Dimity Stephen / Stephan Stahlschmidt / Paul Donner

KB Quality Assurance at the macro level: Comparing the current and previous WoS snapshot

Report on wos_b_2018 and wos_b_2019

Version: 20191029

DZHW
German Centre for Higher Education Research and Science Studies

Editor:
German Centre for Higher Education Research and Science Studies (DZHW) GmbH
Lange Laube 12 | 30159 Hannover | Germany | info@dzhw.eu |www.dzhw.eu
POB 2920 | 30029 Hannover | Germany
phone: +49 511 450670-0 | fax: +49 511 450670-960
Chairman of the Supervisory Board:
Ministerialdirigent Peter Greisler
Scientific Director:
Prof. Dr. Monika Jungbauer-Gans
Managing Director:
Karen Schlüter
Registration Court:
Amtsgericht Hannover | HRB 6489
VAT No.: DE291239300
October 2019

Contents

Motivation

The aim of the report is to identify any potential changes in data between or within database versions that may indicate quality issues. To do so it offers:

- a visual comparison
- between time-series over the last 10 years
- stemming from the current and previous KB database snapshot
- on several key indicators
- for national, sectoral and institutional entities.

The DZHW already conducts quality assurance testing at the micro-level for KB bibliometric databases before the tables enter the production environment. This testing is invaluable to ensuring tables and variables contain the expected content. This report supplements the current micro-level approach by examining changes at the macro-level - institutions, sectors, countries, disciplines - in key variables between the latest two iterations of the databases.

This report is not an exhaustive analysis of the databases' content, nor does it investigate any anomalies identified within the databases. However, this report probes the core variables fundamental to common bibliometric analyses, serves as an overview of the current state of the databases, and highlights changes that may indicate issues with data quality that warrant further investigation to understand or rectify. Changes may arise through several means. For instance, the database provider may add or remove journals from indices, change the discipline classification, or change how the classification is applied. The KB may identify new or decommissioned institutions, which can affect publication output for particular disciplines, or countries may implement policies regarding publication practices that can exert a substantial influence on the content published over time. This report aims to provide users of the KB databases with an overview of potential changes soon after the databases enter the production environment, allowing these factors to be considered in analyses.

Set of indicators

The indicators we have chosen reflect the core variables in the database that are fundamental to key bibliometric analyses and indicators. We provide context to the selection of variables and what information can be determined from their analysis in each of the following sections.

We make two sets of comparisons in this report. For indicators where it is important to consider trends over time, such as whole publication counts, we compare the databases for the 10 years up to the year for which both have complete data. For example, the latest common year with complete data for the wos_b_2018 and wos_b_2019 databases is 2017, as data for the absolute latest year in each database are incomplete. Similarly, where citation-based indicators are used, we present the time-series up to the latest common year with complete citation data, which is 2015 for the wos_b_2018 and wos_b_2019 databases. This comparison highlights any differences in trends between the databases for the most recent decade.

For other indicators, it is most useful to compare changes between just the most recent years of complete data in each database. For instance, we examine the threshold for Excellence Rates in 2015 from the wos_b_2018 database against 2016 in the wos_b_2019 database. Changes between the years are expected given we are comparing two different sets of publications, however this comparison can also provide insight into structural changes between the database iterations, such as the addition or removal of journals from indices, which may influence indicators at the macro-level.

Such comparisons are also helpful in identifying new or removed institutions or discipline categories. Further, although users will likely use the latest database to produce a complete time-series for new analyses, it is important to understand how additional years of a time-series might differ to existing time-series presented in publications and reports.

Set of entities

We have chosen to compare the databases at the national, sectoral, and institutional levels. The countries chosen are based on those most commonly examined by the DZHW due to their status as high-performing countries or as countries against which it is useful and informative to compare Germany.

We also examine the key German sectors: Universities (Uni), Fachhochschulen (FH), Max Planck Gesellschaft (MPG), Fraunhofer Gesellschaft (FHG), Helmholtz Gemeinschaft (HGF), Leibniz Gemeinschaft (WGL), the business sector (Econ), non-university hospitals (Klinik), and combined Ressortforschung-Bund and Ressortforschung-Länder (Gov). The remaining smaller sectors, such as research associations, clubs, and international and foreign organisations are grouped into an "other" category. Individual institutions are also examined, however only for Germany due to the unavailabilit of institutional coding for other countries. Further, given the large number of institutions, we present only the institutions that appear to have suddenly stopped or started publishing, and the larger institutions that have shown substantial changes in the indicator of interest.

Methodological details

Please note the following methodological details. First, we focus on articles and reviews published in journals as these are the most common documents used in bibliometric analyses. As previously noted, we supply a shortened time-series for citation-based indicators to allow for a 3-year citation window. Wang [2] determined that at least 3 years is required for 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. As such, citation-based indicators include all citations received within the publication year and the subsequent two years.

Whole counting is used throughout the report. Although it is most common to use fractional counting, analysing variables using whole counts will still reveal potential changes in the variables, negating the need to spend the additional time required to set up the necessary tables to perform fractional counting before this report can be run.

Data for disciplines are presented based on either the sc_traditional, sc_extended or Research Areas (RA) classification. Sc_traditional is the fine-grain classification more commonly used in analyses by the DZHW. However, as it contains over 250 categories, it is sometimes useful to use a coarse-grain approach to present an overview of the disciplines. As such, we present some data on the RA classification, which collapses the disciplines into five broad groups. RA are based on the sc_extended classification and so, as Clarivate only provides mapping between the RA and sc_extended classifications, supplementary tables presenting underlying data for RA are presented using the sc_extended classification. Each section containing data about disciplines notes which classification is used.

This report is automated and so tables are created regardless of whether any data fit the criteria, as such blank tables may appear in this report and are nonetheless informative about the indicator under examination.

Analysis

Whole count of articles and reviews: Selected countries and German sectors

The count of items produced by selected entities is the most fundamental bibliometric indicator. Given publication counts form the basis of many indicators, understanding the time-series trend within and between databases can inform expectations about potential changes that may arise in other indicators. In Figures 1 and 2 we present the whole counts of articles and reviews published in journals over the last 10 years by selected countries and sectors. Please note that the panels have different axes.

Changes in publication counts over time may reflect changes made by countries, the database provider, and/or administrative decisions. For example, it is expected that the wos_b_2019 database contains a higher number of publications for the most recent years than the wos_b_2018 database due to the continued indexing of items by Clarivate past the annual point in April at which data is cut to create the KB databases.

Increases in publications over time also result from both the continued growth of the national science systems and WoS' ongoing indexation over time, while sharp increases for a particular country may represent an actual increase in the number of a country's articles published in WoS-indexed journals, such as due to policy decisions, or reflect the recent indexing of a region- or country-specific journal. Decreases may reflect the de-indexation of a discipline-specific journal in which an entity commonly publishes or the stagnation of a sector, such as due to funding or policy decisions or the de-commissioning of an institution. Substantial deviations between databases – particularly in earlier years – or decreases in the current database in recent years may warrant investigation.

