competence centre for bibliometrics

German Centre for Higher Education Research and Science Studies

Performance and Structures of the German Science System 2020

Dimity Stephen, Stephan Stahlschmidt and Sybille Hinze

Studies on the German Innovation System
No. 5-2020

German Centre for Higher Education Research and Science Studies (DZHW)

May 2020 (update)

Diese Studie wurde im Auftrag der Expertenkommission Forschung und Innovation (EFI) erstellt. Die Ergebnisse und Interpretationen liegen in der alleinigen Verantwortung der durchführenden Institute. Die EFI hat auf die Abfassung des Berichts keinen Einfluss genommen.

Studien zum deutschen Innovationssystem
Nr. 5-2020
ISSN 1613-4338

Herausgeber:
Expertenkommission Forschung und Innovation (EFI)

Geschäftsstelle:
c/o Stifterverband für die Deutsche Wissenschaft
Pariser Platz 6
10117 Berlin

Kontakt und weitere Informationen:
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)
Dr. Stephan Stahlschmidt
Schützenstr. 6a
10117 Berlin

Tel: +49 (0)30 2064177-18
Fax: +49 (0)30 2064177-99
E-Mail: stahlschmidt@dzhw.eu

Contents

Executive Summary

This study updates the annual analysis of the performance and structures of the German science system in international comparison. Bibliometric indicators are presented and discussed for the period 1995-2018, and citation-based indicators are presented for publications until 2016. In addition, contextual information is provided in order to better grasp the influence of the underlying data in relation to actual changes in the national performance. Thus, differences regarding absolute publication numbers between the two source databases are due to different coverage and also lead to differences in the actual indicator values, yet still lead to similar results when it comes to relative performance. The analysis also shows that the quality of the data covered in both databases has improved over time, meaning that the share of relevant data elements missing, e.g. information about the discipline a journal is assigned to, author affiliation, document type information, and information on the publication year, has decreased. Also, the degree of concordance regarding data elements between the two databases has improved. However, differences remain with regard to database coverage which is to be expected due to the different strategies pursued by both database producers as discussed in the previous edition of this study [7].

Overall, the trend of an increasing number of publications continues for most countries. This increase is due to increased publication output, but also reflects the increased coverage of journals by Web of Science (WoS) and Scopus. The growth is particularly high for China and India. While for countries with well-established systems such as the USA, the UK, Germany and France, growth rates remain rather low and thus result in decreased publication shares. China not only continues to grow in terms of total number of publications, but continues to improve its international position as regards the share of scientific output it contributes. Now, based on WoS data, China takes the leading position. Thus, starting from being positioned amongst the lowest ranking countries with regard to the scientific output in 1995, China now contributes approximately 20% of global publications. At the same time, the USA, which contributed nearly a third of global publications in 1995, is reduced to 18.6% (WoS) by 2018. Germany's share of publications in the WoS over the three decades also decreased from 6.3% in 1995 to 4.3% in 2018, which continues to represent the fourth-largest share of global publications after China, the USA, and Great Britain.

As regards visibility and impact of the scientific publications, Germany continues to publish in highly visible journals, which is shown by the above average values for the International Alignment (IA) and, within these journals, German publications attract citations which are slightly above the expected values. However, due to the higher expectations set by the highly visible journals, the indicator values for the Scientific Regard (SR) are slightly decreasing. While the USA continues to publish in very highly-cited journals, its values for both the IA and SR indicators, are decreasing. At the same time China also improves the visibility and the impact of its scientific publications. The value for SR, based on WoS data, slightly exceeds the expected average value, while the value for IA is approaching the expected value. This trend is not paralleled by other BRICS countries, which, while improving their SR indicator values and thus reaching close to expected impact in the journals they selected to publish, still continue to publish in less visible journals.

The picture is similar when looking at the Excellence Rate (ER) indicator. Also with regard to these highly valued publications, China continues to improve its performance, almost reaching the expected value. The other BRICS countries do not show similar performance increases. ERs for the USA also continue to decrease. Germany's position is rather stable above the threshold, however there has been a slight downward trend since 2010.

In general, the patterns observed in high- and low-performing countries were largely the same across

the databases, despite the differences in figures, e.g. differences in ER indicator values between WoS and Scopus can be explained by coverage policies, especially Scopus' greater coverage of journals with a regional focus. These journals are relevant to a narrower audience, thus potentially less cited, which reduces the ERs of some of the BRICS countries in Scopus. An opposite effect can be observed for many OECD countries.

Foreword

This report provides the latest iteration of a series of bibliometric indicators used to analyse the performance of Germany against 22 selected countries and benchmarks it within the international science system. In the first section, we present four indicators over a time-series beginning in 1995: the annual fractional share of global publications to gauge each country's level of scientific contribution; the complementary Scientific Regard and International Alignment indicators that assess the impact of the scientific contribution based on the citedness of a country's publications in relation to the journals in which they are published, and the citedness of the journals themselves; and Excellence Rates, which are the percentage of a country's publications that were in the 10% most highly cited publications per discipline. We provide these indicators based on data extracted from both the Web of Science (WoS) and Scopus databases to enable a macro-level comparison of the indicators calculated from two key bibliometric data sources.

In analysing a more than 20-year time-series, not only are changes in the national science systems relevant, but so too is the constantly-changing nature of the WoS and Scopus databases, which are regularly updated to expand their coverage of journals, both current and historical. As such, in the second section of this report, we provide a suite of contextual information relevant for the interpretation of the time-series for indicators presented in Section 1. This includes analyses of the coverage of the databases, missing metadata, agreement on metadata between databases, and the composition of the database in terms of document type, discipline and the languages in which publications are written. These data highlight apparent changes in the underlying databases, allowing a descriptive distinction to be drawn between changes in the national science systems and changes in the databases.

In this report, we examine articles and reviews published in journals from the entire Scopus database and from the WoS indices Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Arts and Humanities Citation Index (A&HCI). We present data from 1995 to 2018 for WoS and from 1996 to 2018 for Scopus, as Scopus data are only available to the German Competence Centre for Bibliometrics from 1996 onwards. We use a three-year citation window for citation-based indicators, and thus include citations occurring in the year of publication or in the following two years. Consequently, citation-based indicators are provided for publications up until 2016. We have not excluded self-citations.

We have primarily used fractional counting for all indicators and data that are disaggregated by country. We applied fractional counting at the level of the organisation for WoS data from 1995 to 2008 due to data quality issues, and at the level of the author in WoS from 2009 onwards and throughout the time-series for Scopus data. Similarly we have applied fractional counting to aggregate publications from their native disciplines in Scopus and WoS to the OECD's Fields of Science and Technology (FOS) classification, the classification we have used to present disciplinary data. Further details about the methodology used in this report are available in Appendix 2.

Section 1: Indicators

This Section presents four indicators – national fractional shares of global publications, Excellence Rates, Scientific Regard, and International Alignment for 23 countries over a time-series. These indicators are provided with the aim of describing the performance of Germany's science system and evaluating it against the 22 countries selected for comparison (see Appendix 1).

National shares of global publications

National shares of global publications are a useful indicator of a country's international standing regarding scientific output. We present in Figures 1 and 2 the fractionalised percentage of global publications held by each of the 23 countries in WoS and Scopus. In these figures, the width of each country's band represents its share of global publications. Further, the ordering of the bands within each group indicates the country's ranking from the largest share at the top to the smallest share at the bottom. Please note the different scales of the panels in each figure.

One key trend is the increase in China's share over time from 1.6% of global publications in 1995 to 20.4% in 2018, taking it from 12th place to 1st for the first time in 2018. Most countries' shares have correspondingly decreased. The USA, for instance, produced nearly a third of global publications in 1995 (32.5%), however this was reduced to 18.6% by 2018. Germany's share decreased by 2 percentage points from 6.3% in 1995 to 4.3% in 2018, which now constitutes the fourth-largest share of global publications after China, the USA, and Great Britain. France, Israel and Sweden's shares also declined over time, with France losing nearly half of the 4.9% it held in 1995 by 2018. Conversely, Poland's rank rose rapidly from 1995, overtaking Switzerland, Sweden, Belgium, and Israel to hold a similar share of publications as the Netherlands in 2018.

We see a similar pattern in Scopus: China's share of global publications increased from 2.6% in 1996 to nearly 20% in 2018, with a noteworthy jump from 8% to 11% in 2005 alone, pushing it from 9th place to 1st by 2016. The USA's share declined accordingly from 26.9% to 16.8% in 2018, placing it 2nd behind China from 2016. India also increased its share in both databases, but to a larger extent in Scopus: India held around 2% of global publications in 1996 in both databases, but had grown to only 3.6% by 2018 in WoS compared to 4.6% in Scopus, giving it the third-largest share in Scopus. Scopus has better coverage of most BRICS countries, with these being the only countries with a higher share in Scopus than WoS in 2018, which aligns with the providers' business models that WoS indexes the most relevant journals, while Scopus covers a larger number of regional journals [7]. Great Britain and Germany hold the fourth and fifth largest shares in Scopus, respectively, both having lost approximately 2 percentage points between 1996 and 2018. After the USA, Japan had the greatest decrease in its share, losing 4 percentage points in both databases.

Also, Figure 10 shows that a large number of journals were indexed in WoS in 2007-2008, which may have been a reaction to Scopus' introduction in 2004, challenging WoS' monopoly to that point. At this time, China's share of publications in WoS increased from 7.3% in 2006 to 8.4% in 2008, as did shares for Brazil (1.6% to 2.2%), India (2.5% to 3.0%), Korea (2.2% to 2.6%), and Spain (2.5% to 2.7%), suggesting the journals may have been added to improve coverage for these countries.

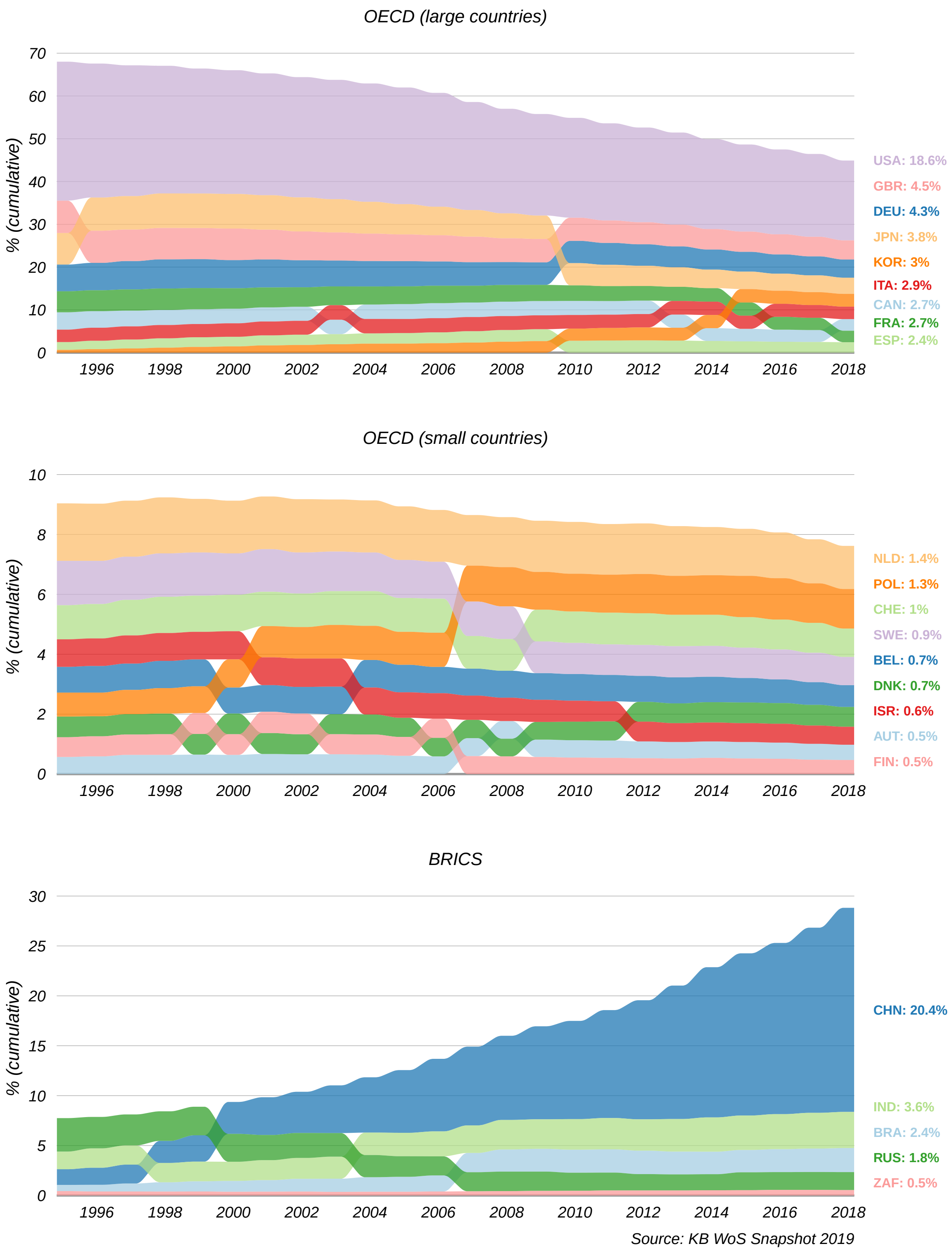

Figure 1: Countries' fractional shares of world publications in WoS over time. Shares are presented cumulatively. The width of each country's band represents its fractional share of publications, and the ordering of the bands shows the country's ranking within its group with larger shares at the top.

