber von den Articles sind nur 420 in WOS und 400 in Scopus ebenfalls Articles. Die anderen 23 bzw. 53 verteilen sich auf 5-6 andere Dokumenttypen.

Quellen und Publisher

Für 20 Publikationen von Dimensions sind keine Quellen verzeichnet. Dabei handelt es sich ausnahmslos um Konferenzpublikationen, für die aber Publishereinträge vorhanden sind (IEEE, SPIE, Electrochemical Society). Bis auf diese Fälle sind jedoch die Angaben für die Quellen und Publisher sowohl bei Dimensions als auch bei WOS vollständig.¹⁹

Abweichungen der Datenbanken bezüglich Quellen oder Publisher gibt es in geringem Maße, die dieselben Gründe wie in der FZJ-Analyse habe. Die Quellen mit den meisten Publikationen haben in allen drei Datenbanken entsprechende Publikationszahlen.

Tabelle 13: Top 8 Quellen aller Datenbanken der IKTS-Publikationen

Journal of the European Ceramic Society	52
Advanced Engineering Materials	19
International Journal of Refractory Metals and Hard Materials	18
Ceramics International	14
Chemie Ingenieur Technik	14
Journal of the American Ceramic Society	14
Electrochimica Acta	12
Journal of Thermal Spray Technology	12

¹⁹ Auffällig ist wieder die große Anzahl von Publikationen in Scopus, für die kein Publisher angegeben ist (169), vermutlich aufgrund der fehlerhaften Datenlieferung an das KB.

Sprachen

In allen drei Datenbanken werden die Publikationen mit den Sprachen Englisch und Deutsch erfasst. Die geringen Abweichungen erscheinen nicht relevant, i.d.R. handelt es sich um Beiträge deutschsprachiger Zeitschriften, die wohl aufgrund englischsprachiger Abstracts in wenigen Fällen mit der Sprache Englisch indexiert wurden.

Kategorien

Alle Datenbanken weisen eine hohe Abdeckung für die Publikationen mit Kategorienvergabe auf. Dimensions hat mit 450 von 473 und damit 95 % noch den geringsten Wert, Scopus mit 100 % den höchsten. Grob betrachtet ist die häufigste Kategorie aller drei Quellen die Materialwissenschaft (DIM: Materials Engineering, WOS: Materials Science, Scopus: Materials Chemistry).

Anzahl Autoren

WOS, Scopus und Dimensions geben insgesamt eine Summe von 2.716 Autoren für alle 473 Publikationen an und haben zu jeder einzelnen Publikation eine identische Zählung.

Corresponding Authors

WOS benennt für fast alle (472) Publikationen einen 'Reprint' bzw. 'Corresponding Author', Scopus für 440 und Dimensions für nur 275 Publikationen. Das Ergebnis bestätigt die deutlich geringere Erfassung dieses Autorentyps bei Dimensions im Vergleich zu den beiden anderen Datenbanken: nur 58 % der Publikationen enthalten diese Information. Im Vergleich zur Auswertung der FZJ-Publikationen liegt der Anteil allerdings um knapp 14 % höher.

Anzahl Adressen

Dimensions erfasst in der Untersuchungsmenge 1.223 GRID IDs, WOS zählt insgesamt 1.280 Organisationen. Wie in allen Auswertungen wird jede Organisation je Publikation nur einmal gezählt, auch wenn mehrere Autoren einer Einrichtung beteiligt sind. Berücksichtigt man fehlende GRID IDs für vorhandene Adressen bei Dimensions und eine mögliche Ungenauigkeit der Institutionenzählung bei WOS, so kann eine hohe Übereinstimmung der Ergebnisse vermutet werden.

Länder

Die Länderauswertung ergibt eine relativ hohe Übereinstimmung für die Top 8 Länder mit den häufigsten Publikationsbeteiligungen. Aus den GRID-IDs können 42 unterschiedliche Länder für Dimensions ermittelt werden, aus den WOS- und Scopus-Adressen jeweils 44. Die Anzahl je Land ist bei WOS und Scopus fast identisch, zu Dimensions gibt es aber teilweise deutliche Abweichungen.

Tabelle 14: Top 8 Länder der Publikationen des Fraunhofer IKTS in den Datenbanken

	DIM	WOS	SCP
Germany	473	473	473
South Africa	28	22	22
United States	12	16	16
Switzerland	14	13	13
Austria	12	12	11
United Kingdom	17	11	11
China	12	10	9
Italy	12	9	10

Funding Information

In Dimensions haben 219 und damit 46 % der Publikationen dieser Untersuchungsmenge mindestens eine Förderorganisation in ihren Datensätzen und von diesen haben 71 (15 %) zusätzlich ein Förderprogramm verzeichnet. Nachgewiesen sind 74 unterschiedliche Förderorganisationen und 63 Programme über eine GRANT ID. WOS hat einen höheren Anteil an Publikationen, die eine Fundingorganisation angeben, es sind 293 (62 %). Auch die Zahl der Publikationen mit Grantinformationen ist entsprechend höher und liegt mit 223 bei 46 % Anteil in der Untersuchungsmenge.