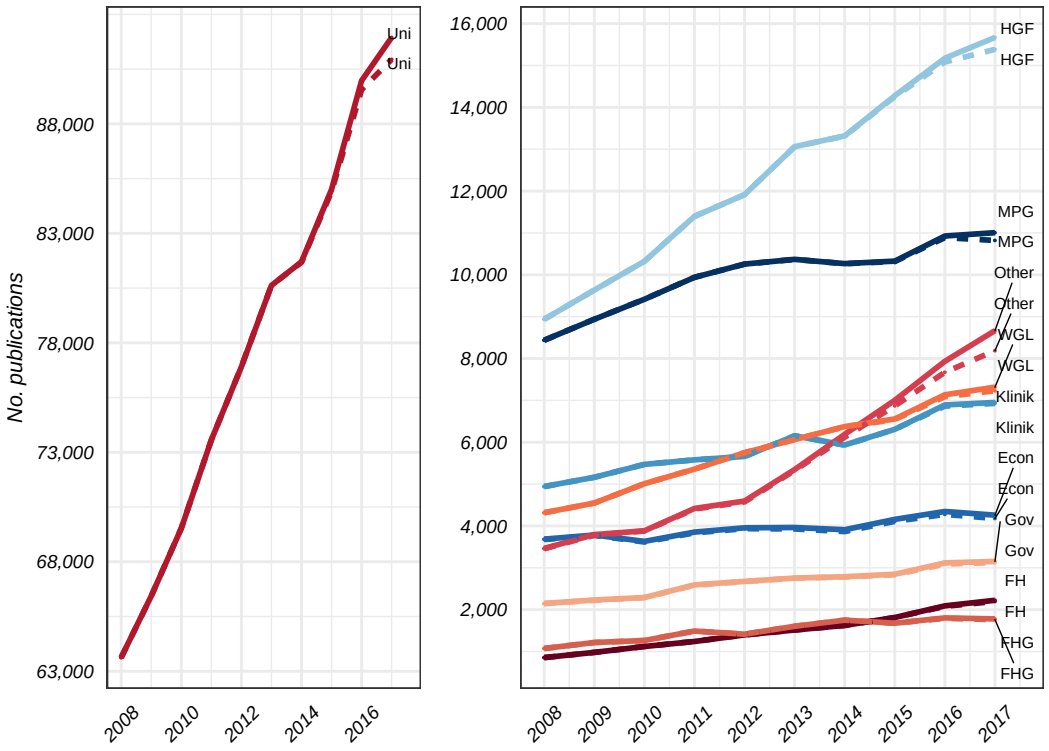

Figure 1: Whole counts of sectoral publications by database, where dashed lines show the previous database and full lines show the current database. Please note the panel's different scales.

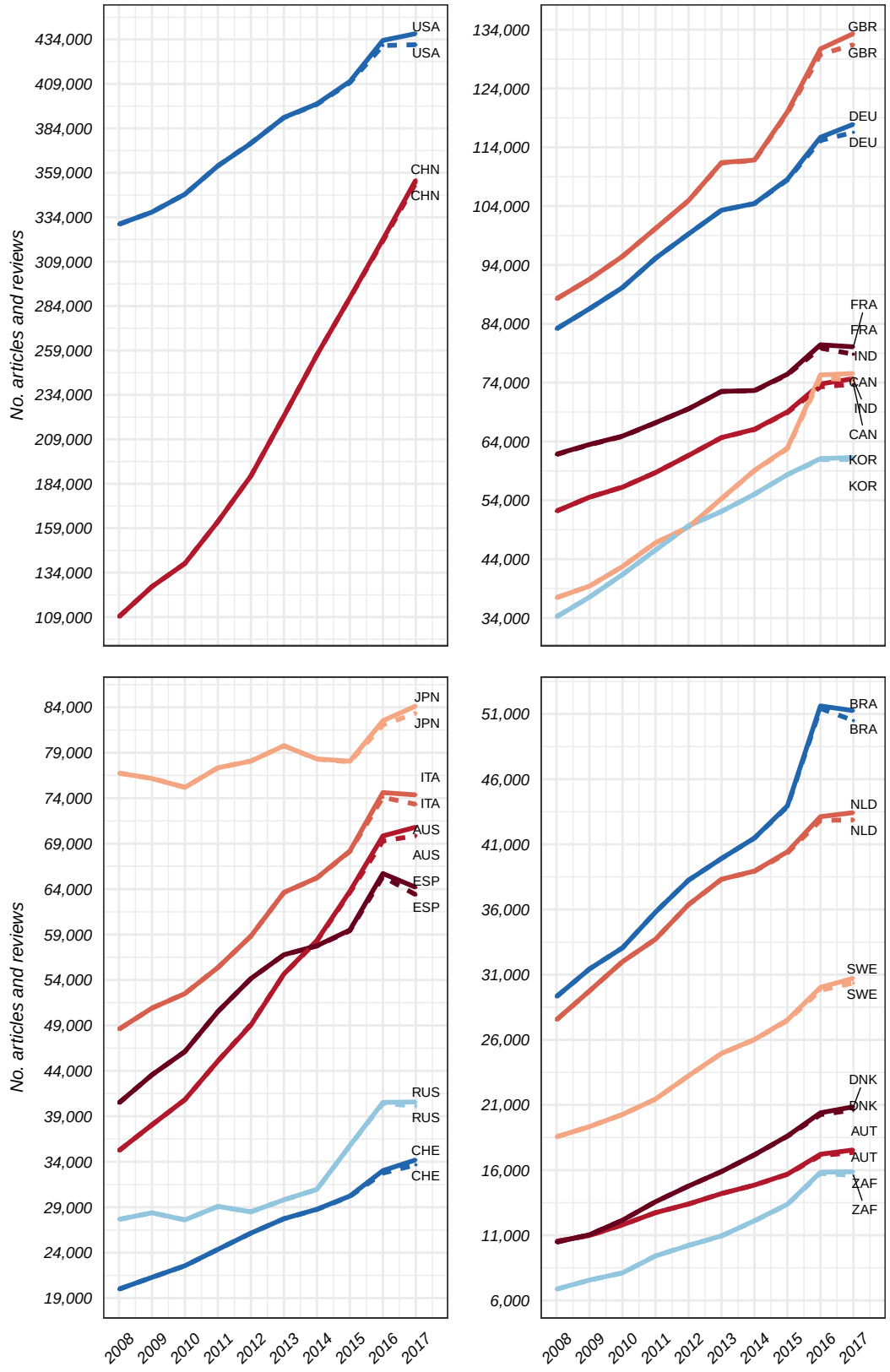

Figure 2: Whole counts of national publications by database, where dashed lines show the previous database and full lines show the current database. Please note the panels' different scales.

Excellence Rates: Selected countries and German sectors

Excellence Rates (ER) identify the percentage of an entity's publications that are in the 10% most highly cited publications from each discipline and could be considered of excellent quality on this basis. ERs are a common indicator used to assess an entity's performance, with an ER exceeding the expected 10% threshold interpreted as better than expected performance. ERs are calculated here based on the sc_traditional discipline classification. The ERs for the common years of the two databases up to 2015 are presented for German sectors in Figure 3 and for countries in Figure 4. As with whole counts of publications, we would expect general agreement between the databases, particularly in the earlier years of the time-series, so substantial deviations may warrant further analysis.