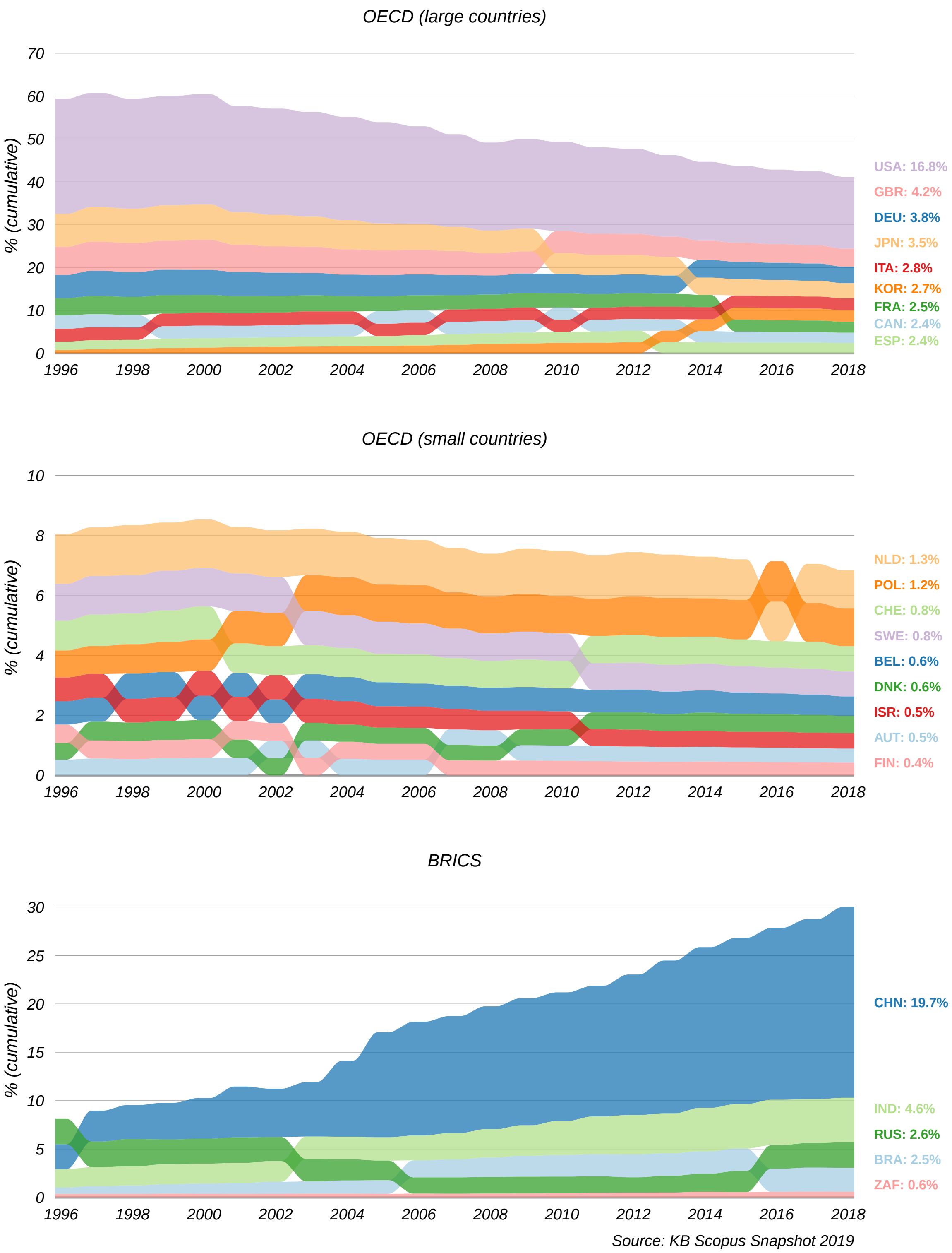

Figure 2: Countries' fractional shares of world publications in Scopus over time. Shares are presented cumulatively. The width of each country's band represents its fractional share of publications, and the ordering of the bands shows the country's ranking within its group with larger shares at the top.

Excellence Rates

Excellence Rates (ERs) are the percentage of a country's publications that were in the 10% most highly cited publications from each discipline and are thus considered highly impactful. In general, a country could be expected to achieve a 10% ER. ERs above 10% indicate the country had a greater than expected percentage of publications in the 10% most highly cited publications and therefore performed better than expected, while ERs below 10% indicate poorer than expected performance. The ERs for the 23 selected countries are presented in Figure 3 for WoS and Figure 4 for Scopus. ERs are calculated up to the publication year 2016 to allow for a 3-year citation window, and are based on the native classifications of disciplines within each database.

The majority of OECD countries performed at or above the expected 10% level in WoS, with only Korea, Poland, Japan, Spain, Israel, and France by a small margin, falling below 10%. The highest ERs were consistently held throughout the time-series by Switzerland (14.9% in 2016), Great Britain (14.0%), the Netherlands (13.8%), the USA (13.6%), and Denmark (13.5%). Despite its high ER, the USA experienced a slow decline over time from approximately 15% in the late 1990s to its current 13%, which, along with Japan, was amongst the largest decreases for the selected countries. Only a small number of countries demonstrated continuous improvements in ERs, including China which rapidly improved its ERs from approximately 4% in the late 1990s to 9.8% in 2016, and India which also gained approximately 3 percentage points in ER from 2.3% to 5.5% over the same period. ERs for Great Britain and Italy also improved, but with more abrupt increases in 2008 and 2012 respectively, which coincided with important changes in the evaluation systems in each country.

Germany's ER was relatively stable, fluctuating between 10.6% and 11.7% over time, placing it in a similar ranking as countries such as Sweden, Belgium, Austria, and Italy. However, Germany's ER has shown a slight downward trend since 2010 from 11.6% in 2010 to 11.1% in 2016. This decrease in ERs over time for countries such as the USA and Germany may reflect the influence of China's citation practices on the international science system, as demonstrated in Stahlschmidt and Hinze [6] and Stahlschmidt, Stephen and Hinze [7]

The trends in countries' ERs over time are similar between WoS and Scopus, although ERs tend to be slightly higher in Scopus for most of the OECD countries and slightly lower in Scopus for most of the BRICS countries. For instance, Germany's ER ranged between 11.3% and 12.3% over the time-series in Scopus, compared to 10.6% and 11.7% in WoS, however the trend was similar in both databases – a peak in 2010 with subsequent slight decline to 2016. Conversely, China's ER in Scopus was consistently 1.4 to 3.9 percentage points below its ER in WoS, although it maintained the trend of rapid improvement in Scopus from 1.7% in 1996 to 8.4% in 2016. As in WoS, in Scopus the highest ERs continued to be held by the Netherlands (15.5% in 2016), Switzerland (15.4%), Denmark (15.1%), Great Britain (14.6%), and the USA (14.0%), although the USA also declined in Scopus over time from 15.6% in 1996.

The disparity seen between WoS and Scopus with regard to OECD and BRICS countries likely reflects Scopus' greater coverage of journals with a regional focus, as demonstrated by the higher number of publications in languages other than English in Scopus compared to WoS seen in Figure 17. These journals are less likely to be cited as they are relevant to a more narrow audience and this has the effect of reducing the ERs of the BRICS countries in Scopus. Conversely, WoS, by more strictly selecting only highly-cited journals, captures only the BRICS countries' publications that are in well-cited journals, which generates higher ERs. The effect is reversed for most of the OECD countries, as Scopus' larger collection of journals captures more of the well-cited publications from these countries than does WoS' narrower scope.

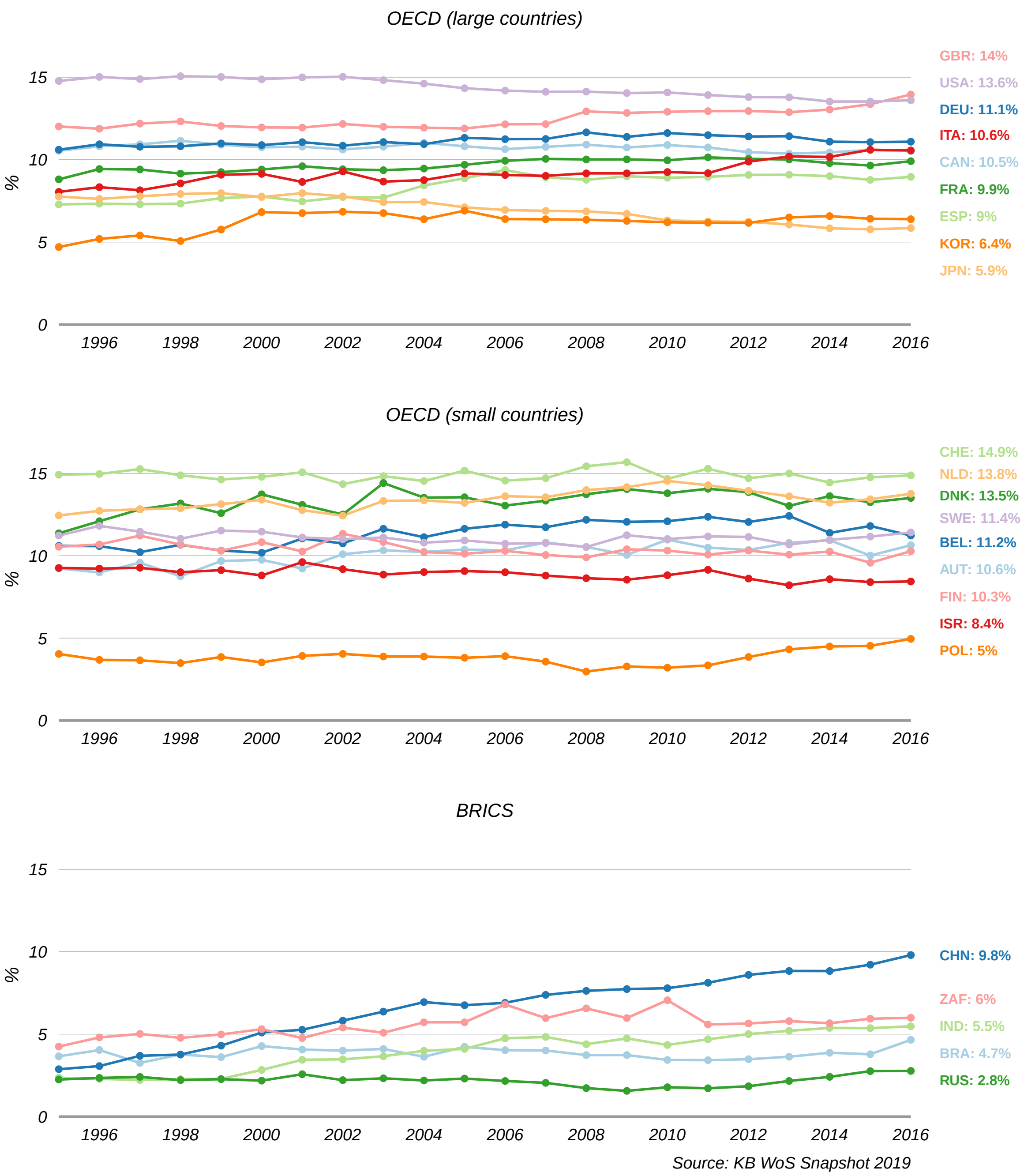

Figure 3: Excellence Rates by country in Web of Science over time.

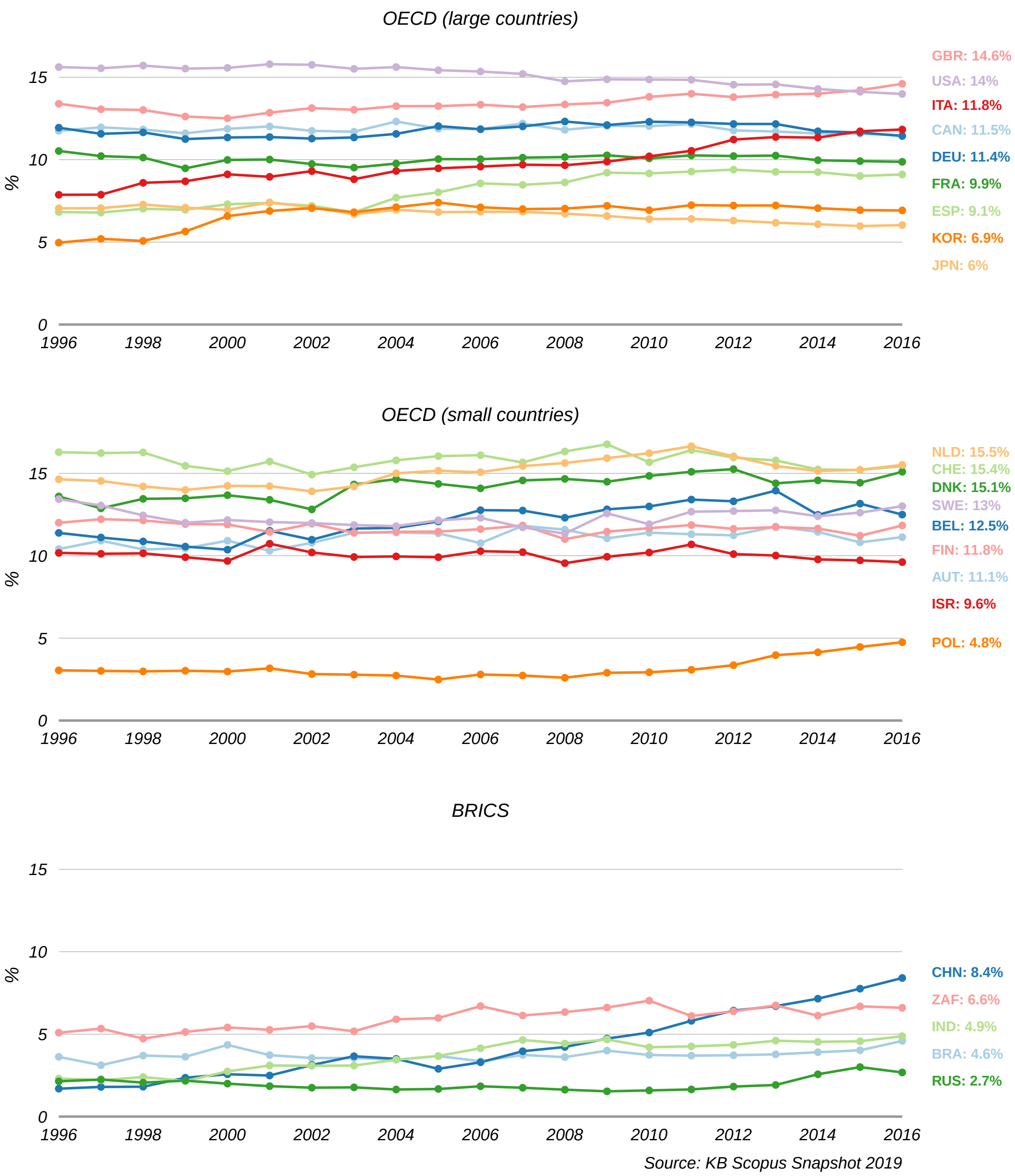

Figure 4: Excellence Rates by country in Scopus over time.