Autorenanalyse: Christian Drost

Neben der Analyse auf Einrichtungsebene wurde zusätzlich eine Case Study auf Personenebene durchgeführt, um auch hier mögliche Unterschiede in den untersuchten Datenbanken zu prüfen. Als aktuelles prominentes Beispiel haben wir uns für Prof. Christian Drost entschieden, mit dessen Researcher Profile Dimensions am 14.9.2020 für seine "Researcher details pages" wirbt.²⁰ Auch aufgrund des relativ eindeutigen Namens haben wir uns für diesen Wissenschaftler entschieden, um den Bereinigungsaufwand möglichst gering zu halten. Die Untersuchungsmenge umfasst alle Publikationen, die unter dem Namen veröffentlicht wurden, und hat keine Einschränkung auf Publikationsjahre.

Insgesamt wurden über einen Abgleich der DOI für Drost 277 Publikationen ermittelt, die in allen drei Datenbanken parallel erfasst sind.

20 app.dimensions.ai/details/entities/publication/author/ur.010432516417.62 (28.07.2020)

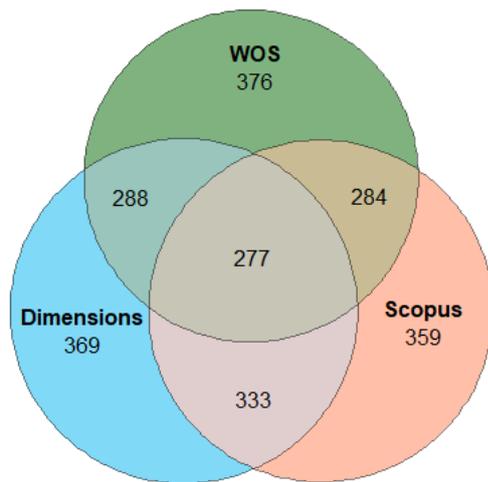


Abbildung 14: Ergebnisse für C. Drost in den Datenbanken -- Schnittmengen

Die tatsächlichen Ergebnisse für den Autor sind leicht unterschiedlich: die meisten Publikationen verzeichnet WOS mit 376, Dimensions weist 369 Publikationen nach und Scopus 359. Die wenigsten DOI hat WOS mit 302 verzeichnet. Eine komplette Überprüfung aller Publikationen in allen Datenbanken ist aufgrund des Aufwands nicht erfolgt, aber durch Stichproben wurden einzelne Gründe für das Fehlen von Publikationen ermittelt. Eine Publikation, die in Dimensions mit DOI nachgewiesen ist, ist z.B. in WOS als Meeting Abstract ohne DOI, aber in Scopus gar nicht verzeichnet (Scopus indexiert diesen Dokumenttyp generell nicht). In einem anderen Fall konnte eine Publikation in Dimensions nicht gefunden werden, weil Nachname und Vorname im Autoren Datensatz vertauscht waren. Und auch gänzlich fehlende Einträge von Autorennamen kamen in allen Datenbeständen vor.

Tabelle 15: Anzahl der Publikationen von C. Drost in den unterschiedlichen Datenquellen

	P	P_doi	P_DIM∩WOS∩SCP	P_exklusiv	Anteil P_exklusiv
Dimensions	369	362	277	92	25%
WOS	376	302	277	99	26%
SCOPUS	359	343	277	82	23%

Häufiger ist eine Publikation zumindest in 2 Datenbanken indexiert, besonders in WOS fehlt aber der DOI für den Abgleich. Und vereinzelt fehlt ein Nachweis für Dimensions auch nur im älteren Datenbestand, nicht aber im aktuelleren. Dass diesbezüglich von einem grundsätzlichen Problem auszugehen ist und Dimensions mit der Aktualisierung hinter WOS oder Scopus zurückbleibt, kann nicht gefolgert werden. Aufgrund der kurzen Existenz der Datenbank fehlt es zu dieser Frage an Erfahrungswerten.

Aufgrund fehlender DOI (vor allem in WOS) oder anderer Gründe gehen je nach Datenbank zwischen 23 % (Scopus) und 26 % (WOS) der Ergebnisse nicht in die gemeinsame Untersuchungsmenge von 277 Publikationen ein. Die Stichprobenergebnisse lassen aber vermuten, dass bei einem vollständigen Abgleich der Publikationen eine hohe Übereinstimmung der Ergebnisse erzielt werden würde.

Publikationsjahre

Zwei Publikationen zeigen abweichende Publikationsjahre bei einer der Datenbanken, so dass insgesamt 4 Jahre nicht-identische Zahlen anzeigen.