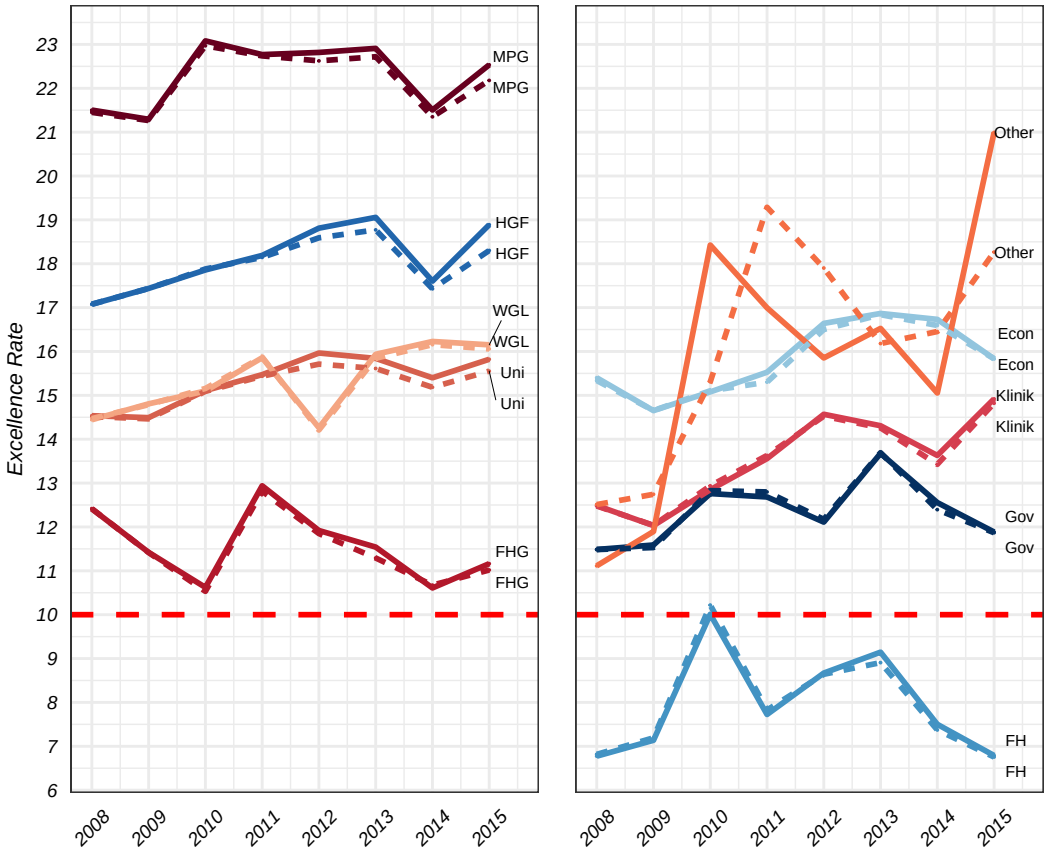

Figure 3: Excellence rates by sector, based on whole counts, where dashed lines show the previous database and full lines show the current database. The expected 10% threshold is shown in red.

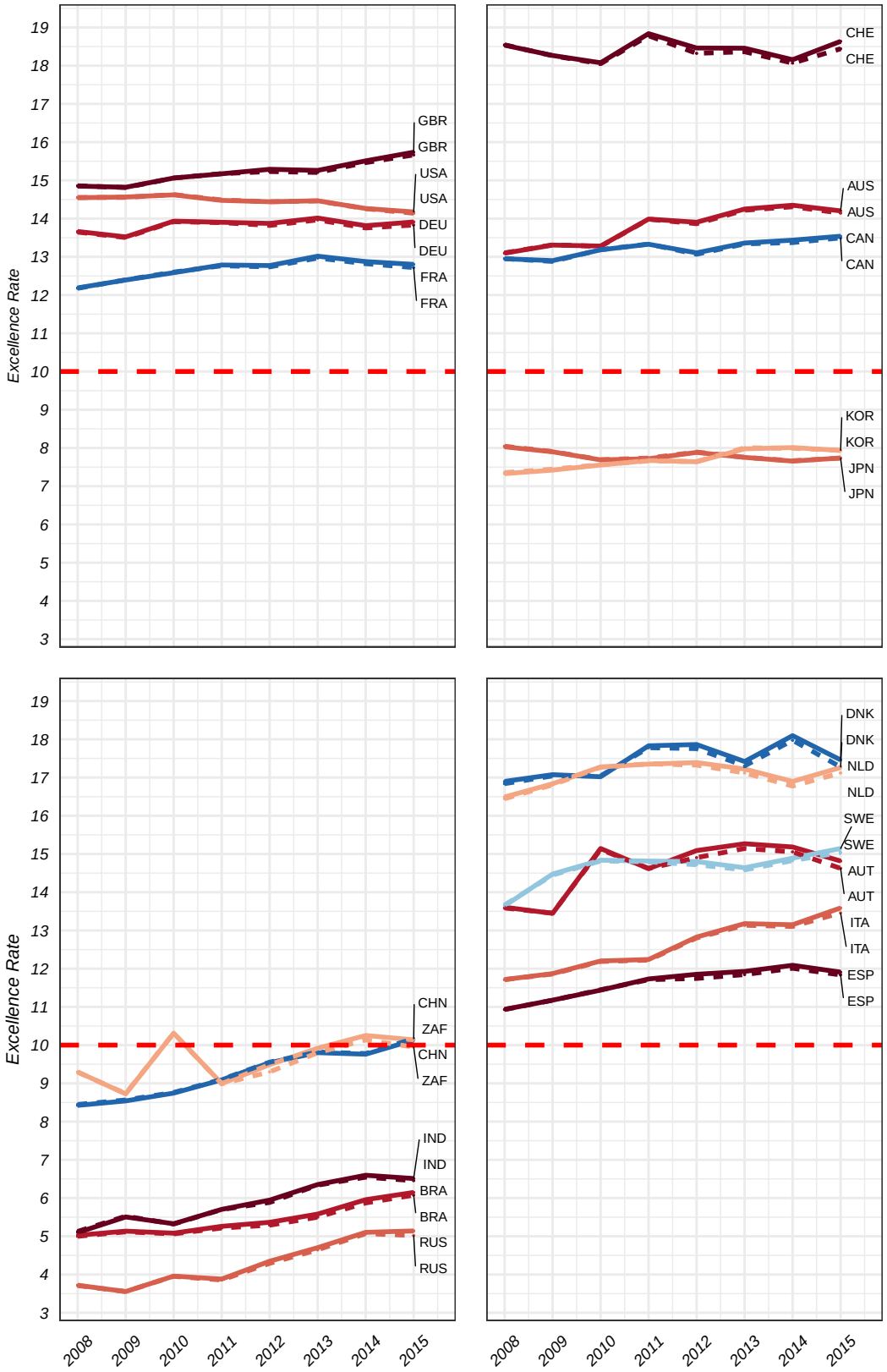

Figure 4: Excellence rates for selected countries, based on whole counts, where dashed lines show the previous database and full lines show the current database. The expected 10% is threshold shown in red.

Excellence Rates: Thresholds by discipline

ERs are dependent on the number of citations a publication receives in relation to the threshold it must exceed to reach the top 10% of the pool of reference publications. A change in the 10% threshold for a discipline can make it more or less difficult for a publication to exceed the threshold, which can have knock-on effects for a sector or country's ER over time. For example, substantial differences in countries' ERs between WoS and Scopus were observed in Stahlschmidt, Stephen and Hinze [1] due to the differences in coverage between the two databases, as Scopus' greater coverage of more sparsely cited journals lowers the ER threshold and allows high-performing countries to receive higher ERs.

While the higher internal consistency of coverage within WoS, compared to between WoS and Scopus, means we would expect less change in the ER thresholds between the iterations of the WoS databases, changes in the journals indexed may raise or lower the ER threshold for disciplines, potentially affecting the ERs of countries or, in particular, sectors due to their stronger disciplinary focus.

To identify changes in thresholds and assess whether changes in ERs for entities are likely, we present in Figure 5 the ER thresholds for articles and reviews in each sc_traditional discipline. We assess articles and reviews separately given the known differences in citation patterns between the document types.