Scientific Regard and International Alignment

Scientific Regard (SR) is an indicator of how cited a country's publications are in comparison to other publications in the same journals. SR is scaled around 0 and so positive values indicate the country's publications were cited more often than average for the journals in which they were published, and negative values indicate they were cited less often than average. International Alignment (IA) complements the SR indicator as, while SR measures the citedness of publications within a journal, IA measures the citedness of the journals in which the country published in comparison to the global average of all journals. As such, higher IA values reflect greater visibility and impact.

In examining SR and IA values here, consideration should be given to the influence of the varying levels of coverage of disciplines by both databases on these citation-based indicators, as shown in Figure 16. Poor coverage of a discipline results in under-representation in citation counts of countries' publications and lower SR and IA values. As such, countries with a disciplinary profile that aligns with WoS' and Scopus' coverage have an advantage over countries with misaligned profiles. A report by the European Commission found European countries often had varied disciplinary focuses [4].

We present in Figures 5 and 6 the SR and IA values based on WoS data for each of the selected countries, and based on Scopus data in Figures 7 and 8. In all figures, darker shades of blue represent more negative values – indicating a country's publications or journals were cited less than average – and darker shades of red indicate higher citation rates than average. Citations are based on those received within 3 years of publishing and so data are presented up to the publication year 2016.

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
CHE	14	14	16	14	15	14	14	12	14	13	16	12	12	14	14	12	12	10	9	9	9	8
DNK	12	12	13	14	10	14	12	10	15	12	13	11	10	12	12	12	12	11	9	9	8	7
FIN	10	11	9	7	9	5	7	6	5	2	1	3	2	0	2	2	2	2	0	0	−1	0
SWE	13	12	12	10	11	9	7	8	6	4	4	4	1	1	4	4	3	2	3	1	1	0
AUT	6	8	4	2	4	5	5	7	9	8	7	6	8	4	5	5	4	5	2	4	2	3
GBR	8	8	9	8	7	6	6	7	6	5	5	4	4	5	6	6	5	4	4	5	6	7
USA	6	6	7	7	7	8	8	8	7	7	6	6	5	5	5	4	4	4	4	3	2	2
BEL	7	6	8	7	4	6	5	8	9	3	7	8	8	10	10	9	12	8	9	8	7	2
DEU	11	13	11	10	11	10	10	9	9	9	9	9	9	9	7	7	7	7	6	5	4	3
NLD	9	10	10	10	8	8	8	6	10	8	8	7	6	8	7	8	9	8	6	5	5	6
ITA	−4	−3	−5	−3	−4	−3	−5	−1	−4	−4	−3	−3	−2	−3	−1	0	1	6	8	8	10	9
CAN	2	0	0	2	2	1	2	0	2	2	1	0	1	0	−1	−1	−1	−1	−2	−2	−2	−2
FRA	2	3	3	1	1	0	1	−1	−1	−1	−1	0	2	0	0	0	0	−1	−2	−2	−4	−4
CHN	−23	−21	−16	−15	−9	−7	−5	−4	−1	1	0	−2	−1	0	0	0	1	2	2	2	2	3
POL	−12	−16	−15	−15	−10	−13	−10	−10	−8	−11	−11	−9	−10	−11	−10	−10	−5	−1	1	1	0	−1
RUS	−2	−1	−3	−5	−2	−3	−6	−4	−4	−6	−5	−6	−6	−6	−7	−5	−5	−5	−2	−1	−2	−1
ESP	−9	−9	−10	−11	−9	−5	−9	−7	−7	−3	−3	−1	−1	−2	0	−1	−1	−2	−2	−2	−4	−4
ZAF	−10	−12	−6	−10	−9	−7	−10	−11	−9	−8	−6	−2	−4	0	−2	1	−2	−5	0	−3	−3	−2
ISR	−9	−12	−11	−13	−12	−9	−7	−8	−12	−9	−12	−10	−12	−13	−13	−15	−14	−15	−17	−15	−15	−17
JPN	−5	−6	−5	−5	−5	−6	−5	−6	−7	−7	−8	−8	−9	−9	−11	−11	−12	−12	−13	−13	−14	−14
IND	−28	−26	−24	−23	−23	−19	−16	−16	−13	−13	−12	−8	−8	−8	−6	−7	−6	−5	−3	−3	−2	−2
BRA	−19	−22	−20	−17	−21	−16	−19	−18	−17	−18	−14	−14	−10	−7	−7	−7	−8	−10	−11	−12	−10	−9
KOR	−20	−18	−15	−15	−13	−11	−12	−9	−11	−10	−8	−9	−8	−7	−8	−8	−9	−9	−9	−10	−12	−13

Source: KB WoS Snapshot 2019

Figure 5: Scientific Regard by country in Web of Science over time.

Taking 2016 as a cross-section of the most recent standings, nine countries held positive values for both indicators in WoS, including Germany, Austria, Belgium, Denmark, Great Britain, Italy, the Netherlands, the USA, and Switzerland. As of 2016, Germany ranked sixth-highest of the 23 countries for IA and seventh-highest for SR, with higher IAs held by Switzerland (27.7), the USA (19.9), the Netherlands (19.1), Denmark (15.2), and Great Britain (14.6), and higher SRs held by Italy (8.8), Switzerland (8.2), Denmark (7.4), Great Britain (6.6), the Netherlands (5.6), and China (3.1). Austria held an SR equal to Germany. These countries with both high IA and SR values then both published in highly-cited journals and also received a higher number of citations than average for those journals.

Conversely, Korea, Japan, Poland and the BRICS countries except China all published in less visible journals and received fewer citations than average for their publications in these journals. Five countries – Spain, Sweden, France, Canada and Israel – all published in highly visible journals (positive IA), but received average or fewer than average citations for their publications in these journals (negative or zero SR). The converse was true for China, which published in slightly less visible journals (-0.6 IA), but received above average citations. Finally, in 2016, Finland performed as average for both indicators.

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
NLD	16	12	16	15	18	20	18	19	21	21	22	23	24	25	25	27	26	24	23	21	20	19
CHE	31	29	30	28	27	26	29	27	26	27	26	25	26	28	29	29	29	30	30	28	29	28
USA	31	32	31	31	30	30	30	30	29	28	27	27	28	29	29	28	27	26	25	23	22	20
DNK	8	8	9	10	13	14	13	12	16	15	15	15	17	19	19	19	17	17	17	17	15	15
GBR	7	7	8	10	9	10	11	14	13	13	14	14	16	19	18	18	17	17	16	17	16	15
DEU	1	1	2	1	3	4	4	5	7	6	9	9	10	13	15	16	16	15	15	15	14	14
CAN	1	5	6	6	7	7	8	8	8	7	8	8	10	10	11	11	11	10	9	8	8	7
ISR	8	4	6	6	8	6	8	8	8	9	10	10	11	13	12	14	15	14	11	11	8	9
SWE	7	9	9	7	8	10	10	10	10	10	11	11	14	14	14	15	13	11	12	13	12	12
ESP	−9	−7	−9	−9	−8	−7	−8	−8	−7	−4	−3	−2	−2	−4	−2	0	1	1	0	2	0	1
JPN	−6	−7	−4	−4	−3	−3	−3	−3	−3	−2	−3	−4	−2	0	2	1	2	1	0	−2	−3	−4
FIN	3	1	6	2	5	8	4	5	5	3	5	4	5	6	6	7	6	5	6	4	3	0
ITA	−1	2	2	2	6	5	3	5	4	5	7	6	6	7	6	6	5	4	3	3	2	2
AUT	−8	−5	−2	−4	1	2	−2	3	2	4	5	6	7	6	4	8	7	5	7	7	4	5
BEL	1	−4	1	−1	3	3	2	1	4	5	6	7	7	7	12	8	10	10	9	10	8	9
FRA	−3	0	0	−1	−1	1	2	1	2	2	3	4	6	7	8	10	12	11	12	12	11	11
RUS	−78	−78	−77	−78	−78	−78	−76	−76	−75	−75	−74	−74	−74	−76	−77	−75	−74	−70	−67	−64	−62	−60
ZAF	−51	−51	−47	−49	−47	−44	−43	−45	−42	−40	−38	−38	−39	−38	−39	−33	−40	−37	−35	−35	−34	−36
BRA	−37	−36	−38	−41	−40	−36	−38	−36	−37	−38	−36	−36	−43	−48	−49	−48	−48	−47	−43	−41	−39	−35
POL	−48	−45	−47	−47	−45	−44	−48	−43	−43	−45	−40	−41	−45	−53	−50	−48	−47	−48	−42	−39	−36	−33
KOR	−43	−39	−40	−38	−33	−26	−28	−26	−25	−24	−20	−21	−21	−21	−21	−19	−16	−13	−11	−9	−6	−6
CHN	−61	−60	−58	−58	−57	−50	−51	−46	−42	−41	−38	−35	−31	−26	−23	−19	−16	−12	−9	−7	−3	−1
IND	−58	−59	−59	−60	−58	−55	−53	−52	−52	−46	−45	−40	−40	−42	−37	−38	−35	−30	−29	−25	−23	−20

Source: KB WoS Snapshot 2019

Figure 6: International Alignment by country in Web of Science over time.

The interaction of these indicators should also be considered when examining the temporal patterns of each over time. For example, Germany mostly maintained a stable SR of between 9 and 11 until 2008, after which point it declined to 3 by 2016. However, its IA increased from 1 to 14 between 1995 and 2016. Alone, the SR suggests Germany's publications lost impact over time, however when viewed in relation with the IA, we can see that German researchers are now publishing in more impactful journals than before, but they exceed the average citations of publications to a lesser extent than they did in less impactful journals. Countries such as France and Belgium showed similar patterns. The USA however is decreasing in both indicators, suggesting it is losing visibility for both its publications and journals, although it continues to publish in very highly-cited journals.

Generally, the BRICS countries received fewer citations than average for both their publications and the journals they published in, however this improved over time. All of the BRICS countries but Russia improved their SR values, although only China began receiving more citations than average, and these countries maintained some of the lowest IA values, although both India and China demonstrated rapid improvement.

The trends were similar when looking at Scopus data. The highest SRs in 2016 were held by Italy (12.5), Denmark (8.9), Switzerland (8.6), Great Britain (7.8), the Netherlands (7.5), and Russia (7.0), with Germany in seventh (4.4). The presence of Italy and Russia amongst the highest-ranking countries for SR reflects recent rapid improvements for both countries over the past decade, and the disparity in Russia's ranks between Scopus and WoS extends from Scopus' greater coverage of regional journals. Conversely, Korea, Brazil, India, Japan, Israel, France, Canada, Spain, and Poland all had negative

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
CHE	15	18	19	17	14	15	12	14	13	14	14	13	13	14	11	12	10	9	10	8	9
DNK	14	20	14	13	15	13	11	14	13	13	11	12	12	12	11	11	13	8	10	9	9
BEL	9	9	9	6	5	8	11	10	8	7	10	8	10	10	8	10	9	11	6	9	4
DEU	14	13	11	11	10	10	10	10	10	11	10	10	9	8	8	8	8	6	6	5	4
NLD	14	13	11	10	11	10	7	11	10	9	8	10	10	10	9	10	8	6	7	8	8
AUT	2	7	7	6	10	5	6	8	6	10	10	9	6	8	4	3	5	5	4	2	2
USA	5	6	6	7	8	9	8	8	8	7	6	6	5	5	5	4	4	4	3	3	2
FIN	9	10	8	7	8	9	7	4	3	2	3	4	1	4	3	4	4	3	2	1	2
GBR	9	10	9	9	8	7	8	7	7	6	6	5	6	6	6	6	6	5	6	7	8
SWE	15	14	11	12	12	9	9	9	6	7	6	3	2	7	6	6	4	5	3	4	2
POL	−9	−9	−8	−5	−10	−9	−7	−5	−5	−6	−6	−7	−9	−8	−8	−2	−2	1	0	−1	−2
ZAF	−10	−3	−11	−5	−3	−2	−5	−5	−1	0	1	−1	−1	−2	3	−2	−3	3	−5	0	0
CHN	−11	−9	−6	−4	−6	−4	−4	−1	0	0	−1	0	0	0	0	0	0	0	0	0	1
ESP	−5	−6	−7	−5	−5	−5	−5	−5	−3	−2	−1	−1	0	0	−1	−1	−1	−2	−2	−3	−2
CAN	0	2	2	3	2	3	3	3	3	2	2	2	2	1	1	1	0	0	−1	0	−2
FRA	5	4	3	1	2	2	1	2	0	0	1	2	1	1	1	1	0	0	−1	−3	−2
ITA	−2	−3	0	−2	−2	−2	0	−4	−3	−2	−1	1	0	1	2	4	9	11	12	14	12
RUS	−4	−2	−1	−3	−2	−4	−4	0	−3	−1	−6	−8	−6	−7	−4	−4	−2	3	11	10	7
ISR	−11	−11	−12	−11	−7	−5	−7	−9	−7	−10	−7	−10	−12	−11	−13	−11	−12	−15	−13	−14	−14
JPN	−2	−3	−3	−4	−4	−4	−4	−6	−5	−5	−6	−6	−7	−8	−7	−9	−9	−10	−10	−12	−12
IND	−23	−22	−16	−17	−13	−11	−13	−9	−9	−8	−5	−4	−3	−2	−2	−1	0	0	−1	−4	−2
BRA	−14	−17	−12	−15	−13	−12	−12	−11	−12	−11	−8	−6	−5	−4	−5	−6	−6	−8	−8	−7	−8
KOR	−20	−17	−16	−13	−13	−15	−13	−12	−11	−9	−10	−8	−7	−8	−7	−8	−8	−8	−10	−11	−13

Source: KB Scopus Snapshot 2019

Figure 7: Scientific Regard by country in Scopus over time.