Tabelle 16: Publikationsjahre der Publikationen von C. Drosten in allen drei Datenbanken

Jahre	DIM	WOS	SCP
2001	1	1	1
2002	1	1	1
2003	4	4	4
2004	10	10	10
2005	4	4	4
2006	5	5	5
2007	8	8	8
2008	8	8	8
2009	15	15	15
2010	20	20	20
2011	16	16	16
2012	20	20	20
2013	28	28	28
2014	37	37	37
2015	34	33	34
2016	15	16	15
2017	18	18	18
2018	27	26	26
2019	6	7	7

Dokumenttypen

Wie aufgrund der bisherigen Erkenntnisse nicht anders zu erwarten, sind die Publikationen des Autors in Dimensions fast ausschließlich als Article gekennzeichnet. WOS unterscheidet 5 Dokumenttypen, Scopus sogar 9 (und führt eine Publikation ohne Dokumenttyp). In eine Analyse mit ausgewählten Dokumenttypen (z.B. Article, Review und Letter) gingen aus den Ergebnissen jeweils eine unterschiedliche Anzahl von Publikationen ein: 267 aus Web of Science, 261 Scopus und 276 aus Dimensions. Für die weiteren möglichen Auswertungen, z.B. Indikatoren auf Basis von Zitationszahlen, können solche Unterschiede von Bedeutung sein. Im konkreten Fall haben sich die Differenzen als gering erwiesen.²¹

Tabelle 17: Drogen-Publikationen:Gegenüberstellung der Dokumenttypen aller drei Datenbanken

WOS	SCP	DIM	Anzahl Publikationen
Article	Article	Article	231
Article	Review	Article	8
Article	Conference Paper	Article	4
Article	Note	Article	2
Article	--	Article	1
Article	Short Survey	Article	1
Correction	Error	Article	3
Editorial Material	Editorial	Article	3
Editorial Material	Article	Article	2
Editorial Material	Letter	Article	1
Editorial Material	Note	Article	1
Letter	Letter	Article	17
Review	Review	Article	2
Review	Chapter	Chapter	1

Quellen

Unter Verwendung der Quellenliste aus dem Dimensions-Ergebnis lassen sich knapp 90 % der Publikationen aller drei Datenbanken vergleichend zuordnen. Die übrigen 10 % (28) bei Scopus bzw. 11 % (31) bei WOS ließen sich nicht ohne weitere Prüfung und Bearbeitung der Titel abgleichen. Betroffen sind 7 Titel bei WOS und 7 bei Scopus, davon 4 mit demselben Titel. Die *Proceedings of the National Academy of Sciences* sind in WOS und Scopus beispielsweise ergänzt um 'of the United States of America'. Den Zeitschriften *Journal of Infectious Diseases* und *ISME Journal* sind in Dimensions richtigerweise der (englische) Artikel vorangestellt.

Das *Filaria Journal* heißt bereits seit 2008 *Parasites & Vectors*, wie WOS und Scopus für eine Publikation von 2018 richtig erfasst haben. Dimensions verwendet dagegen den Vorgängertitel der Zeitschrift.

Der umgekehrte Fehler fiel bei der Zeitschrift *mSphere* auf. Der Beitrag war eigentlich in der Zeitschrift *Clinical and Vaccine Immunology* erschienen, die 2018 in *mSphere* aufgegangen war. Dimensions

21 Zur jeweiligen Publikationsauswahl wurden in allen drei Datenbeständen (Dimensions mit Datenstand 2019) die Zitationszahlen ermittelt und die Zitationsrate berechnet. Der geringste Wert (20) wurde durch die Scopus-Publikationen in WOS ermittelt, die höchste Zitationsrate von 23 wurde von jeder Auswahl in mindestens einem Datenbestand ermittelt.

verwendet den aktuellen Titel für eine Publikation der Vorgängerzeitschrift. Somit konnten in allen Datenbanken kleine Fehler aufgezeigt werden, die aber wohl nicht systematisch vorkommen.

Publisher

Für knapp die Hälfte der Publikationen (136) fehlen Angaben zu den Verlagen in den Scopus-Daten des KB, während alle WOS-Daten der Untersuchungsmenge diesbezüglich vollständig sind. In den entsprechenden Daten von Dimensions fehlen viermal Informationen zu den Verlagen. Dadurch können nur für 141 Publikationen über alle drei Datenbanken Vergleiche zu den Publisher-Einträgen durchgeführt werden.

Publisherwechsel stellen vereinzelt ein Problem dar, das zu unterschiedlichen Angaben der Datenbanken bezüglich einzelner Publikationen führt. Eine Publikation des Jahres 2003 aus dem *Journal of NeuroVirology*²² wird in Dimensions dem Verlag Springer zugewiesen, obwohl dieser die Zeitschrift erst 2011 übernommen hat.

Die Zeitschrift *Emerging Microbes and Infections* wechselte 2019 von Springer zu Taylor & Francis,²³ Dimensions zählt aber drei Publikationen aus den Jahren 2018 und eine aus dem Jahr 2013 schon zum neuen Verlag. Bei einer Überprüfung hat sich herausgestellt, dass die Zeitschrift in Dimensions insgesamt nur dem letzten Verlag zugeordnet ist.