In the top panels of Figure 5 we see the ER thresholds for each discipline in 2015 in both the wos_b_2018 and wos_b_2019 databases. The colour denotes the number of disciplines with each combination of thresholds, from fewer in blue to more in red. These panels depict the changes in ER thresholds in the same year between databases, providing context for any differences observed in 2015 in Figures 3 and 4.

In the bottom panels we present again the thresholds for each discipline in 2015 in the wos_b_2018 database but now compared against the threshold in 2016 in the wos_b_2019 database. These panels highlight changes between the latest years in each database, indicating whether we could expect to see changes in ERs between the databases.

The outlying disciplines with the greatest change in thresholds based on the bottom panels of Figure 5 are shown in Tables 1 and 2. Disciplines with a change in the ER threshold for articles of more than 20% between 2015 in wos_b_2018 and 2016 in wos_b_2019 are shown in Table 1, along with disciplines where the previous threshold was zero, highlighting potentially new or emerging disciplines.

For reviews, the disciplines with a change of more than 40% or that were previously zero are shown in Table 2. A higher change in thresholds is used for reviews as they tend to receive more citations than articles and the thresholds are more volatile. We also implement a cut-off of a threshold of at least 10 citations to reduce the number of disciplines with low thresholds presented, which are more susceptible to spurious changes due to their size. Data are based on the sc_traditional classification.

Large increases in the threshold would require publications to achieve substantially more citations to exceed the 10% threshold and be included in the ER, while a decrease in the threshold means publications require fewer citations than previously.

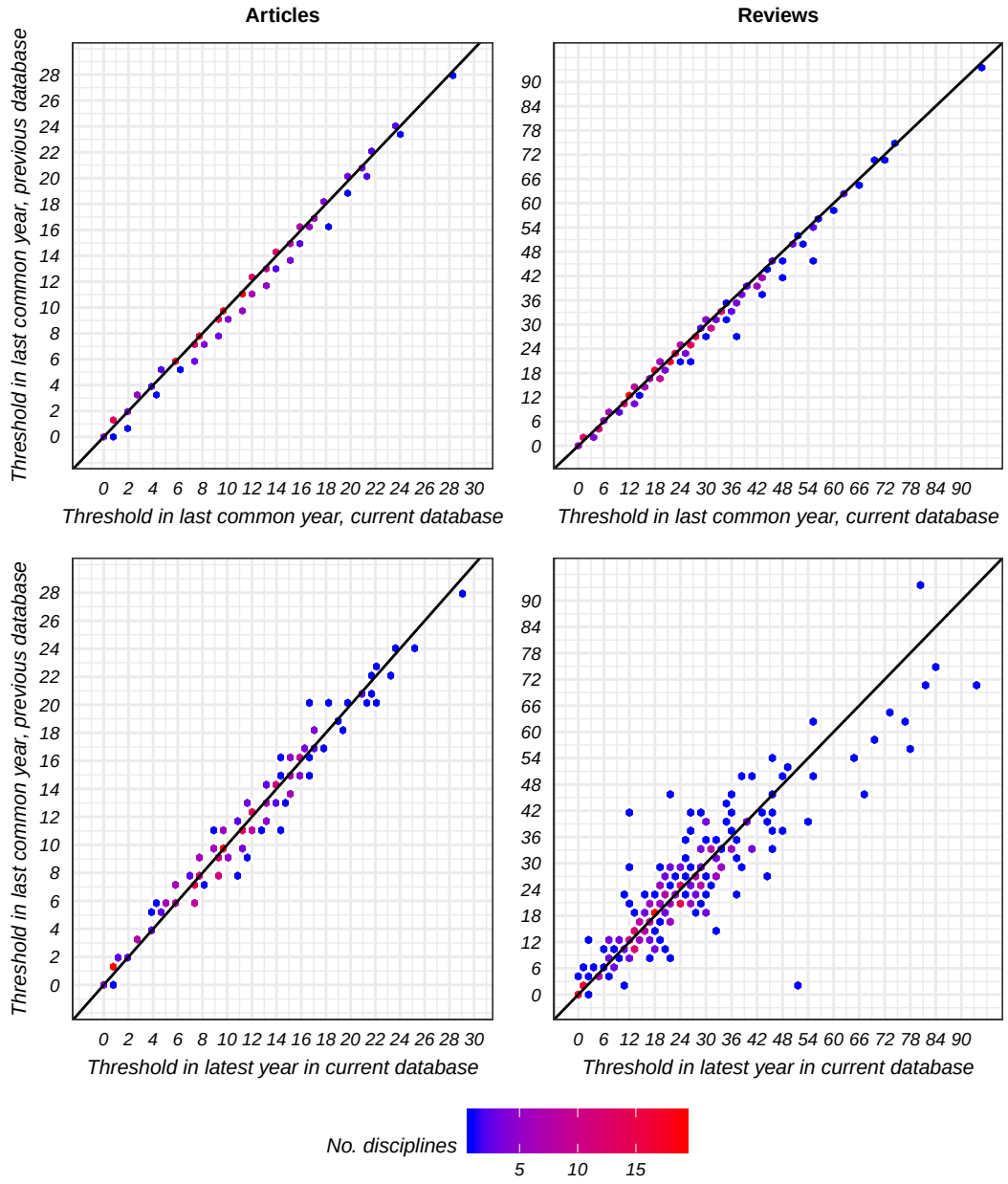

Figure 5: The ER threshold for articles and reviews in each discipline (sc_traditional) between databases, where colour denotes the number of disciplines with this combination of thresholds.

Table 1: Articles: Disciplines where the ER threshold changed by more than 20% between the latest year in each database, or the previous threshold was 0.

Discipline	Previous threshold	Current threshold	No. diff.	Perc. diff
Literary Theory & Criticism	0	1	1	Inf
Psychology, Mathematical	8	11	3	37.5
Planning & Development	9	12	3	33.3

Discipline	Previous threshold	Current threshold	No. diff.	Perc. diff
Transportation Science & Technology	11	14	3	27.3
Ethnic Studies	6	4	-2	-33.3
Film, Radio, Television	2	1	-1	-50.0
Religion	2	1	-1	-50.0

Table 2: Reviews: Disciplines with a current ER threshold of at least 10, where the threshold changed by more than 40% between the latest year in each database, or the previous threshold was 0.

Discipline	Previous threshold	Current threshold	No. diff.	Perc. diff
Classics	0	1	1	Inf
Film, Radio, Television	2	52	50	2500.0
Mathematics	3	10	7	233.3
Materials Science, Paper & Wood	8	22	14	175.0
Materials Science, Textiles	15	33	18	120.0
Statistics & Probability	11	20	9	81.8
Psychology, Educational	10	18	8	80.0
Industrial Relations & Labor	9	16	7	77.8
Materials Science, Ceramics	18	31	13	72.2
Physics, Particles & Fields	22	37	15	68.2
Psychology, Mathematical	18	30	12	66.7
Urban Studies	11	18	7	63.6
Physics, Nuclear	28	44	16	57.1
Engineering, Ocean	13	20	7	53.8
Education, Special	8	12	4	50.0
Mathematical & Computational Biology	18	27	9	50.0
Automation & Control Systems	20	29	9	45.0
Chemistry, Inorganic & Nuclear	46	66	20	43.5
Nanoscience & Nanotechnology	56	79	23	41.1
Social Sciences, Mathematical Methods	22	11	-11	-50.0
Limnology	46	22	-24	-52.2
Computer Science, Cybernetics	28	12	-16	-57.1
Demography	41	12	-29	-70.7

Citations: Mean 3-year citations of articles and reviews by discipline

The number of citations a publication could be expected to receive is dependent on the discipline from which it originates. As such, we examine here the mean 3-year citations of articles and reviews

by discipline. Mean 3-year citations (MC3) are the mean citations publications in each discipline accrue in the first 3 years after publication. As we did with ERs, we compare here the last common year in both databases to assess the retroactive effects stemming from changes made in the latest database, and the latest complete year in both databases to assess potential structural changes and updates to the time-series. Data are based on the sc_traditional discipline classification.