SRs. The BRICS countries and Poland also had the most negative IA scores, although China, India and Poland all improved over time. The highest IAs in 2016 were held by Switzerland (35.4), the Netherlands (29.9), the USA (29.3), and Denmark (27.6), with Germany in 8th place (19.6).

Germany exhibited the same downward trend in SR and upward trend in IA over time in Scopus as seen in WoS. In both databases, many countries experienced a peak in IA circa 2008-2010 followed by a decrease over time to 2016, including the USA, the Netherlands, Finland, Israel, Canada, Great Britain, and Denmark. The reason for this trend is unclear, however its presence in both databases suggests it reflects the science systems of these countries rather than a change in the databases.

The rapid improvement in Italy's SR values should be considered in light of recent research by Baccini, De Nicolao and Petrovich [2] noting an upward trend in intra-national citations subsequent to changes in Italy's national evaluation system that took effect in 2011.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
GBR	17	15	16	15	15	20	23	22	22	24	23	23	26	23	24	24	24	23	24	23	23
DEU	15	12	14	13	14	16	16	18	19	21	20	21	25	23	23	23	22	23	22	21	20
SWE	22	21	19	18	17	22	23	24	24	27	26	28	28	25	25	25	24	26	24	23	24
ISR	19	20	20	20	18	22	23	23	23	26	25	27	27	25	27	28	25	25	24	21	22
CAN	21	21	20	21	20	23	22	23	23	23	24	24	24	23	23	23	22	22	21	20	19
DNK	20	15	17	19	19	23	24	28	27	29	29	31	32	27	28	27	27	27	26	26	28
FRA	17	14	14	11	12	14	13	13	13	16	14	15	17	14	15	15	15	16	15	15	13
AUT	7	9	10	11	12	13	16	15	16	18	14	19	17	14	17	15	14	16	16	13	15
BEL	9	11	5	8	10	13	12	14	16	18	18	20	19	21	20	21	20	20	20	19	19
FIN	8	16	15	17	18	17	18	20	20	21	22	22	22	18	19	19	17	17	16	16	14
ITA	1	2	6	8	8	11	14	13	15	17	17	16	17	15	16	15	13	12	12	10	10
USA	40	40	40	39	37	40	39	39	39	39	39	39	39	36	36	35	34	34	33	31	29
CHE	39	39	39	35	32	37	36	35	38	36	36	36	38	37	37	37	38	36	36	36	35
NLD	25	26	24	27	27	29	31	32	34	35	35	36	37	35	38	37	35	34	32	31	30
JPN	−7	−8	−5	−5	−6	−1	0	1	1	2	1	3	4	2	0	1	0	1	1	−1	−2
ESP	−6	−7	−5	−5	−5	−3	−2	−2	1	2	2	5	4	5	7	6	6	5	6	3	2
KOR	−26	−31	−28	−22	−15	−12	−11	−8	−6	−2	−3	−5	−6	−5	−5	−2	−1	−1	−1	2	1
RUS	−77	−77	−79	−77	−78	−81	−81	−81	−80	−80	−75	−76	−76	−76	−76	−76	−71	−70	−70	−68	−71
CHN	−80	−82	−84	−81	−76	−79	−69	−64	−70	−76	−70	−63	−57	−51	−46	−37	−32	−28	−20	−13	−8
ZAF	−41	−41	−41	−42	−41	−39	−42	−39	−42	−35	−33	−27	−28	−29	−27	−32	−29	−32	−39	−31	−33
BRA	−33	−36	−37	−35	−32	−35	−35	−34	−32	−31	−39	−37	−39	−41	−42	−43	−44	−42	−39	−39	−36
POL	−53	−55	−55	−54	−53	−52	−53	−54	−58	−55	−54	−50	−49	−51	−49	−48	−47	−44	−37	−33	−31
IND	−53	−57	−57	−58	−57	−53	−52	−52	−46	−45	−40	−37	−38	−38	−43	−45	−44	−41	−40	−41	−37

Source: KB Scopus Snapshot 2019

Figure 8: International Alignment by country in Scopus over time.

Section 2: Contextual information

This Section provides contextual information for the interpretation of the indicators presented in Section 1. Here we examine key aspects of the databases, such as the distribution of the content across document types, languages and disciplines over time, the extent to which important metadata is missing or differs between the databases, and the internal coverage of the databases, which is relevant to citation-based indicators. These data help to draw a descriptive distinction between changes attributable to the national science systems and to the underlying databases.

The number of publications and journals indexed

The number of publications and journals indexed in each database and the databases' rate of growth is important in interpreting the country-level performance indicators presented in Section 1. For example, the size of a database is relevant as publication counts for countries would be higher from a database with a larger number of items. Similarly, a larger collection increases the likelihood that any specific item is cited within the database, which is important as the citation-based indicators SR, IA and ERs are derived from intra-database citations. Further, the rate at which publications and journals are indexed could influence a country's time-series. As such, we examine here the databases' indexed items and journals over time to understand any relevant influences.

Throughout the time-series, Scopus held a consistently higher number of journals than WoS, as seen in the top panel of Figure 9. However, both databases grew at similar rates over time: the number

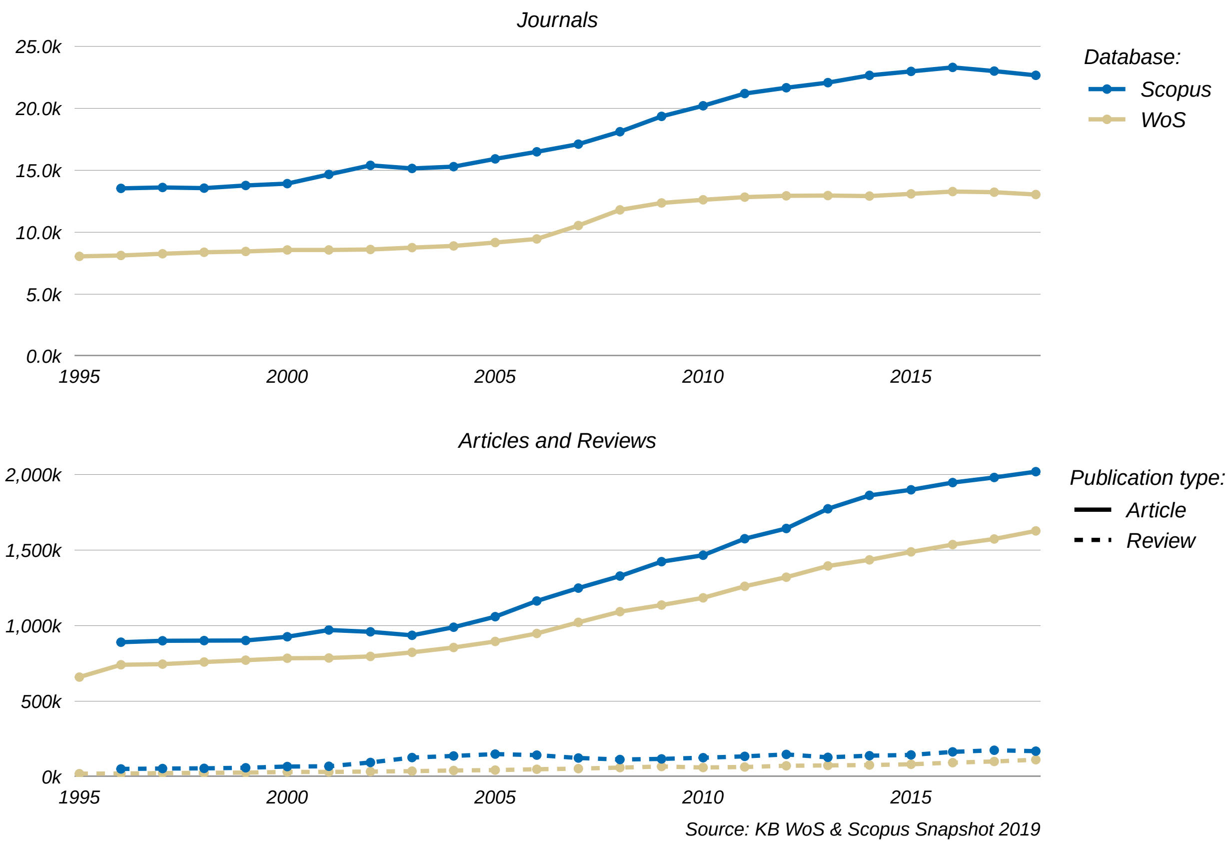

Figure 9: The total number of journals, articles and reviews indexed in each database over time.

of journals indexed in Scopus rose by 67.4%, from 13,500 journals in 1996 to 22,600 journals in 2018, and by 61.9% in WoS, from 8,000 in 1995 to 13,000 in 2018. As would be expected, this translated to a higher number of both articles and reviews in Scopus over the time-series – in the order of an additional 392,000 articles and 57,000 reviews in Scopus in 2018, as shown in the bottom panel of Figure 9. Both databases steadily grew over time, with both more than doubling their articles between 1995/6 and 2018 – an increase of 147% in WoS and 127% in Scopus – while WoS increased its coverage of reviews by 449%, and Scopus' review coverage rose by 227%. Note, however, that as shown in Figure 18, there is often disagreement between the databases about the classification of items as reviews. The increased rate of growth in both databases from 2004 may reflect the effect of Scopus coming to market and the challenge it posed to WoS as the market leader at the time.

This potential reaction of Clarivate Analytics to Scopus' entrance is also reflected in Figure 10, which shows the number of journals added or removed from each database annually. The addition of over 1,200 journals to WoS in both 2007 and 2008 may have been spurred by Scopus' larger content. Overall, we can see that the content of Scopus is typically more variable than that of WoS, with more journals added or removed per year than in WoS. This likely relates to Elsevier's more inclusive approach to Scopus, allowing more journals to be indexed each year, but also requiring the removal of journals that no longer meet the inclusion criteria, which may happen more often when the original indexation criteria in Scopus are more easily met than in WoS.

We have calculated the data for the journals added or removed over time based on the annual number of publications published by the journal. The baseline is 1995 for WoS and 1996 for Scopus, with

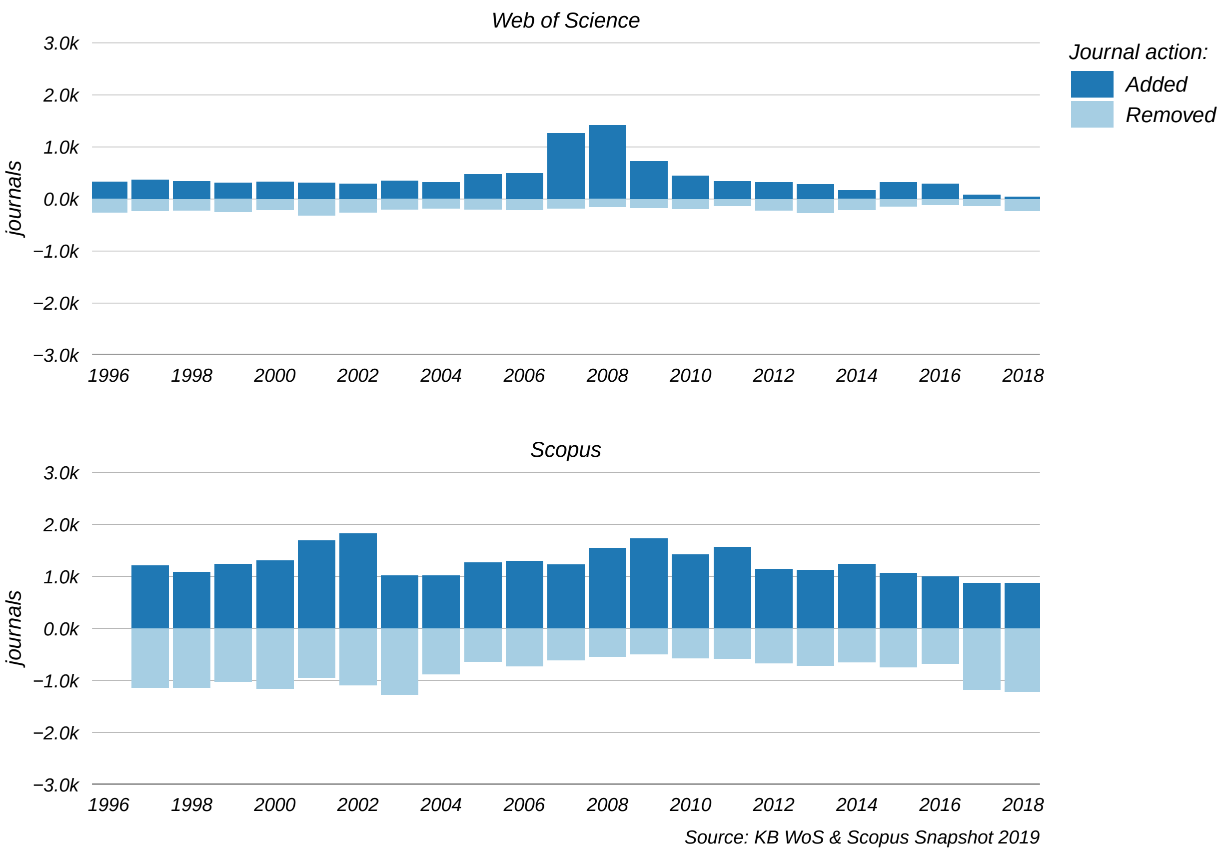

Figure 10: The yearly number of journals added and removed from each database over time.

the number of journals added or removed based on changes from the previous year. A journal has been counted as "added" in the first year after 1995/6 that it begins having publications recorded. Conversely, if an existing journal publishes no new content for a year, it is counted as removed from the database for the first year in which there is no content. On this basis, it is possible a journal may have stopped publishing rather than having been de-indexed from the database. However, the policies of Clarivate Analytics [1] and Elsevier [5] toward both the initial indexing and continued indexation include that a journal must publish regularly. As such, if there is no new content for a journal, it either has been de-indexed or likely would be de-indexed soon after it stops publishing.