WOS benennt für die ältere und eine der Publikationen aus 2018 die Nature Publishing Group als herausgebenden Verlag (die laut Wikipedia seit 2015 zu Springer Nature gehört) und für die beiden anderen Tylor & Francis. Alle drei Publikationen des Jahres 2018 sind übrigens aus demselben Zeitschriftenband²⁴ und damit die Benennung unterschiedlicher Verlage unerwartet. Scopus weist diesen drei Publikationen die Nature Publishing Group zu, die ältere hat keine Angaben zum Publisher.

Die Vielfältigkeit von Verlagsbenennungen und auch strukturelle Zugehörigkeiten und Veränderungen führen zu heterogenen Ergebnissen und zeigen, dass eine Normierung dieser Daten nicht oder kaum vorgenommen wurde bzw. im Falle von Dimensions durchaus zu hinterfragen ist. Die Schreibweisenvereinheitlichung bei Dimensions führt immerhin zu einer übersichtlicheren Ergebnismenge von 23 Verlagsnamen, während es bei Scopus 39 und bei Web of Science 44 sind. Den 49 Publikationen z.B., die in Dimensions unter dem Verlag Elsevier geführt werden, stehen bei Web of Science sechs Varianten des Namens Elsevier gegenüber und außerdem zwei weitere Verlage, die aber auch der Elsevier-Verlagsgruppe angehören. Ähnlich gestalten sich die Angaben bei der Datenbank Scopus, die außerdem für eine Publikation der Zeitschrift *Clinical Microbiology and Infection* aus dem Jahr 2009 den Verlag Blackwell Publishing benennt. Die Zeitschrift ist erst 2015 zu Elsevier übergegangen. Blackwell Publishing ist allerdings auch bereits seit 2007 Teil von John Wiley & Sons.

22 DOI: 10.1080/13550280390173319

23 Angaben laut EZB: [http://ezb.uni-](http://ezb.uni-regensburg.de/ezeit/detail.phtml?bibid=FZJUE&colors=7&lang=de&jour_id=187874)

[regensburg.de/ezeit/detail.phtml?bibid=FZJUE&colors=7&lang=de&jour_id=187874](http://ezb.uni-regensburg.de/ezeit/detail.phtml?bibid=FZJUE&colors=7&lang=de&jour_id=187874), 21.07.2020

24 DOI: 10.1038/s41426-018-0121-2, 10.1038/s41426-018-0193-z, 10.1038/s41426-018-0199-6

Sprachen

Als Sprache der Publikationen ist in allen drei Datenbanken entweder komplett Englisch erfasst (WOS) oder es fehlen die Sprachangaben (Dimensions: 3, Scopus: 1).

Kategorien

Nur eine der 277 ausgewählten Publikationen hat in Dimensions keine Kategorie erhalten: es handelt sich um einen Article, der in WOS und Scopus jedoch den Dokumenttyp correction bzw. error erhalten hat.²⁵ Ebenso wie der Originalartikel²⁶ ist diese Korrektur in Nature erschienen und hat damit in WOS und Scopus keine spezifische Kategorie zugewiesen bekommen (Multidisciplinary). Dimensions ordnet den Originalartikel unter die Division History and Archaeology ein. Der Titel "Ancient hepatitis B viruses from the Bronze Age to the Medieval period" weist tatsächlich auf eine historische Komponente hin.

Web of Science und Scopus haben alle Beiträge der Vergleichsauswahl ihren Kategoriensystemen zugeordnet.

Virology und Infectious Diseases sind in WOS und Scopus jeweils die Kategorien mit den häufigsten Publikationen. In Dimensions existieren diese Kategorien nicht, auch nicht auf der 2. Ebene. Dort sind beispielsweise die 178 Publikationen, die in WOS zu diesen Kategorien gehören, größtenteils im Fachbereich Medical Microbiology (125 Publikationen) eingeordnet.

Tabelle 18: Publikationen von C.Drosten nach First-Level-Kategorien in DIM20

Division	P
Medical and Health Sciences	224
Biological Sciences	82
Agricultural and Veterinary Sciences	7
Environmental Sciences	2
History and Archaeology	2
Physical Sciences	2
Technology	2

Tabelle 19: Publikationen von C. Drosten nach Second-Level-Kategorien in DIM20

Group	P
Medical Microbiology	175
Clinical Sciences	42
Genetics	30
Public Health and Health Services	27
Biochemistry and Cell Biology	20
Microbiology	20
Immunology	10

25 DOI: 10.1038/s41586-018-0406-6

26 DOI: 10.1038/s41586-018-0097-z

Tabelle 20: Publikationen von C. Drosten nach Subject Category in WOS

Subject Category	P
Virology	94
Infectious Diseases	85
Immunology	67
Microbiology	64
Multidisciplinary Sciences	29
Biotechnology & Applied Microbiology	17
Parasitology	10

Tabelle 21: Publikationen von C. Drosten nach Fields in Scopus

Field	P
Infectious Diseases	122
Virology	105
Microbiology (medical)	85
Microbiology	53
Epidemiology	50
Immunology	46
Insect Science	25
General Biochemistry, Genetics and Molecular Biology	17
Immunology and Allergy	16
Parasitology	15
Multidisciplinary	14
Molecular Biology	14
General Agricultural and Biological Sciences	13
Public Health, Environmental and Occupational Health	10

Wichtigstes Ergebnis für die Untersuchung der Kategoriensysteme ist jedoch die hohe Vollständigkeit der Daten. Und obwohl es zwischen Dimensions einerseits und WOS und Scopus auf der anderen Seite Unterschiede in den Kategoriensystemen gibt, so weist das Ergebnis für den engen Fachbereich der vorliegenden Untersuchungsmenge dennoch auf eine nachvollziehbare Kategorisierung bei Dimensions hin.