The top panels of Figure 6 show the MC3 for each discipline in 2015 in both the wos_b_2018 and the wos_b_2019 databases for articles and reviews respectively. The bottom panels show the MC3 in 2015 in the wos_b_2018 database and 2016 in the wos_b_2019 database. A greater deviation of disciplines from the central line indicates a greater degree of change in the mean citations of a discipline's items between years.

The outlying disciplines from the bottom panels of Figure 6 are shown in Tables 3 and 4. We include in Table 3 the disciplines with a greater than 20% change in mean citations of articles between 2015 in the wos_b_2018 database and the 2016 in the wos_b_2019 database, and also disciplines where the previous threshold was zero.

As previously noted, citations for reviews tend to be higher and more volatile than articles. As such, we show in Table 4 the disciplines with a greater than 40% change in mean citations between 2015 in the wos_b_2018 and wos_b_2019 databases, and also disciplines where the previous threshold was zero.

Disciplines with low MC3s require only a small change to exceed the threshold of change to be reported. As such, we have included a threshold of a current MC3 of at least 1 for articles and 3 for reviews to remove disciplines with spurious changes due to their low level of citations.

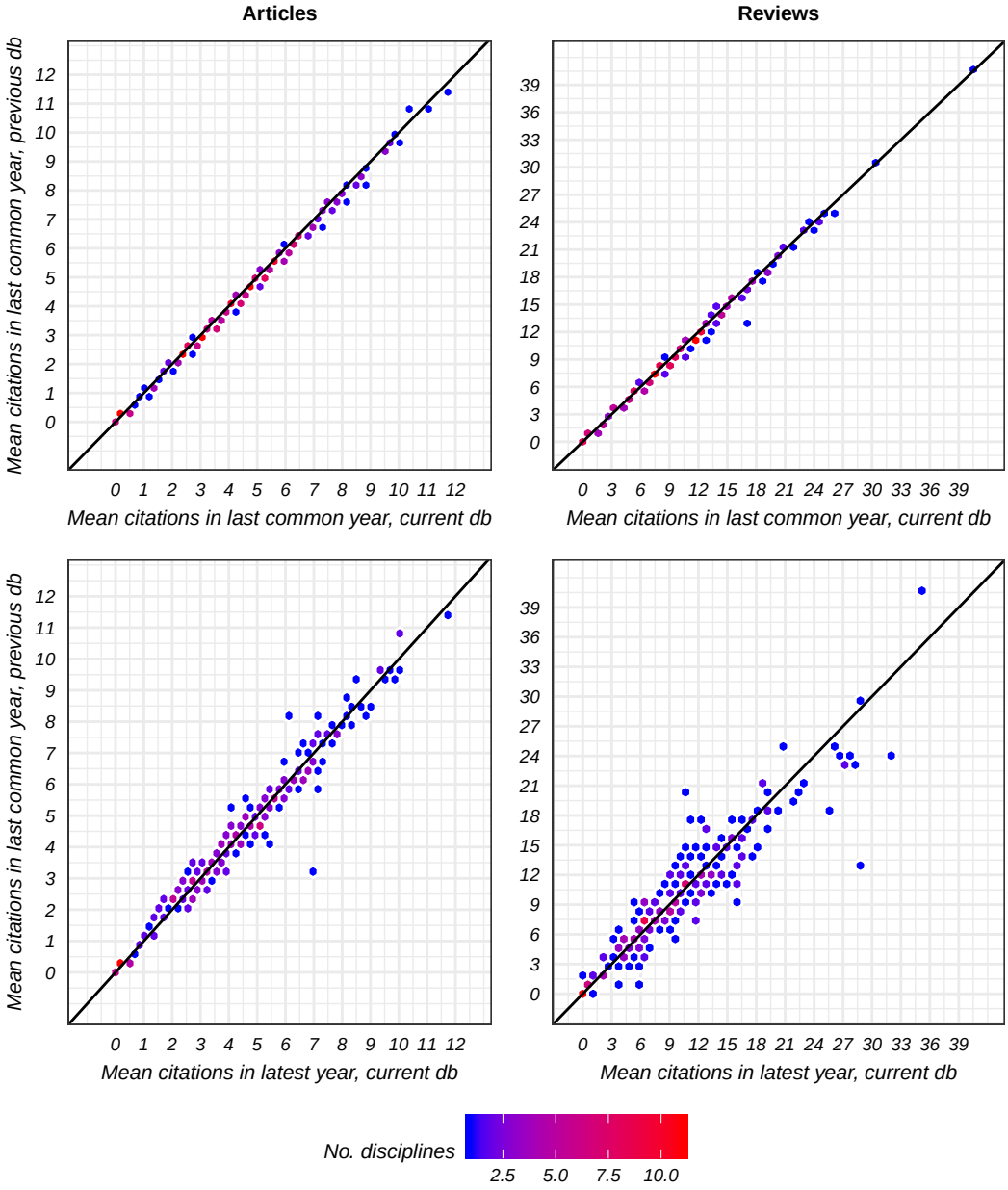

Figure 6: The MC3 for articles and reviews in each discipline between databases, where colour denotes the number of disciplines with this combination of citations.

Table 3: Articles: Disciplines with a current MC3 of at least 1, where the MC3 changed by more than 20% between the latest year in each database, or the previous MC3 was 0.

Discipline	Previous cit.	Current cit.	No. diff.	Perc. diff.
Planning & Development	3.3	6.8	3.4	103.3
Psychology, Mathematical	4.0	5.3	1.3	33.1
Engineering, Petroleum	2.1	2.6	0.5	22.9
Andrology	3.3	4.0	0.7	21.6

Discipline	Previous cit.	Current cit.	No. diff.	Perc. diff.
Transportation	3.3	4.0	0.7	20.1
Demography	2.7	2.1	-0.6	-21.5
Social Sciences, Interdisciplinary	2.2	1.7	-0.5	-21.6
Education & Educational Research	2.0	1.6	-0.5	-23.3
Medicine, General & Internal	8.1	6.1	-2.0	-25.0
Law	1.7	1.2	-0.5	-28.1
Ethnic Studies	2.4	1.7	-0.7	-28.8

Table 4: Reviews: Disciplines with a current MC3 of at least 3, where the MC3 changed by more than 40% between the latest year in each database, or the previous MC3 was 0.