Database composition by document type

Bibliometric analyses commonly focus on articles and reviews as the document types of interest. However, over a long time-series it is relevant to examine whether the vehicles of scientific communication may have changed and influenced bibliometric indicators. Further, indicators are normalised by document type to account for differing patterns of citations for different documents. We show in Figure 11 the percentage of items by document type in WoS and Scopus.

There are notable differences in the composition of each database. While both databases are comprised of about 65% articles in the most recent years, there is a substantially larger percentage of "Other" document types in WoS – 19.3% per year on average – than the 4.5% in Scopus. This occurs as WoS indexes all documents from the journals it includes, while Scopus indexes only key document types. In particular, WoS includes meeting abstracts, while Scopus does not. Also, there are about double

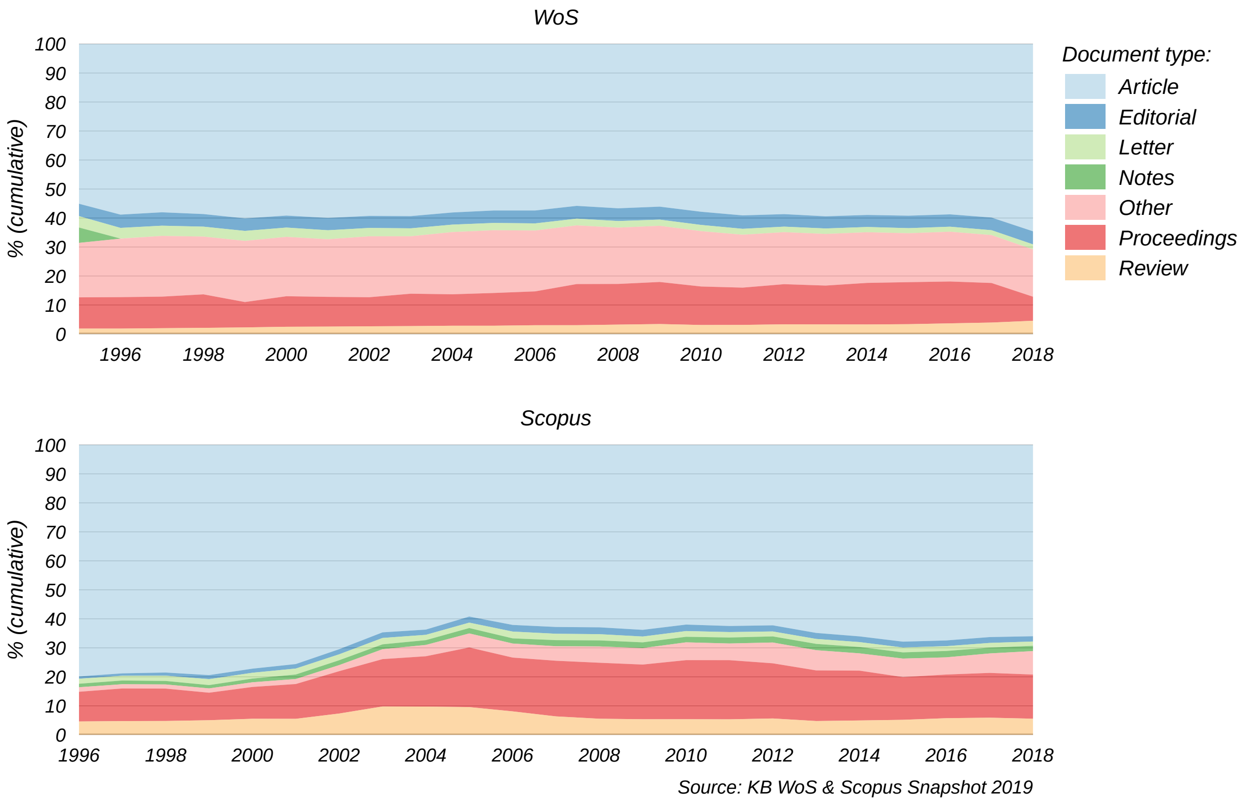

Figure 11: The percentage of publications indexed in Web of Science and Scopus by document type over time.

the percentage of reviews in Scopus (6.1% per year on average) than WoS (3.0%). Proceedings make up a similar and substantial percentage of each database – approximately 16% of Scopus documents and 12% of WoS documents. In analysing only articles and reviews then, we utilise on average 74% of the documents in Scopus and 62% of the documents in WoS.

Database composition by discipline

We also examine the composition of the two databases by means of the number of publications in each field and discipline. Bibliometric indicators are often normalised against disciplines to account for differences in citation practices between disciplines. As such, the composition of the database and its disciplinary coverage can be important for countries with a strong focus on particular disciplines as poor coverage of a discipline can influence the country's performance assessment. As previously noted, it has been determined that countries often vary in their disciplinary focus [4], which may present an advantage for countries with discipline profiles well-covered by the databases. This section describes the composition of the databases by field and discipline, while the following section examines the coverage of each of the fields.

Data are presented here based on the OECD's Fields of Science and Technology (FOS) classification, having been fractionally assigned from the native classifications used in WoS and Scopus. In Figure 12 we show the number of publications in each OECD field in each database, while Figures 13 and 14 break this down to the number of publications in each discipline of the three larger fields and three smaller fields.

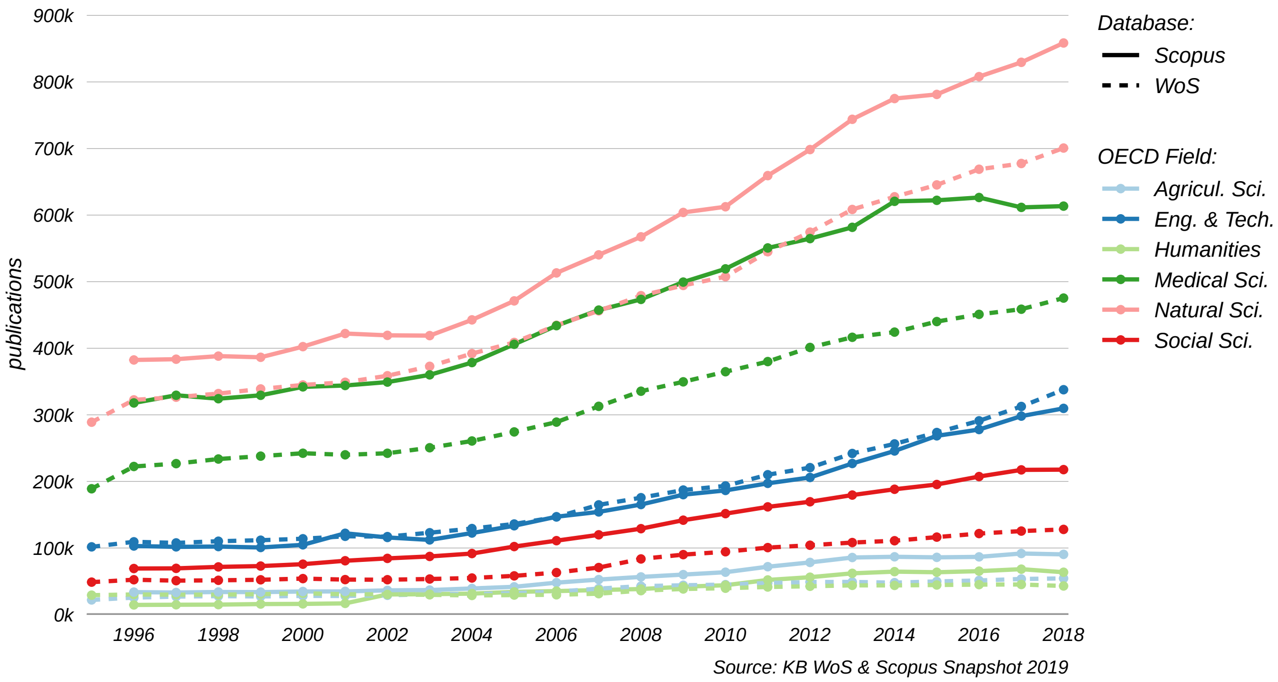

Figure 12: The number of publications in each OECD field and database over time.

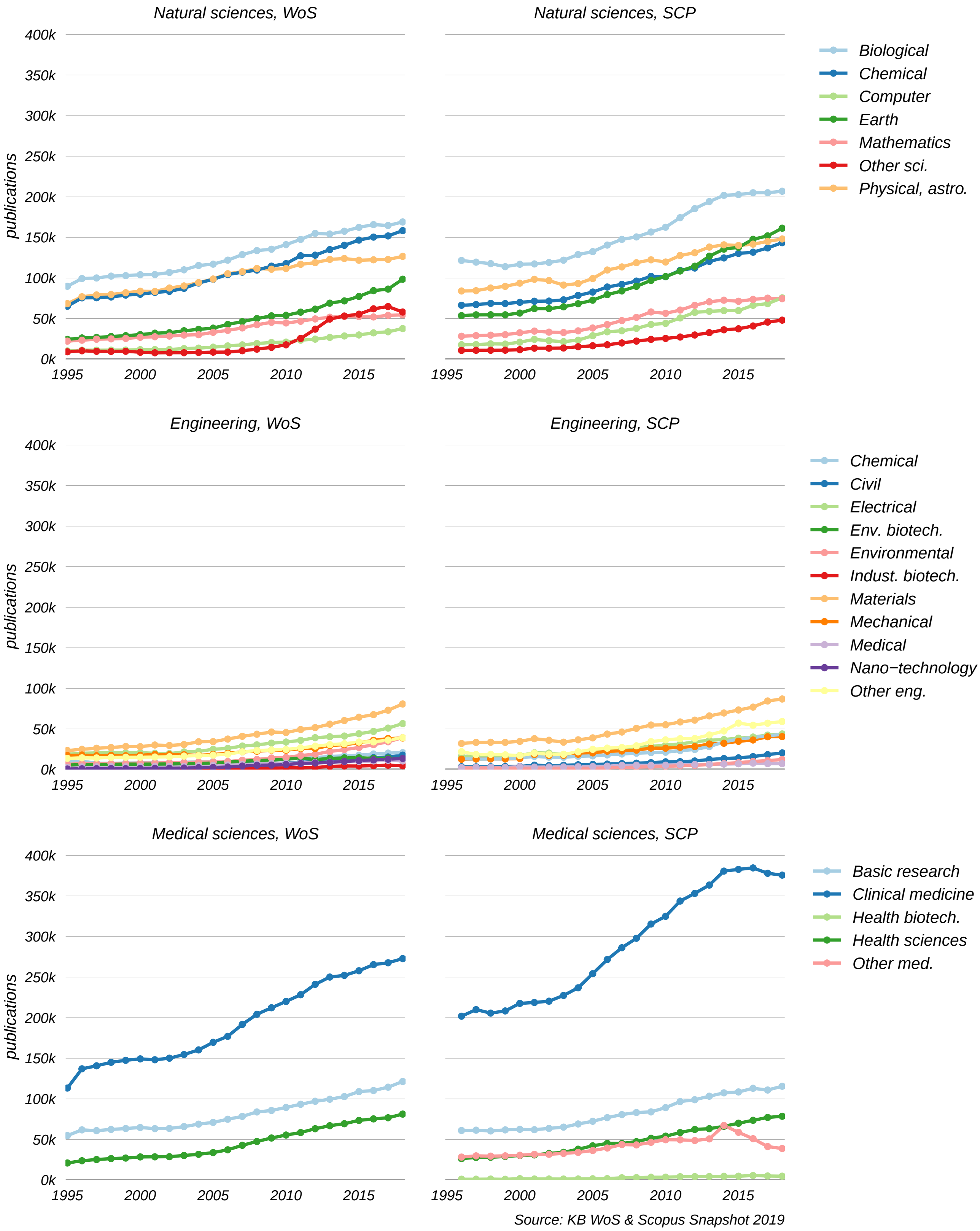

Figure 13: The number of publications in each OECD discipline of the larger fields by database over time.