Anzahl Autoren

Bei den Autorenanzahlen je Publikation gibt es eine hohe Übereinstimmung im Vergleich der Datenbanken. Gezählt wurde jeder Nachname einmal, evtl. mit einer geringen Zählungenauigkeit bei Namensgleichheiten. Auffällig war lediglich die Zählung bei einer Publikation, für die Dimensions die Autoren eines beteiligten Konsortiums einzeln mitgezählt hat, während dies bei Web of Science und Scopus nicht der Fall ist. Dimensions hat dadurch 30 Autoren mehr gezählt als die beiden anderen Datenbanken.

Corresponding Authors

Bei der Kenntlichmachung der Corresponding oder Reprint Authors gibt es sehr unterschiedliche Ergebnisse. Bei den WOS-Publikationen gibt es nur zwei ohne RP-Angaben, bei Scopus fehlen 13 und Dimensions verzeichnet für 195 und damit 70 % der Publikationen keinen Corresponding Author.

Anzahl Adressen

Die Anzahl der Publikationen mit Autorenadressen ist für die Publikationen aus WOS und Scopus gleich hoch (275). Dimensions hat aber 20 Publikationen weniger mit Adressangaben. Auch in der durchschnittlichen Adressenanzahl liegt Dimensions unter WOS und Scopus, auch wenn die Zahlen der beiden letztgenannten Datenquellen teilweise zu hoch ausfallen: gezählt wurden die vollen Adressangaben (ohne RP-Adressen), deren Varianten innerhalb einer Publikation zu einer höheren Zählung führen können. Im Schnitt werden zu den Dimensions-Publikationen 4,2 Adressen angegeben, bei den Scopus-Publikationen 6,7 und bei den WOS-Publikationen 6,8.

Für Auswertungen von Autoren- und Adressdaten wird diesen Ergebnissen zufolge relevant sein, ob für Dimensions ausschließlich die GRID IDs genutzt werden, oder die vollen Adressangaben zur weiteren Bearbeitung zur Verfügung stehen. Der Anteil Publikationen ohne jede Adressangabe liegt in diesem Datensatz bei 7 % und würde damit ebenfalls Einfluss auf das Ergebnis haben. Informationen zu den Corresponding Authors sind für Dimensions zu lückenhaft und damit ungeeignet für Auswertungen, während der Erfassung von Autoren in den drei Datenbanken generell unterschiedliche Kriterien zugrunde zu liegen scheinen. Eine genauere Untersuchung der Daten wäre für diesen Bereich sinnvoll.

Länder

Die größere Anzahl fehlender Adressen bei Dimensions wirkt sich auch auf die Länderauswertung aus. Zusätzlich stellt das Fehlen von GRID IDs, auf deren Daten die Auswertung beruht, ein Problem dar.

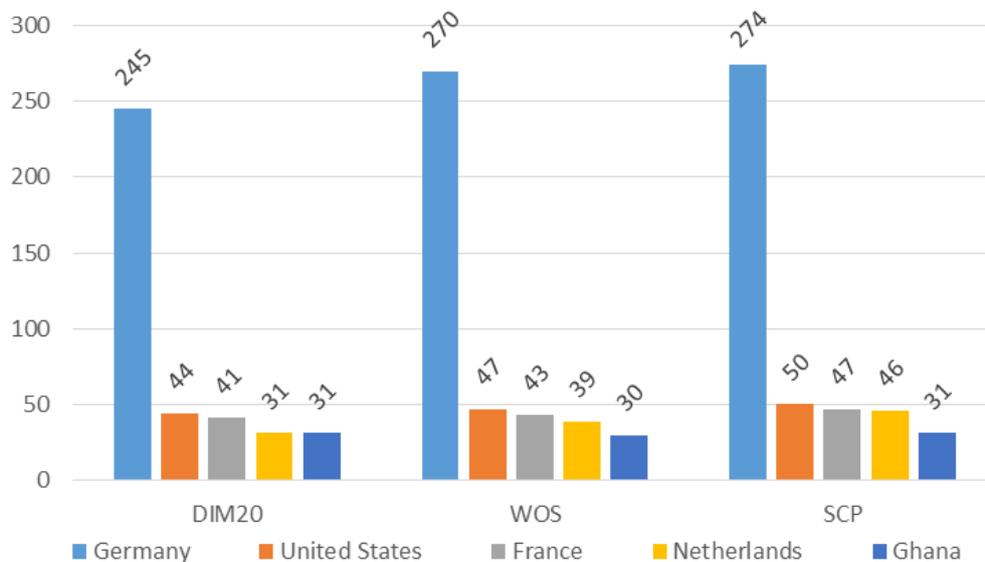


Abbildung 15: Die 5 Länder mit den höchsten Anteilen in der Analyse von C. Drostens in allen 3 Datenbanken im Vergleich

Die Reihenfolge der Länder nach Anzahl Beteiligungen an den Publikationen ist zwar auch über die Top 5 hinaus gleich, aber die absoluten Zahlen aller drei Datenbanken weichen teilweise deutlich voneinander ab. Gezählt wurde jedes Land je Publikation nur einmal.