Discipline	Previous cit.	Current cit.	No. diff.	Perc. diff.
Classics	0.0	0.3	0.3	Inf
Film, Radio, Television	1.0	5.8	4.8	477.8
Mathematics	1.1	3.3	2.2	190.2
Ethics	2.4	5.2	2.9	122.0
Imaging Science & Photographic Technology	13.5	28.7	15.2	112.3
Physics, Particles & Fields	9.5	16.2	6.7	71.2
Materials Science, Paper & Wood	3.1	5.4	2.2	71.0
Physics, Mathematical	4.0	6.9	2.8	70.5
Psychology, Educational	5.8	9.3	3.6	62.0
Mathematical & Computational Biology	7.2	11.5	4.3	59.5
Industrial Relations & Labor	3.4	5.4	1.9	57.0
Engineering, Petroleum	3.8	5.9	2.1	55.9
Gerontology	7.9	11.8	3.9	49.3
Materials Science, Textiles	4.6	6.7	2.1	45.8
Optics	18.3	26.1	7.8	42.7
Mathematics, Applied	6.5	3.9	-2.6	-40.4
Limnology	19.9	10.6	-9.4	-47.0

Uncited articles and reviews: Percent by selected countries and German sectors

While ERs represent the most highly cited publications and mean citations tell us about what's average, the percentage of uncited publications can tell us about the entities at the tail end of the citation distribution. When examining uncited publications, we expect to see a decreasing trend in uncited publications over time. This occurs because citation counts are based on the items indexed in each database and so, as Clarivate continues to index journals, it increases the likelihood that any publication will have been cited by the indexed items. In particular, we would expect that the percentage of uncited publications in the last common year would be lower in the current database than the previous database, as data added in the latest iteration "complete" the incomplete last year of the previous database. An increase in uncited publications in the latest year may reflect processing issues that require investigation.

We present in Figures 7 and 8 the percentage of articles and reviews per German sector and selected country that remained uncited 3 years after they were published.

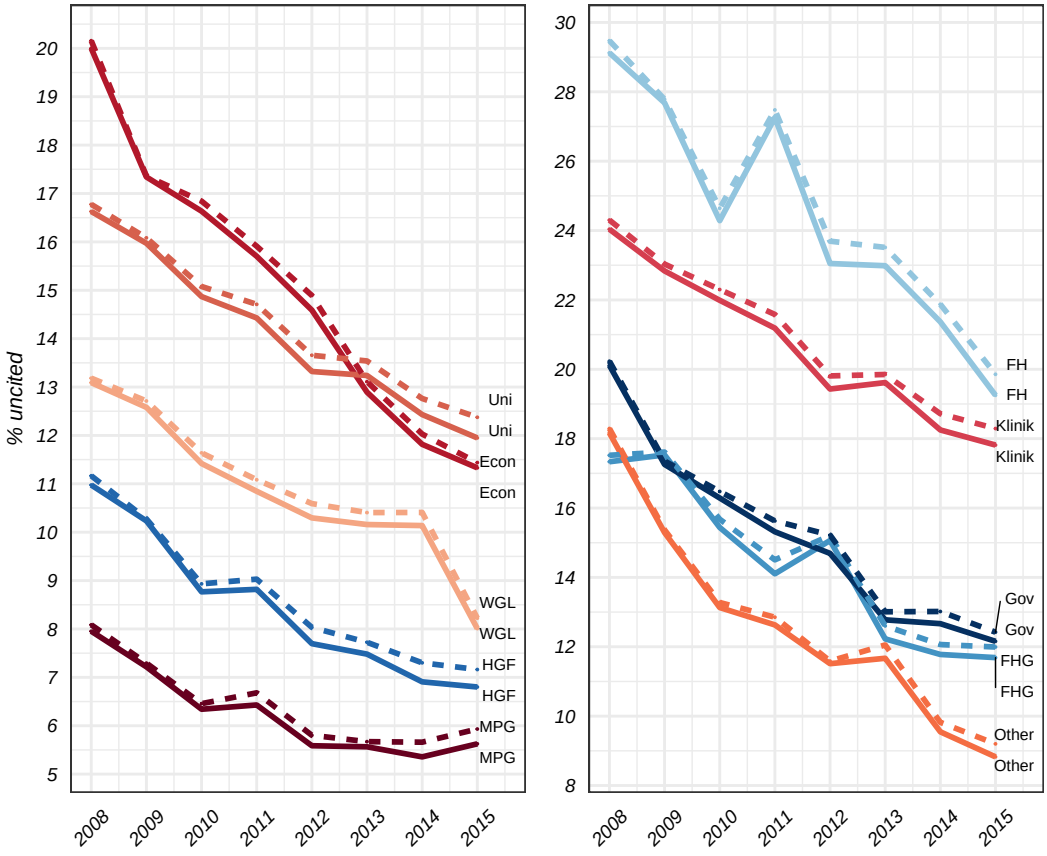

Figure 7: The percentage of uncited publications by German sector, based on whole counts, where dashed lines show the previous database and full lines show the current database.

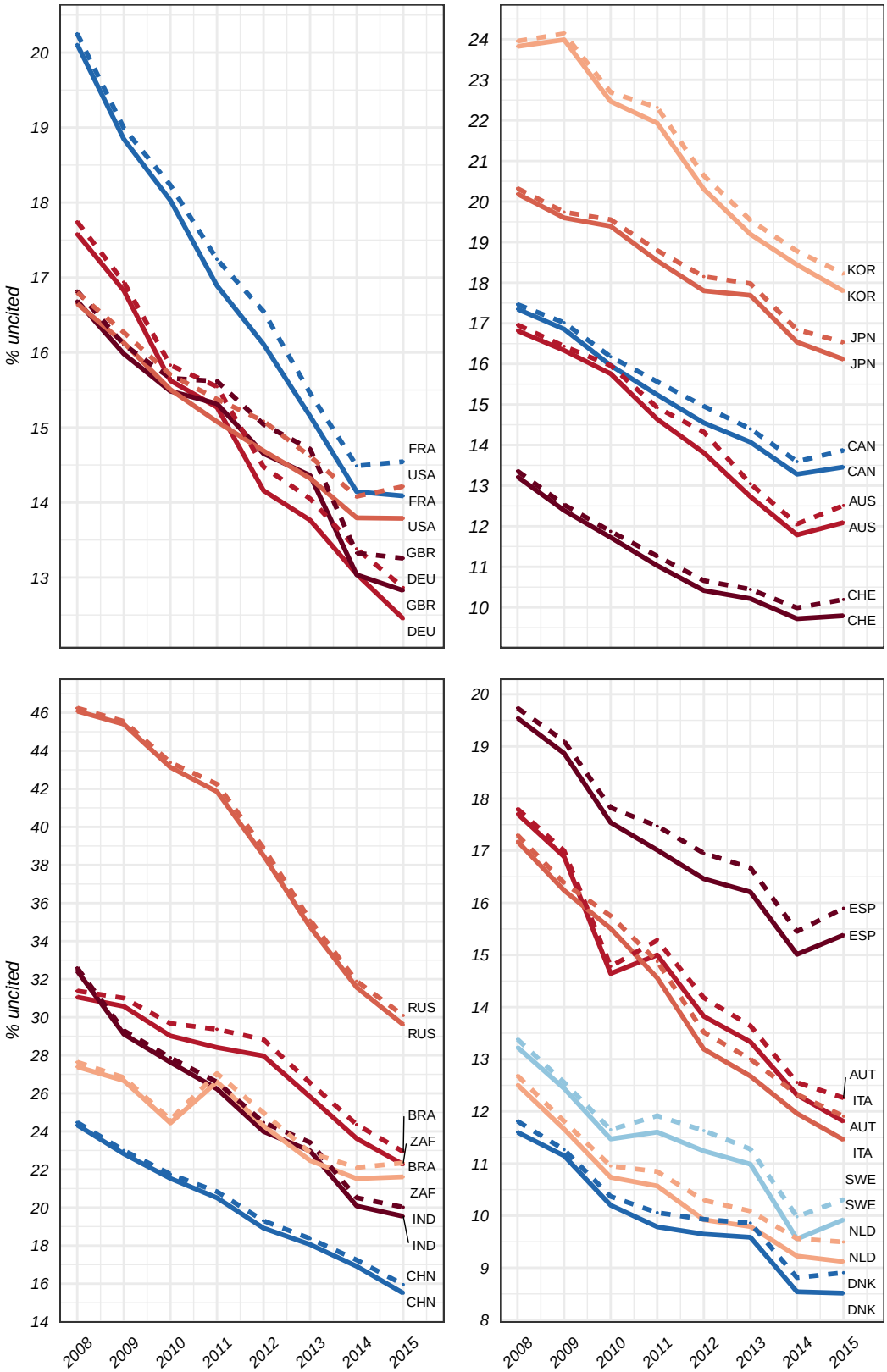

Figure 8: The percentage of uncited publications by selected countries, based on whole counts, where dashed lines show the previous database and full lines show the current database.