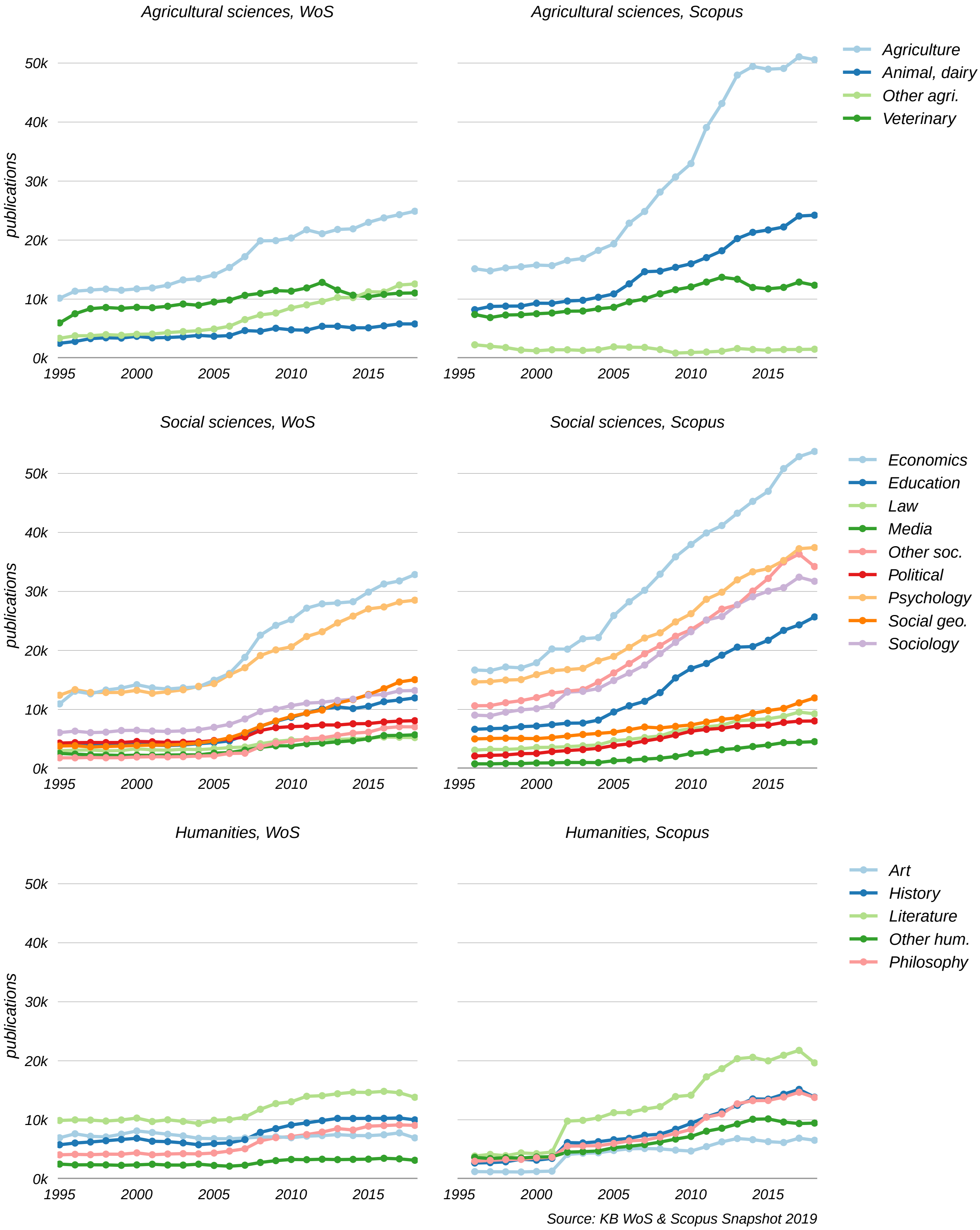

Figure 14: The number of publications in each discipline of the smaller fields by database over time.

Figure 12 highlights the emphasis of both databases on the medical and natural sciences, with these disciplines having the highest numbers of publications by a large margin. Also, Scopus contained higher numbers of publications in all disciplines than WoS across the time-series, except for engineering and technology and the humanities prior to 2002. The number of publications in all disciplines continued to grow throughout the time-series for both databases.

Figures 13 and 14 suggest the disciplines making the largest contributions to those fields are also similar between databases, although publication numbers are higher in Scopus. For example, biological sciences and clinical medicine are the two disciplines contributing the largest number of publications to the natural and medical sciences fields, respectively, while economics and psychology drive the increases in the social sciences. In both databases clinical medicine is the discipline with the largest number of publications, with nearly 400,000 publications in Scopus and over 270,000 in WoS. There are differences in the mappings of disciplines to the FOS classification however, as can be seen in the medical sciences where Elsevier maps its publications to five categories whereas Clarivate Analytics uses only three categories.

The percentage of uncited publications

The annual percentages of publications that received no citations within 3 years of publication are shown in Figure 15. As would be expected, the percentages decrease over time as the database providers continue to index journals, increasing the likelihood of publications having at least one citation from within the indexed corpus. Despite its larger coverage however, Scopus contains a higher percentage of uncited items, which likely reflects its greater number of regional journals from which publications are less likely to be cited, and also its current lower internal coverage depicted in Figure 16.

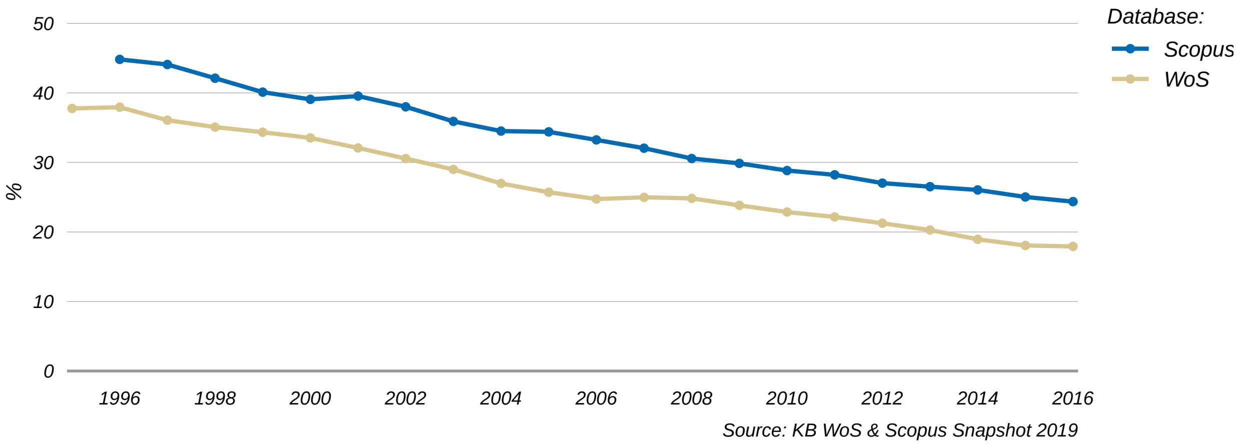

Figure 15: The percentage of uncited publications in each database over time.

The internal coverage of the databases

Internal coverage is a measure of the databases' coverage of a discipline's corpus. Internal coverage is the percentage of an indexed item's references that are also indexed in the database (also called source items), as opposed to not indexed. This reflects the database's coverage of a discipline because, if the majority of a source item's references are other source items, the database could be said to adequately cover the majority of key journals in the discipline, or that there is a good match between what the database providers deem relevant to index and what researchers deem relevant to cite. Conversely, if most references are non-source items, the coverage of the discipline is insufficient and raises questions about the validity of bibliometric indicators derived for these disciplines. For instance, if a large percentage of an indexed item's references are to non-source items, the reverse is also likely true, that the source items are missing a large percentage of their citations from non-source items. This is important because only citations between source items are included in citation-based indicators. Therefore, the coverage of a database is integral to the validity of citation-based indicators.

The percentage of references that are source items is expected to increase over time as the database providers continue to index journals and improve the coverage of under-represented disciplines, however the percentage is unlikely to ever reach 100% as authors will continue to cite items outside of the scope or coverage of the databases.

We present in Figure 16 the percentage of publications' references that are to source items over time by database and the OECD fields. This figure highlights that WoS is a well-established database, having had its initial foundations set in the 1960s, while Scopus has rapidly developed its index since its emergence in 2004. As such, Scopus' coverage is lacking prior to at least 2010, with coverage of less than 45% for all disciplines. WoS' corpus of medical and natural sciences journals is firmly established, with internal coverage over 80% since 2004 for the medical sciences and over 75% since 2010 for the natural sciences. The growth of both disciplines has slowed to an average of less than 1% per year in the last decade. Coverage of the agricultural sciences and engineering and technology continues to grow over time by on average 1.7% per year. These four fields could be considered to have good coverage with all over 65% since 2010. However, the social sciences and particularly the humanities have notably poorer coverage with only 21% of humanities and 56% of social sciences references to source items in 2018. This may be in part because a large proportion of humanities output is published in books, while this analysis examines only articles and reviews. These fields exhibited the highest growth rates however, with humanities coverage increasing by on average 5.1% and social sciences by 2.4% per year over the last decade.

As is evident from Figure 16, the growth of all disciplines is higher in Scopus with all disciplines increasing by more than 8.6% on average per year since 2009. The overall coverage remains lower than WoS however, with the medical sciences having the highest coverage (69.9% in 2018), followed by engineering and technology (68.1%), the natural sciences (65.8%), agricultural sciences (57.3%), social sciences (48.5%), and the humanities (25.9%).

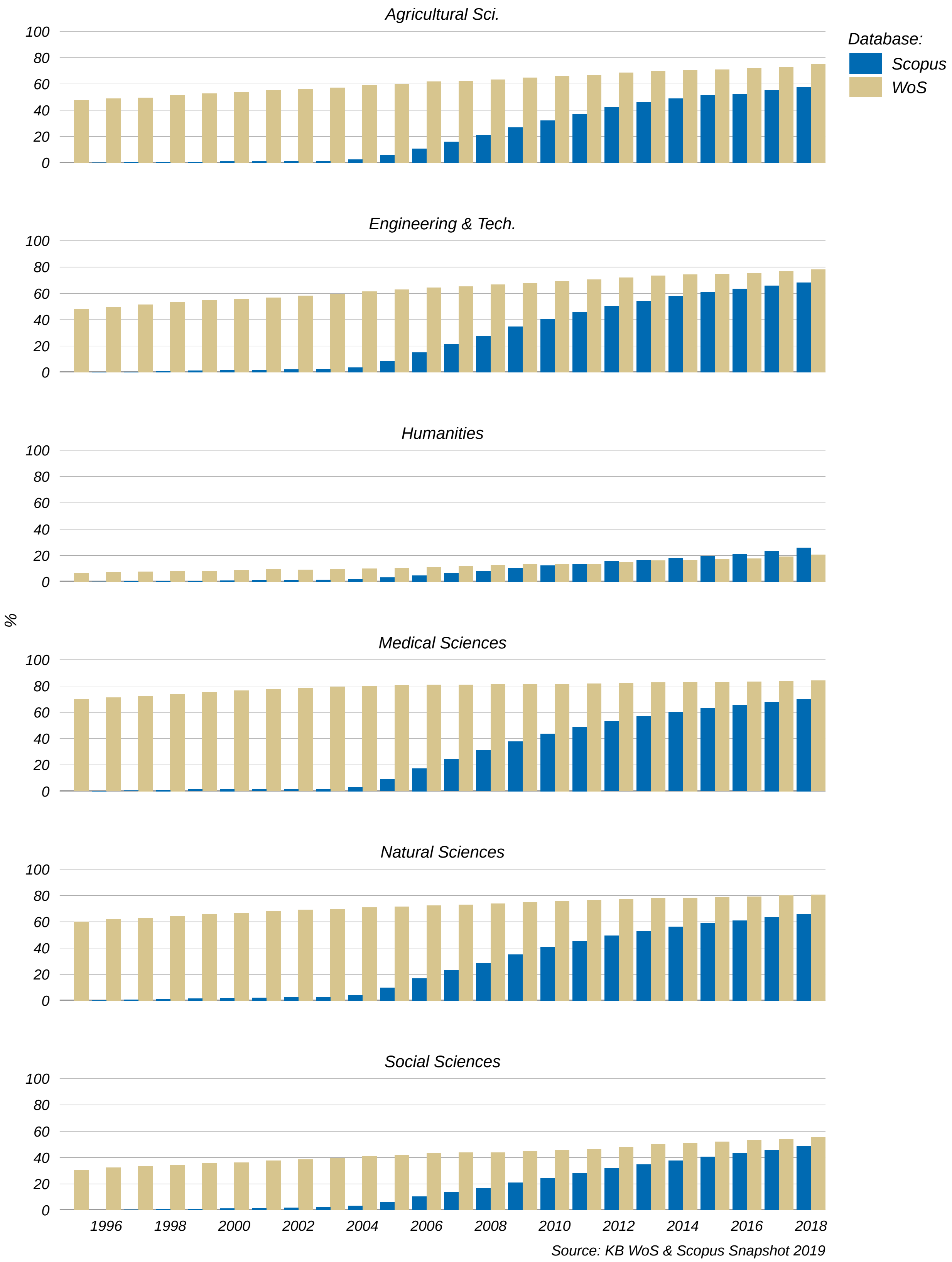

Figure 16: The percentage of source items' references that cite other source items by database over time (internal coverage).