Funding Information

Für 178 (64 %) Publikationen in Dimensions und 179 (65 %) in WOS ist mindestens eine Förderorganisation verzeichnet. Die Überschneidungsmenge dieser Ergebnisse umfasst 112 Publikationen und damit rund 40 %. Scopus weist lediglich zu 33 (12 %) der Publikationen Förderorganisationen nach, 29 davon befinden sich auch in der Überschneidungsmenge von Dimensions und WOS.

Stichproben der exklusiven Publikationen aus Dimensions, also solcher mit Funding Information, die bei WOS und Scopus gänzlich fehlen, zeigen, dass die vorhandenen Informationen auch nicht immer vollständig übernommen wurden. In einem Beispiel wurden vier von fünf in Acknowledgements oder unter Funding aufgelistete Förderorganisationen in die Metadaten übernommen, in einem anderen nur eine von fünf.²⁷

Entsprechend der Vorgehensweise beim institutionellen Datenvergleich wurden in den jeweiligen Publikationsdaten diejenigen ermittelt, die durch das BMBF gefördert worden waren. Bei Dimensions waren 50 Publikationen durch die eindeutige GRID ID gut zu ermitteln. In den WOS-Daten konnten nach Zusammenführung der Schreibvarianten insgesamt 52 Publikationen mit einer BMBF-Förderung identifiziert werden. Die Schnittmenge beträgt 43 Publikationen. Die Scopus-Daten wurden in diesem Vergleich aufgrund der geringen Informationen nicht berücksichtigt. Mehrere Publikationen von Dimensions und WOS, die außerhalb der Schnittmenge liegen, wurden überprüft: die Benennung des BMBF war jeweils korrekt und damit die Nichterfassung in der jeweils anderen Datenbank fehlerhaft.

Der Anteil der Publikationen mit Informationen zu einer Förderung fällt in dieser Untersuchung in allen drei Datenquellen geringer aus, als in der Untersuchung der FZJ-Publikationen. Während sich die Scopus-Daten aufgrund der mangelnden Informationen gar nicht zu einer Analyse der Funding Information eignen, weisen die Stichprobenergebnisse von Dimensions und WOS trotz einiger Differenzen auch Ähnlichkeiten auf.

5. Zusammenfassung der Ergebnisse und Fazit

Die Studie hat die Daten der Datenbank Dimensions quantitativ und qualitativ hinsichtlich ihrer Verwendbarkeit in bibliometrischen Auswertungen untersucht. Dabei stand der Vergleich mit den etablierten Datenbanken Web of Science und Scopus im Vordergrund. Diese Studie ist nur ein Teil der Evaluierung und ist mit den weiteren Teiluntersuchungen, die vom DZHW und von der Uni Bielefeld durchgeführt wurden, im Gesamtkontext zu betrachten.

Im Hinblick auf den Gesamtbestand wurden für Dimensions zum Teil mehrere Datenstände betrachtet (April 2019 und März 2020) um auch die Veränderungen, die diese schnelllebige Datenbank macht, aufzeigen zu können. Während in der Datenbank im April 2019 nur gut die Hälfte aller Publikationen einem Fachgebiet zugeordnet war (55 %), waren es ein Jahr später bereits 72 % aller Publikationen. Die Klassifikationsysteme der Datenbanken sind grundsätzlich komplett unterschiedlich, was einen Vergleich auf dieser Ebene erschwert. Der Ansatz der Klassifizierung variiert ebenfalls: während in WOS und Scopus eine Fachkategorie auf Basis der entsprechenden Zeitschrift vergeben wird, klassifiziert Dimensions seine Publikationen basierend auf dem jeweiligen Artikel selbst. Interessante Erkenntnis war hier die hohe Abdeckung von Publikationen mit Kategorien in den Case Studies im Unterschied zum Gesamtbestand.

27 DOI: 10.1016/s1473-3099(15)70090-3 und 10.1016/s1473-3099(15)70090-3

Eine Länderangabe ist auch in den aktuellen Daten im Jahr 2020 nur in jeder zweiten Publikation vorhanden (48 %), wird aber wohl zum großen Teil der Tatsache geschuldet sein, dass nur Adressen mit einer GRID ID ausgewertet wurden. Ergebnisse im Online-Zugriff weisen auch Adressen auf, die keiner GRID ID zugeordnet sind, aber sowohl Länder- als auch institutionelle Informationen enthalten. Bei beiden untersuchten Datenständen lag der Anteil von Dokumenten ohne einen Quellentitel bei gut 20 %. Einen überwältigenden Unterschied zeigt die Information zur Sprache der Publikationen: während im Jahr 2019 nur 0,25 % Publikationen Sprachangaben in Dimensions besaßen, war der Anteil ein Jahr auf 99,3 % angestiegen.