Disciplines: Changes in discipline classification

The two tables in this section highlight simply whether any changes have been made to WoS' discipline classifications, the sc_traditional and sc_extended. This could include splits, aggregations or removals of a discipline, or the inclusion of a new discipline to reflect new and emerging topics. Here we identify changes in the classification structure by comparing the number of articles and reviews attributed to each discipline in the latest year of each database and selecting those disciplines where the number was zero in one year but not in the other.

Disciplines with no prior publications but some in the current year suggest the discipline may have been recently added, while the opposite suggests the discipline may have been removed or merged. Changes may also reflect changes in spelling or punctuation of the discipline name. Any changes should be checked with WoS' published classification structure. Changes in the structure of the sc_traditional classification are shown in Table 5 and changes in the sc_extended classification in Table 6.

Table 5: Changes in the sc_traditional discipline classification structure between the previous and current databases.

Classification	Previous pubs	Current pubs
Development Studies	NA	695
Green & Sustainable Science & Technology	NA	6,205
Quantum Science & Technology	NA	901
Regional & Urban Planning	NA	3,223

Table 6: Changes in the sc_extended discipline classification structure between the previous and current databases.

Classification	Previous pubs	Current pubs
Development Studies	NA	695

Disciplines: Changes in articles and reviews by discipline

This section identifies the disciplines that had a substantial change in the number of publications assigned to them between the latest years in each database. Changes in counts of publications per discipline reflect changes in the journals indexed, the classification structure, and any potential processing issues. As such, any large changes shown here may be worth examining.

We show in Figure 9 the 20 disciplines with the highest percentage increases and decreases in publication counts between 2017 in wos_b_2018 and 2018 in wos_b_2019. The number shown next to each bar is the numerical change in publication counts. We have used whole counting and the disciplines are based on the sc_traditional classification. Disciplines previously identified as being new or removed have not been included here.

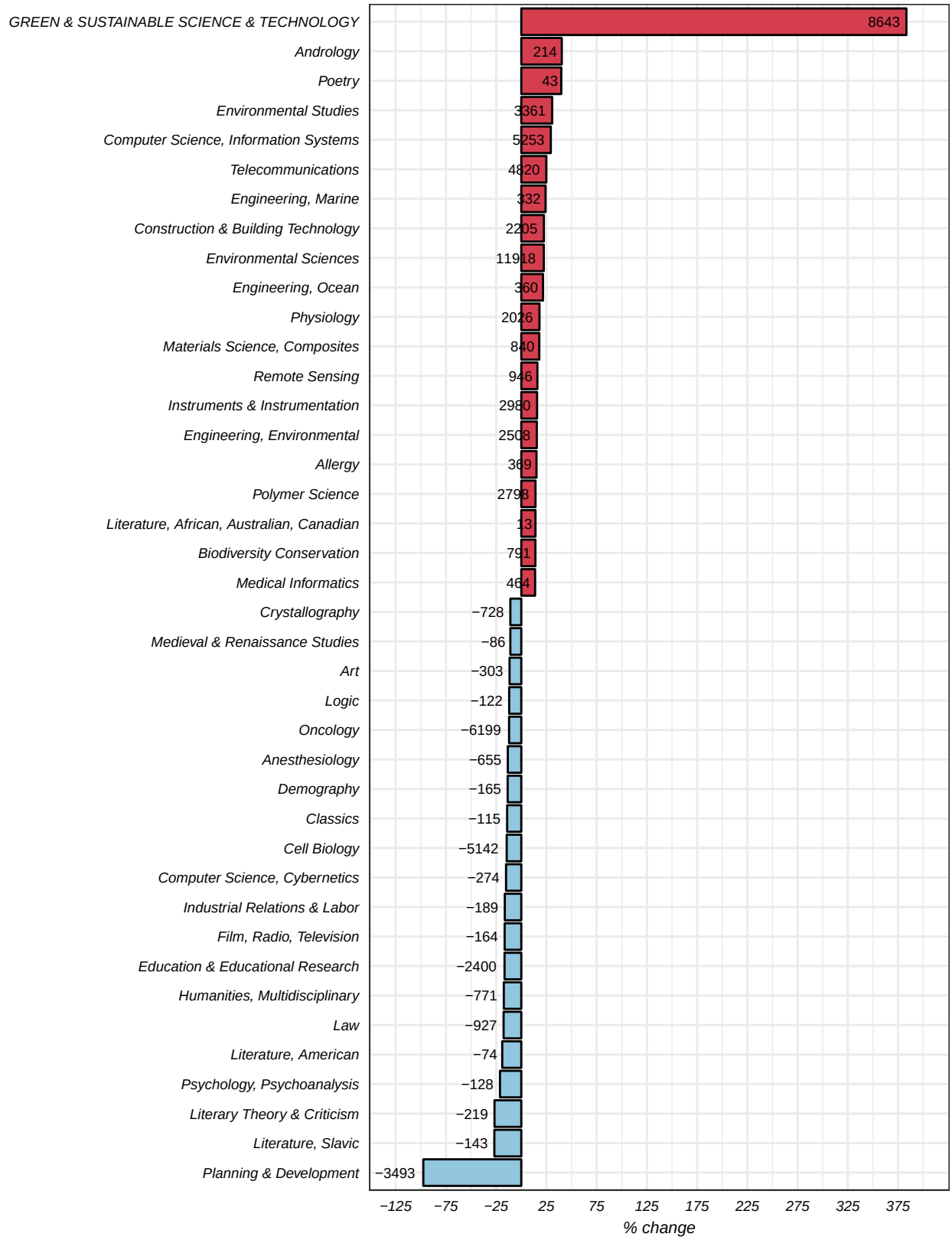

Figure 9: The 40 disciplines with the highest percentage change in publication counts between the latest complete years in each database, with numerical difference in counts.

Metadata: Changes in pubyear, doctype and pubtype

This section details the number of items for which changes were made to key metadata in the latest iteration of the database. We look at changes in the recorded publication year, document type and publication type as these three variables are typically the key inclusion criteria for bibliometric analyses. A change in metadata for a large number of items may be problematic, particularly if the changes are not randomly distributed, such as adjustments having been made to items from a particular journal or set of publications, which may affect counts and indicators for specific entities. Some changes can be expected as Clarivate updates or corrects items, however a large number of items or a change in a time-series may require investigation.

We identify changes in the metadata of in-scope items by first matching items between the wos_b_2018 and wos_b_2019 databases using the UT_EID identifier and then counting the number of instances where matched items do not have the same publication year, document type (i.e. an article or review has been changed to a different document type) or publication type (i.e. the publication type changed from journal to another type) between databases. As such, Table 7 shows the number of items that have had their metadata changed between the previous and current databases. Data are presented based on the publication year recorded in the previous database.

Table 7: The number of items with changes in metadata between the previous and current database versions.