The percentage of publications in languages other than English

The percentage of indexed publications written in languages other than English is a useful approximation of the inclusion of regional journals in the databases. In Figure 17, we show the percentage of publications in each database that are written in six key languages and an “other” grouping. In both databases, English-language publications comprised the majority of publications. German-language publications held the next largest share in both databases during the late 1990s, however this was quickly overtaken by Chinese-language publications in Scopus. The Chinese share increased from 1.5% in 1996 to 6.8% in 2006, before declining over the following years to 3.8% in 2018. Chinese-language publications followed a similar trend in WoS, however at a small magnitude, increasing from only 0.2% to 0.7% before declining again to 0.3% in 2018. Overall, in both databases, the percentage of publications in most non-English languages decreased steadily over time. Spanish and Portugese are exceptions to a point as both of these languages increased slightly in both databases circa 2006, although Portugese began to decline again in 2013.

These changes in the percentage of publications in languages other than English reflect authors' choices in the language used to write in and also changes in the journals covered by the databases. Disciplines are also influential, as medical and natural sciences are generally published in English given they address internationally-relevant research questions, while social sciences and humanities, with their greater focus on local populations and research questions, are more likely to be published in local languages. As such, while coverage of the humanities and social sciences is improving in both databases (see Figure 16), the growing medical and natural sciences literature continues to comprise the majority of indexed publications (see Figure 12). Notably, Scopus allows publications to be assigned to multiple languages, which may partly explain the substantially larger percentage of non-English-language publications in Scopus compared to WoS, however Scopus also holds a greater number of regional journals than WoS.

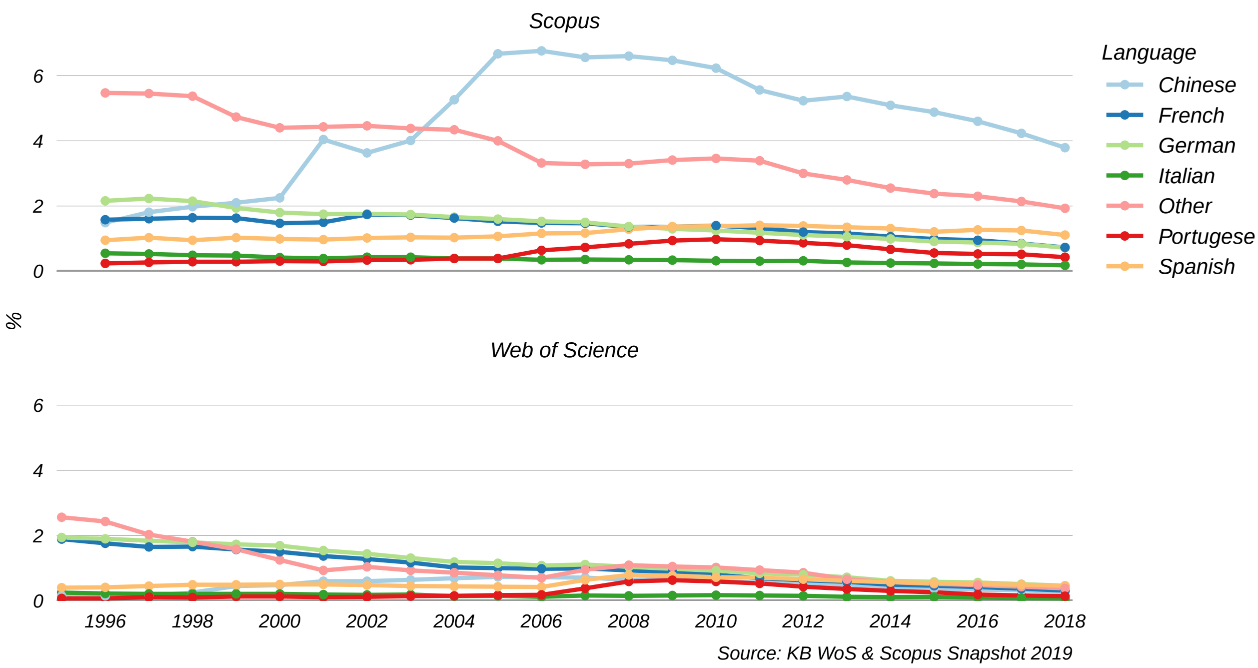

Figure 17: The percentage of publications written in languages other than English over time.

Agreement on key publication metadata

In the following section we compare three key metadata for publications indexed in both databases, referred to here as "duplicate" publications. These metadata are publication year, document type, and the number of countries with which the publications' authors are affiliated. A large percentage of disagreement between the databases in the metadata recorded for the duplicate publications could result in differences in indicators that are not reflective of the science system but database-particular processing data. We present data in this section by publication year based on the year recorded in WoS, however this does not signify that the data in WoS are presumed to be correct.

Publication year

Agreement on the year in which a duplicate publication was published is important in ensuring consistency in the annual publication counts between databases. In Table 1 we present the annual percentage of duplication publications with no difference in the recorded publication years, and the percentage with a different publication year in Scopus than in WoS. Overall there was a very high level of agreement between the databases with 99% of duplicate publications having the same year recorded in both databases, except in 2015 when 1.5% of publications had an earlier publication year in Scopus.

Table 1: The percentage of publications with differences in publication years between databases over time.

Year	Earlier in Scopus	No difference	Later in Scopus
1996	0.0	99.9	0.1
1997	0.2	99.7	0.1
1998	0.3	99.5	0.2
1999	0.3	99.6	0.2
2000	0.2	99.6	0.2
2001	0.2	99.5	0.3
2002	0.2	99.6	0.2
2003	0.2	99.7	0.2
2004	0.1	99.8	0.2
2005	0.1	99.7	0.2
2006	0.1	99.8	0.1
2007	0.1	99.8	0.1
2008	0.1	99.7	0.1
2009	0.1	99.7	0.2
2010	0.4	99.5	0.1
2011	0.2	99.7	0.1
2012	0.1	99.7	0.1
2013	0.2	99.7	0.1
2014	0.3	99.5	0.2
2015	1.5	98.4	0.1
2016	0.7	99.2	0.1
2017	0.2	99.7	0.1
2018	0.2	99.8	0.0

Document type

The correct classification of documents is important as citation counts are normalised by document types. We show in Figure 18 the annual percentage of agreement between the databases on the document type of duplicate articles and reviews. As these data pertain only to articles and reviews, where there is a mismatch and WoS records the publication as an article, it is recorded as a review in Scopus and vice versa. In general, agreement on articles is quite high with more than 93% agreement each year. There is more disagreement between the databases in classifying reviews however, with up to 48% of reviews in WoS classified as articles in Scopus in 1996, although this improves over time to around 20% since 2010. As such, there is an element of difference in how reviews are classified in WoS and Scopus, which is also relevant when considering Figures 9 and 11, which indicate Scopus contains a greater number and percentage of reviews than WoS.

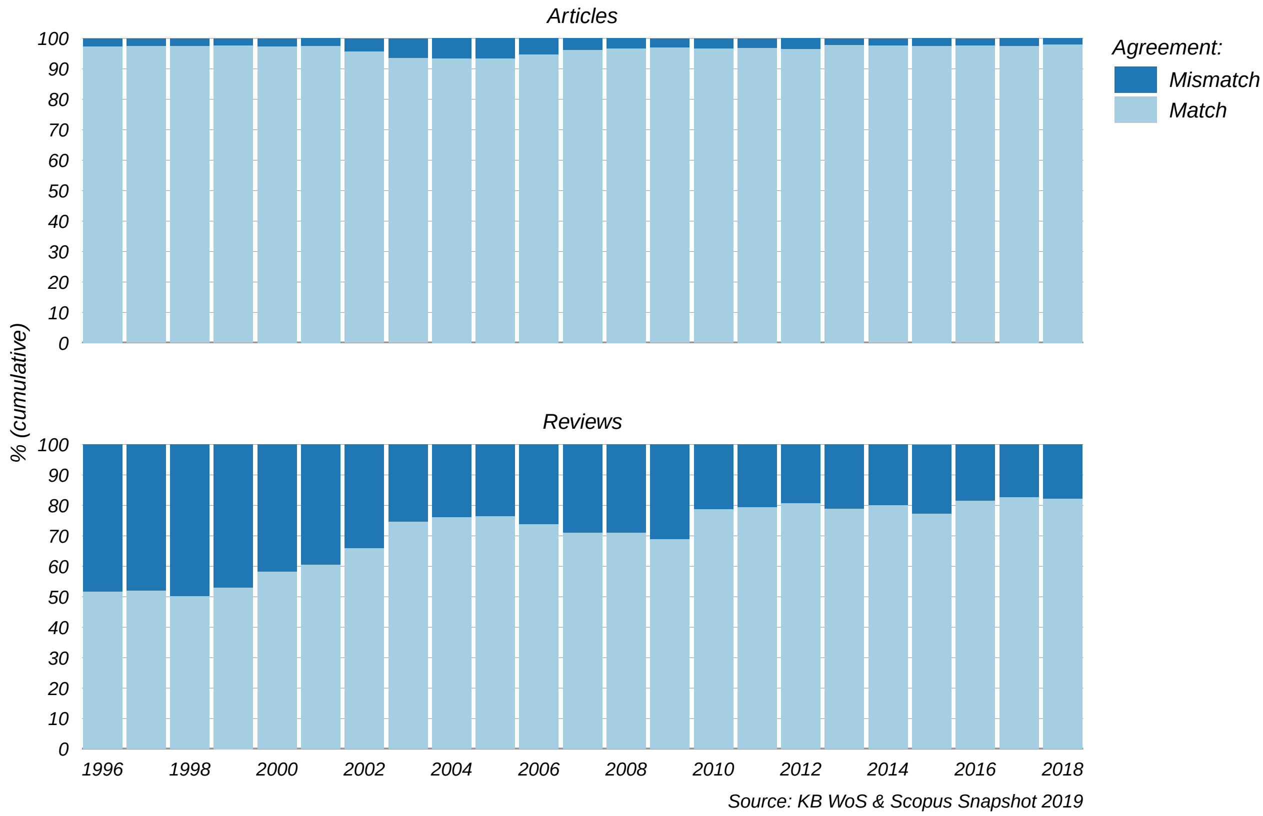

Figure 18: The percentage of publications with matched or mismatched document types between databases over time.

Number of authoring countries

Differences between databases in counts of the number of countries with which authors are affiliated could have important implications for fractional counting. We present in Table 2 the percentage of publications each year for which the count of countries is the same or different. Agreement between databases on country counts improves over time, increasing from 90% in 1996 to 98% in the most recent years. This small level of disagreement is unlikely to exert a serious influence on bibliometric indicators.

Table 2: The percentage of publications with differences in the number of authoring countries between databases over time.

Year	More in Scopus	No difference	Fewer in Scopus
1996	4.7	89.7	5.7
1997	4.8	91.1	4.1
1998	3.2	92.5	4.3
1999	2.8	92.9	4.3
2000	2.8	93.2	4.1
2001	2.6	92.3	5.1
2002	2.6	92.6	4.8
2003	2.8	94.8	2.4
2004	2.9	95.3	1.8
2005	3.1	95.4	1.6
2006	2.8	95.7	1.5
2007	2.6	95.3	2.1
2008	2.3	96.0	1.7
2009	2.3	96.3	1.4
2010	2.4	96.4	1.3
2011	2.1	96.8	1.1
2012	2.1	97.0	0.9
2013	2.0	97.0	0.9
2014	1.9	97.1	1.0
2015	1.8	97.1	1.0
2016	1.1	97.9	1.1
2017	1.0	97.9	1.1
2018	1.2	97.9	0.9

The extent of missing data

A publication is typically excluded from analysis when it is missing the required information to allocate it to a group, such as a country or discipline. Consequently, it is important to understand the extent to which key data are missing and the influence this missingness could have on indicators. Further, Elsevier and Clarivate often undertake projects to improve aspects of their databases, including known issues with missing data, which affect data for only a portion of the time-series. Examining the extent of missing data can provide important contextual information about changes in indicators over a time-series. We examine here four key variables: missing discipline information, missing author-affiliation links, and missing institution and country information.

Missing discipline data

Many scientific practices vary between disciplines, such as citation behaviours. As such, the calculations of bibliometric indicators often normalise by or otherwise take into account disciplinary information. For instance, ERs are based on the most highly cited publications per disciplines. Items that are missing discipline information would then be excluded from the calculation of ERs. In the top right panel of Figure 19, we present the percentage of publications in each database that were missing discipline information. WoS has maintained a low level of missing discipline information throughout the time-series. The percentage of items missing discipline information in Scopus was higher in previous years, however has decreased over time and is essentially zero from 2007 onwards. As such, the recent level of missing discipline data is very low in both databases.