Im Vergleich der Datenbanken Dimensions, WOS und Scopus existieren große Unterschiede bei der Auswahl der Quellen bzw. den unterschiedlichen Definitionen dieses Begriffs: während es bei WOS und Scopus ausschließlich Zeitschriftentitel, Bücher oder Konferenzberichte sind, ist es bei Dimensions eine Mischung aus Zeitschriften, Preprint-Servern, Abstractservices oder ähnlichen Quellen. Daher ist es auch wenig verwunderlich, dass auch unter den unterschiedlichen Datenständen von Dimensions dynamische Veränderungen zu betrachten sind: im März 2020 ist der Preprint-Server arXiv die häufigste Quelle, während dieser ein Jahr zuvor noch überhaupt nicht zum Bestand der Datenbank gehörte. Dieser Ansatz erschwert eine bibliometrische Analyse auf Zeitschriftenebene, da diese in Dimensions nicht explizit als solche herauszufiltern und in einer großen Sammlung an verschiedenen Quellenarten untergemischt sind. Und es erklärt den deutlich größeren Bestand von Dimensions (für die Publikationsjahre 1996-2018 und Datenstand April 2019 waren es fast 63 Millionen) im Vergleich zu Scopus (gut 48 Millionen) und WoS (gut 43 Millionen).

Die Bezeichnungen und vor allem auch Anwendung der Dokumenttypen ist ebenfalls von großen Unterschieden geprägt: Dimensions unterteilt die verzeichneten Publikationen in lediglich sechs Dokumenttypen, während in Scopus 16 und in WOS sogar 41 Arten von Publikationen indexiert sind. Eine Stichprobe zeigt auf, dass in Dimensions viele Dokumente des Typs Article im WOS strenggenommen nicht als wissenschaftliche Publikation gelten (z.B. Book Review). Für die Verwendung dieser Informationen im Rahmen bibliometrischer Analysen, die häufig eine feinere Einschränkung auf bestimmte Dokumenttypen erfordern, ist das eine Schwachstelle von Dimensions.

Dimensions-Publikationen haben nahezu ausnahmslos einen Digital Object Identifier (DOI). Eine Stichprobe für das Jahr 2018 ergab, dass nur 0,6 % der Einträge keinen DOI besitzen, während es im selben Jahr bei Scopus rund 10 % und bei WoS sogar 27 % der Publikationen ohne DOI-Eintrag sind. Für vergleichende Analysen, die alle drei Datenbanken inkludieren, ist Dimensions daher mit dem universellen Identifier DOI am besten geeignet. Alle Einträge in WOS und Scopus besitzen zwar eine eigene ID (UT-Code bzw. Scopus ID), diese sind jedoch nur innerhalb der jeweiligen Datenbank nutzbar und können nicht zwischen den verschiedenen Datenbanken verwendet werden.

Zur detaillierteren Betrachtung inhaltlicher Unterschiede wurden nach der Analyse der jeweiligen Gesamtbestände (1996-2018) Case studies auf Einrichtungs- sowie auf Autorenebene durchgeführt. Die Untersuchungsmengen bestanden aus Publikationen des Forschungszentrums Jülich und des Fraunhofer-Instituts für Keramische Technologien und Systeme aus dem Zeitraum 2009-2018 sowie des Wissenschaftlers Christian Drost (ohne zeitliche Einschränkung). Insgesamt kann eine gute Datenlage für Dimensions bescheinigt werden. Für die konkreten Anwendungsfälle liegen im Vergleich zum Gesamtbestand in vielen Bereichen vollständigere Informationen vor. Problematisch bleibt die Dokumenttypenvergabe bei Dimensions, auch wenn im Rahmen der Studien die Überschneidung mit den Dokumenttypen von WOS und Scopus höher war. Bei den Autoredaten fällt die nur geringfügig vorgenommene Kennzeichnung der Corresponding Authors der Dimensions-Daten auf. Diese Information ist in manchen Analysen erforderlich. Die Vollständigkeit der Adressangaben in Dimensions kann abschließend erst dann bewertet werden, wenn auch die Adressen ohne GRID ID

einbezogen sind. Die bisherigen Ergebnisse deuten aber auf eine ähnlich gute Datenlage wie bei WOS hin und es gibt einen wesentlich höheren Anteil von Publikationen mit Adresszuordnungen in den Case Studies als im Gesamtbestand seit 1996. In einigen Fällen fehlen Autoren oder Adressen bzw. sind weitere, als die von den Verlagen benannten, Namen und Adressen aus den Texten entnommen und ergänzt worden.