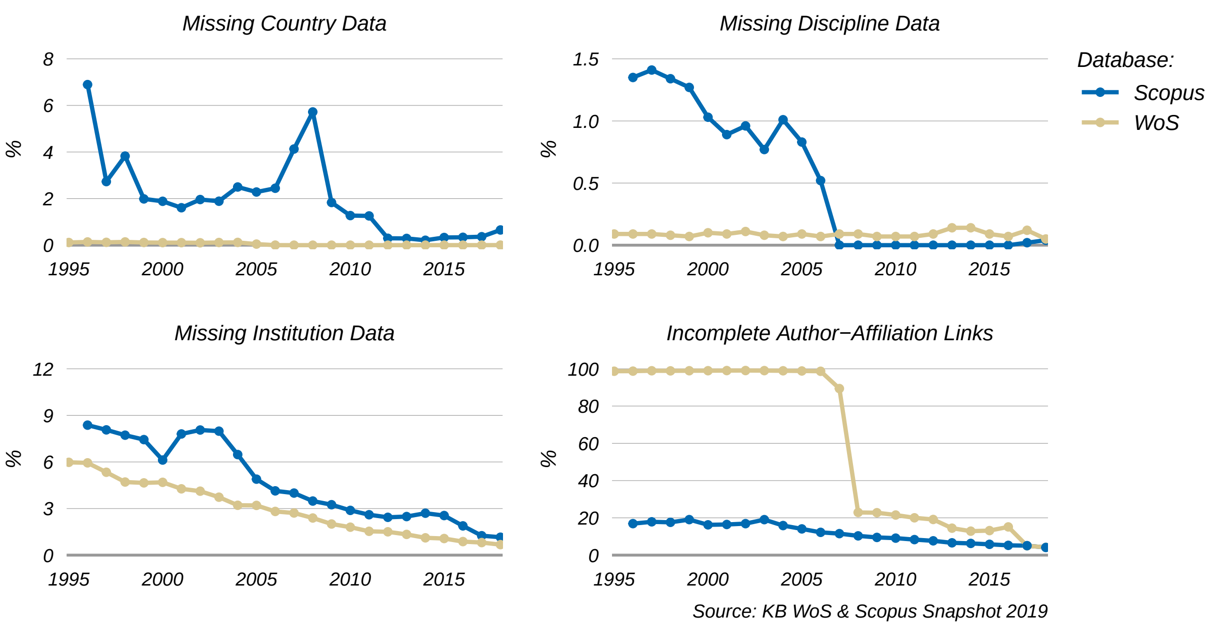

Figure 19: The percentage of publications with missing country data (top left), missing discipline data (top right), missing institution data (bottom left) and incomplete author-affiliation links (bottom right) by database over time.

Missing author-affiliation links

Author-affiliation links refer to the matching of a publication's authors with their organisational affiliations and is important for fractional counting of publications to countries. The percentages of publications with incomplete author-affiliation links in each database are shown in the bottom right panel of Figure 19. Scopus shows a continuous improvement in data quality with a steadily decreasing percentage of publications with missing links from around 17% prior to 2004 to 5% in recent years. Beginning in 2007, Clarivate Analytics made a concerted effort to improve author-affiliation links in WoS, reducing the incomplete links from nearly 100% prior to 2007 to less than 25% in 2008 and steadily decreasing thereafter to similar levels as Scopus in the most recent years.

Missing country or institution information

As could be expected, country information is integral to allocating publications to countries. Missing country data can occur in two ways. First, absolutely no information about the authors' institutions is available for an item, or secondly, institution information is available but there is no information about the country. We present in the bottom left panel of Figure 19 the percentage of publications in each database with no institution information, and in the top left panel the additional percentage of publications with no country data. As such, the total percentage of publications that are excluded from analysis is the percentage missing due to lack of institution information, in addition to the percentage missing due to lack of country data.

In both databases the percentage of publications with missing institution data has decreased over time, from 8% to 1% between 1996 and 2018 in Scopus and from 6% to 1% between 1995 and 2018 in WoS. Throughout the time-series, WoS has also maintained a very low percentage of publications with missing country data, at essentially 0%. Scopus data shows an unstable pattern, but suggests a downward trend, with less than 1% of publications missing country data since 2012.

Appendix 1: Country code list

Table 3: List of names and codes of countries included in this report.

Country	**Code**
Austria	AUT
Belgium	BEL
Brazil	BRA
Canada	CAN
Switzerland	CHE
China	CHN
Germany	DEU
Denmark	DNK
Spain	ESP
Finland	FIN
France	FRA
Great Britain	GBR
India	IND
Israel	ISR
Italy	ITA
Japan	JPN
Korea	KOR
The Netherlands	NLD
Poland	POL
Russia	RUS
Sweden	SWE
United States	USA
South Africa	ZAF

Appendix 2: Methodological details

This appendix discusses key methodological details to be considered in the interpretation of data from this report. This report is predominantly based on document types 'articles' and 'reviews' from the publication type 'journal'. Data are extracted from the Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Arts and Humanities Citation Index (A&HCI) from the Web of Science (WoS), and from the entire Scopus database. Fractional counting is used for all data disaggregated by country or discipline. Publications are examined for the period 1995 to 2018 from WoS and 1996 from Scopus, in accordance with the availability of data to the German Competence Centre of Bibliometrics. Citation data are examined for the period 1995 or 1996 to 2016 based on a 3-year citation window. The following sections describe key features of the methodology used in compiling this report and which should be considered when interpreting these data.

Whole versus fractional counting

There are two methods for counting publications that have more than one author – whole counting or fractional counting. Whole counting assigns a whole count of the paper to each author so one paper is considered as one contribution from each author and the country with which they are affiliated. Consider, for example, one paper with an author from Germany and an author from France. Using whole counting the paper would count as one paper each for Germany and France, for a total count of two papers. Evidently this method of counting inflates the overall number of publications. One method of remedying this inflation is to award each author a proportion of the paper, known as fractional counting. In this example, equal proportions of the paper would each be attributed to Germany and France and the total number of publications remains at one. Fractional counting can however disadvantage countries that regularly collaborate internationally as they lose a proportion of these papers from their totals, and this should be considered in interpreting the data here.

This report uses fractional counting for all data disaggregated by country. For the complete Scopus time-series and the latter part of the WoS time-series, fractionalisation was applied at the level of the author. That is, a proportion of the paper was attributed to each of the author's affiliated institutions and these proportions are summed across the corpus of relevant papers for each country. This approach sufficiently captures the multiple international affiliations authors may have and provides the most accurate counts of each country's publications. However, the quality of the data linking authors with affiliations prior to 2008 in WoS was inadequate to support author-level fractional counting (see Figure 19). As such, for years prior to 2008, fractional counting has been applied at the level of the organsation rather than the author. This would produce somewhat different counts than if author-level counting was applied throughout the time-series. For instance, Table 4 shows an example of a publication's authorship with four authors from 3 organisations in 2 countries. When fractional counting is applied to the authors, each author receives 0.25 of the publication which aggregates to 0.75 for country 1 and 0.25 for country 2. When the fractional counting is applied at the organisational level, each organisation receives 0.33 of the publication which aggregates to 0.66 for country 1 and 0.33 for country 2. In this way, fractional counts of countries' publications will differ between 2008 and later years. For further information, Waltman and Eck [8] provides a useful discussion on counting methods and their impacts on field-normalised indicators.

Table 4: An example of a publication's authorship

Author	Organisation	Country
Author 1	Organisation 1	Country 1
Author 2	Organisation 1	Country 1
Author 3	Organisation 2	Country 1
Author 4	Organisation 3	Country 2

Fractional counting is also applied to data that are disaggregated by discipline. For example, a paper may be published in a journal that Elsevier and Clarivate Analytics have assigned to more than one discipline within their databases' native classifications. We have then applied fractional counting to attribute the proportions of the paper to the relevant disciplines of the native classifications, before summing these fractional counts to the Fields or Disciplines of the OECD's Fields of Science and Technology Classification, as described below.

Disciplines classification

Data disaggregated by scientific disciplines are presented using the OECD's Fields of Science and Technology (FOS) classification. Both Elsevier and Clarivate Analytics have concordances mapping their classification structures - Scopus' All Science Journal Classification (ASJC) and WoS' 'traditional' Subject Categories classification scheme - to the FOS structure which was used in this report. However, neither database's classification uses all of the disciplines in the FOS, nor does the FOS account for all categories in the databases' classifications.

The FOS disciplines 'nano-technology', 'environmental biotechnology', and 'industrial biotechnology' are not used in Scopus. For WoS data, the FOS disciplines 'health biotechnology' and 'other medical sciences' are not used, and neither database uses the 'agricultural biotechnology' discipline. These differences in mapping need to be considered when directly comparing data disaggregated by discipline between WoS and Scopus in that, for example, items which were mapped to 'nano-technology' in WoS will have been classified into a different FOS discipline for Scopus data. Further, items in attributed to the 'multidisciplinary' category of Scopus' ASJC have been excluded from presentation here as this category is not mapped to any discipline in the FOS classification. Also, a small proportion of items was not classified to any category in WoS and so was also excluded here as these items were not able to be attributed to an FOS discipline.

In both databases, publications can be classified to more than one discipline. As such, the publications counts have also been fractionalised based on the number of classifications they were assigned to within their native structure and then aggregated to the FOS categories. The field-normalisation of indicators occurs on the classifications assigned in WoS and Scopus, and data is then presented by FOS discipline.

Citation window

While counts of publications can be reliably calculated as early as the following year, a period of time must elapse during which publications are disseminated, read and accumulated citations before counts of the publications' citations can be calculated. As such, it is typical in bibliometric studies to analyse citations in a window of 3 to 5 years after the publishing year. Wang [10] determined that 3 years is required for most publications to reach their maximum number of citations per year, after which point the number of citations are likely representative of the publication's long-term impact. For this reason this report uses a 3 year citation window, which also ensures the relevancy of the data better than the longer 5 year window. As such, and data and indicators pertaining to citations include all citations received within the year the paper was published and the subsequent two years. Consequently, items published at the end of a year have a slight disadvantage in that they have slightly shorter window in which to accrue citations.

Self-citations

Self-citations can either be included or excluded from citations counts. Self-citations have been retained in the data for this report on the basis that, first, self-citation is a standard means of scientific communication and of building upon one's own previous body of work, and secondly that the patterns of self-citation are likely to be similar with fields so will not present an advantage or disadvantage due to differing citation practices after field normalisation [3].

Excellence Rate

The Excellence Rate identifies the percentage of a country's publications that were in the 10% most highly cited publications from each discipline and that could be considered of excellent quality on this basis. In this report, we employed the method described by Waltman and Schreiber [9] to calculate the 10% most frequently cited publications. Following this method, we identified the publications with citations above the 90th percentile, however there may be a number of publications with citations on the threshold of the 10th percentile which, if included, would exceed the 10% required. As a secondary step then, we proportionally assigned the publications on the 90th percentile threshold to achieve exactly the top 10%. When interpreting Excellence Rates, the expected rate is 10%. Values higher than 10% then indicate the country had a higher than expected percentage of publications in the subset of 'excellent' publications and thus better performance.

Scientific Regard

The Scientific Regard (SR) indicator shows whether a country's publications are cited more or less than average compared to other publications from the same journals. SR is calculated by comparing the observed number of citations for a country's corpus of papers to the number of citations those papers could have been expected to receive, i.e. the average citations of papers in the journals the country published in, and then the scale is transformed to range between -100 and 100. As such, an SR of 0 indicates the countries were cited at the average of the journals they published in, while values over 0 indicate the country's publications are cited more frequently than average, and values below 0 indicate a lower than average citation rate compared to other publications from the same journals. The SR value for a country is calculated as:

$$SR_k = 100\, tanh\, ln\, (OBS_k/EXP_k)$$

where OBS_k is the observed rate of citations of country k's publications, and EXP_k is the expected citation rate based on the average citation rate of the journals in which country k published.

International Alignment

International Alignment (IA) is a measure of the international visibility of the journals in which a country publishes, based on citations. The IA value for a country is calculated as:

$$IA_k = 100\, tanh\, ln\, (EXP_k/OBS_w)$$

where OBS_w is the observed number of citations of all publications in the world, and EXP_k is the expected number of citations for country k based on the number of citations received by the journals country k published in. As such, positive IA values indicate the country's papers were published in journals that were cited more frequently than average, and so reflect higher visibility and impact. Conversely, negative IA values indicate the country published in journals that were cited less frequently than the world average.

References

[1] Clarivate Analytics. *Web of Science Journal Evaluation Process and Selection Criteria*. n.d. URL: https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/.

[2] A. Baccini, G. De Nicolao and E. Petrovich. “Citation gaming induced by bibliometric evaluation: A country-level comparative analysis”. In: *PLOS ONE* 14.9 (Sept. 2019), pp. 1–16. DOI: 10.1371/journal.pone.0221212. URL: https://doi.org/10.1371/journal.pone.0221212.

[3] L. Bornmann et al. “Bibliometric Standards for Evaluating Research Institutes in the Natural Sciences”. In: *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Ed. by Cronin B. and C. R. Sugimoto. Cambridge: The MIT Press, 2014, pp. 201–223.

[4] European Commission. *Third European Report on Science and Technology Indicators*. Tech. rep. Brussels: European Commission, 2003.

[5] Elsevier. *Content policy and selection*. n.d. URL: https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection.

[6] S. Stahlschmidt and S. Hinze. “The dynamically changing publication universe as a reference point in national impact evaluation: a counterfactual case study on the Chinese publication growth”. In: *Frontiers in Research Metrics and Analytics* 3 (Oct. 2018), p. 30. DOI: 10.3389/frma.2018.00030. URL: https://www.frontiersin.org/articles/10.3389/frma.2018.00030/full.

[7] S. Stahlschmidt, D. Stephen and S. Hinze. “Performance and Structures of the German Science System. Studien zum deutschen Innovationssystem.” In: EFI, 2019. Chap. Studie 5-2019.

[8] L. Waltman and N. J. van Eck. “Field-normalized citation impact indicators and the choice of appropriate counting method”. In: *Journal of Informetrics* 9.4 (2015), pp. 872–894.

[9] L. Waltman and M. Schreiber. “On the calculation of percentile-based bibliometric indicators”. In: *Journal of the American Society for Information Science and Technology* 64.2 (2013), pp. 372–379. DOI: 10.1002/asi.22775.

[10] J Wang. “Citation time window choice for research impact evaluation”. In: *Scientometrics* 94.3 (2013), pp. 851–872. ISSN: 0138-9130. DOI: 10.1007/s11192-012-0775-9.