Auch bezüglich der Kategorien ist die Abdeckung für die Publikationen der Case Studies deutlich verbessert gegenüber der Gesamtbestandbetrachtung. Wie zuvor bereits aufgezeigt, sind in Dimensions einige Publikationsarten indexiert, für die aus verschiedenen Gründen eine automatisierte Kategorisierung nicht oder kaum möglich ist, die aber in den Case Studies aufgrund der speziellen Auswahl nicht mit einbezogen sind und zum Unterschied der Ergebnisse beitragen dürfte.

Die Informationen zur Förderprogrammen und fördernden Institutionen können bei Bedarf für bibliometrische Analysen ebenfalls relevant sein. Dimensions weist sie ausdrücklich aus und bietet auch online in den 'Analytical Views' die 'Funders' auf Basis ihrer GRID IDs zur Auswertung an. Im Vergleich zu WOS fällt der Anteil der Publikationen mit Förderinformationen etwas geringer aus. Im direkten Vergleich hat sich gezeigt, dass vorhandene Angaben in beiden Datenquellen auch unvollständig sind und die Definition von Förderinstitutionen nicht einheitlich gehandhabt wird. Da die Ergebnisse im Vergleich noch einige Unterschiede aufweisen, erscheint es notwendig, weitere Untersuchungen zu diesem Bereich vorzunehmen, um sowohl die Vollständigkeit als auch die Qualität der Angaben besser beurteilen zu können.

Nach den bisherigen Ergebnissen bietet Dimensions eine gute Basis für bibliometrische Analysen. Einige Bereiche, wie z.B. die Publikations- oder Dokumenttypeneinteilung und die Sacherschließung, weisen noch zu wenig Differenzierung bzw. Lücken auf, die aber möglicherweise in der Weiterentwicklung der relativ neuen Datenbank verbessert werden. Die für die Institutionen- oder Personenanalysen wichtigen Autorennamen und Adressen scheinen gut erfasst zu sein. Wie groß das Problem fehlender Einträge, auch im Vergleich zu den ebenfalls nicht fehlerfreien Daten in WOS und Scopus, ist, müssen weitere Analysen im Vergleich der Datenbanken zeigen.

6. Literaturverzeichnis

Bode, C., Herzog, C., Hook, D., & McGrath, R. (2018). *A Guide to the Dimensions Data Approach. A collaborative approach to creating a modern infrastructure for data describing research: where we are and where we want to take it*. Retrieved from <https://www.digital-science.com/resources/portfolio-reports/a-guide-to-the-dimensions-data-approach/>: <https://www.digital-science.com/resources/portfolio-reports/a-guide-to-the-dimensions-data-approach/>

Bornmann, L. (2018). Field classification of publications in Dimensions: a first case study testing its reliability and validity. *Scientometrics*, 117(1), 637-640. doi:10.1007/s11192-018-2855-y

Hook, D. W., Porter, S. J., & Herzog, C. (2018). Dimensions: Building Context for Search and Evaluation. *Frontiers in Research Metrics and Analytics*, 3(23). doi:10.3389/frma.2018.00023

Orduna-Malea, E., & Delgado Lopez-Cozar, E. (2018). Dimensions: re-discovering the ecosystem of scientific information. *El Profesional de la Información*, 27(2), 11. doi:10.3145/epi.2018.mar.21

Scopus Content Coverage Guide. (2020). Retrieved from elsevier.com/solutions/scopus/how-scopus-works/content?dgcid=RN_AGCM_Sourced_300005030

- Subject Category. (2020, 2020/06/01). *Web of Science Core Collection Help*. Retrieved from https://images.webofknowledge.com//WOKRS534DR2/help/WOS/hp_subject_category_terms_tasca.html
- Visser, M., van Eck, N. J., & Waltman, L. (2020). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic.
- Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, 10(2), 347-364. doi:10.1016/j.joi.2016.02.003
- Web of Science: List of Subject Classifications for All Databases*. (2014). Retrieved from https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-List-of-Subject-Classifications-for-All-Databases?language=en_US: https://support.clarivate.com/ScientificandAcademicResearch/s/article/Web-of-Science-List-of-Subject-Classifications-for-All-Databases?language=en_US
- What is the complete list of Scopus Subject Areas and All Science Journal Classification Codes (ASJC)? (2020, 2020/05/27). *Scopus: Access and use Support Center* Retrieved from https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/session/L2F2LzEvdGltZS8xNjE2NzYwNjcxL2dlbi8xNjE2NzYwNjcxL3NpZC9mVTFFFTnpaaWhFdTFlaThnX3RMT3hsN0h0aFBCb1dzJTdFTnd0NnZybvdkawhHMDIKJTdFVXFsdW5FRGVVUF9KQldaUDNqdUJ3dzU0SENBYjJJUFZFTV9oNGMzJTdFZUpKd0ZGY0k2Uk9TY3NNOHk0cTZLZktaMmVZQkVBSGclMjEIMjE%3D/
- Which research categories and classification schemes are available in Dimensions? (2020). Retrieved from <https://dimensions.freshdesk.com/support/solutions/articles/23000018820-what-are-fields-of-research-and-other-classification-systems-and-how-are-they-created->
- Williams, C. (2018). Dimensions from Digital Science. *Insights*, 31(33). doi:doi.org/10.1629/uksg.420