Year	Pub. year	Doc. type	Pub. type
2007	0	65	0
2008	0	88	0
2009	0	3	0
2010	11	112	0
2011	8	129	0
2012	37	175	0
2013	44	242	0
2014	25	452	0
2015	48	2,074	0
2016	221	873	0

Institution and country data: Number of articles and reviews with missing data

Bibliometric analyses often examine indicators at the level of institutions or countries. Further, fractional counting can be applied based on institutions, with articles apportioned according to authors' affiliations. As such, it is imperative for accurate indicators that most, if not all, items have institution and country data, as missing information removes otherwise valid items from analyses.

The Items table of the bibliometric databases holds a record of all available items, while the associated data about authors' affiliations are held, in part, in the Institutions table. We have operationalised missing institution information here as publications that appear in the Items table but have no corresponding information in the Institutions table. We present in the top panel of Figure 10 the number of items in each database between 2008 and 2017 with no institution information. Additionally, items can have institution information but no country code – from which country counts are derived – and these are shown in the bottom panel of Figure 10. Large disparities between the databases or substantial increases in missing information should be investigated.

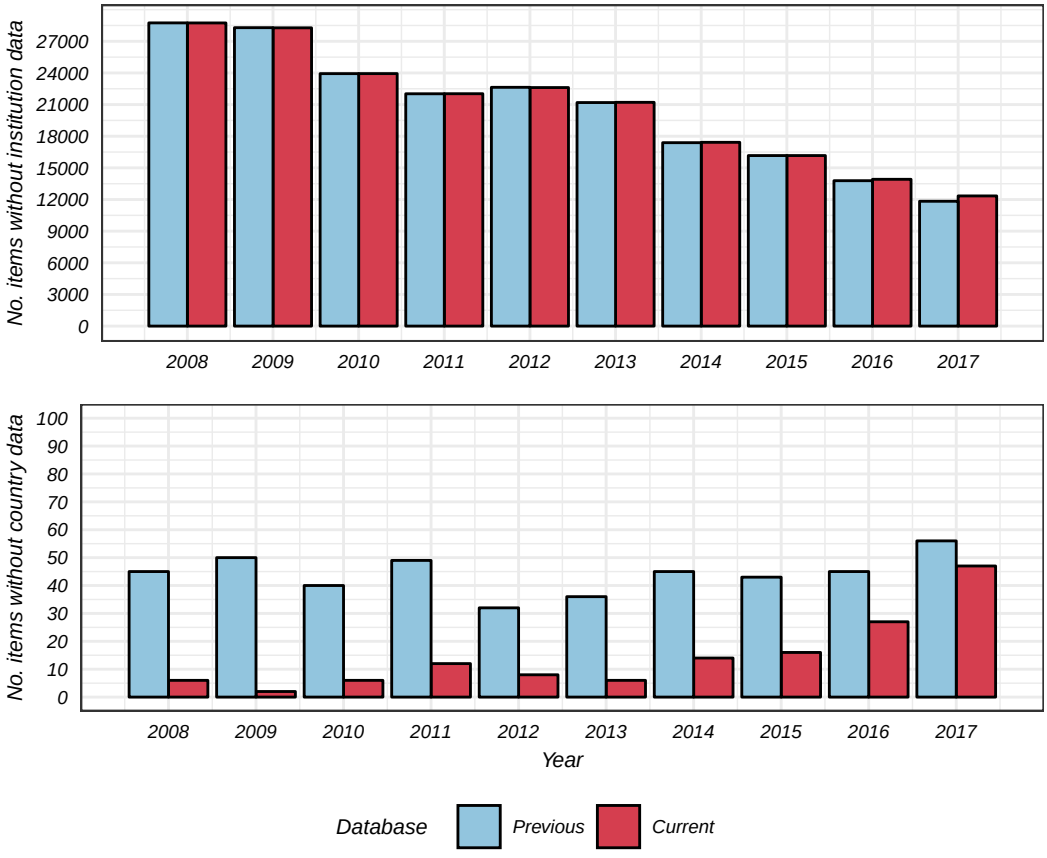

Figure 10: The number of items with missing institution information (top) and the additional items that have institution information but no country code (bottom) over time by database.

Author-institution links: Percentage complete by Research Area and discipline

Similarly to ensuring that all or most items have institution and country information, it is important for allocating publications to entities that authors' affiliations with institutions have been assigned for the majority, or ideally all, items. As such, we examine here the percentage of items in each sc_extended discipline with complete links between authors and institutions.

In Figure 11, we see in the left panel the percentage of complete links for 2017 data in both the previous and current databases, highlighting any retroactive changes that may have been made in the current database. In the right panel is again the percentage of complete links made in 2017 in the wos_b_2018, now compared with the 2018 in the wos_b_2019, indicating potential changes between the latest year in each database.

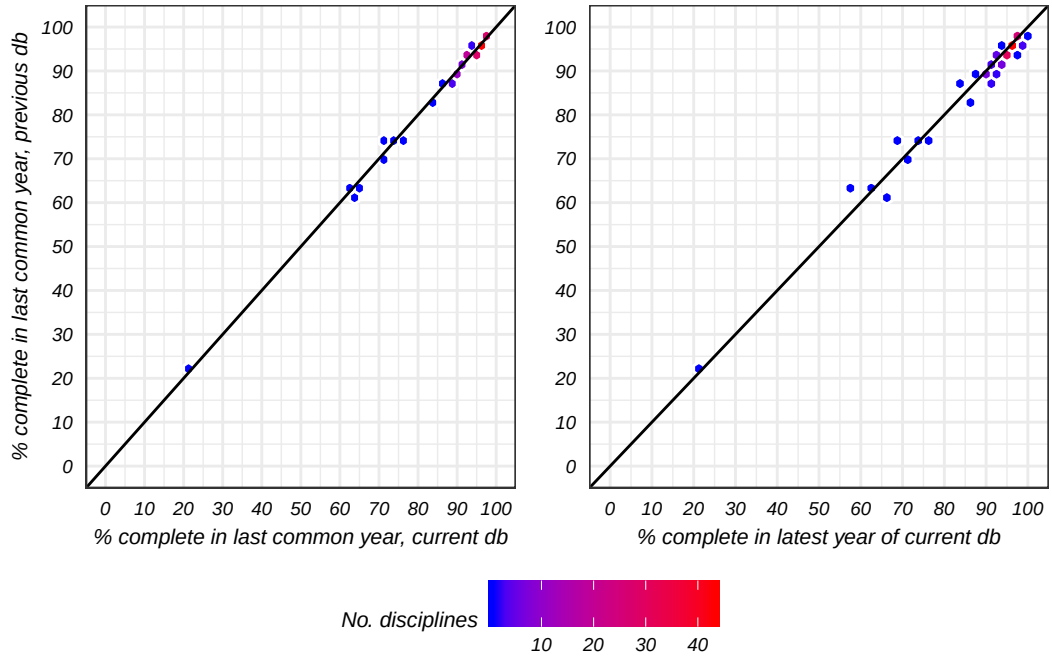

Figure 11: The percentage of complete author-institution links by disciplines (sc_extended).

The outlying disciplines observed in the right panel of Figure 11 that have a change of more than 5 percentage points in the percentage of complete author-institution links between databases are shown in Table 8.

Table 8: Disciplines (sc_extended) with a change of more than 5 percentage points in missing links between latest year in current database and last common year in previous database.

Discipline	Prvs items	% prvs complete	Crrnt items	% crrnt complete	Change
Film, Radio & Television	1,083	64.5	935	57.6	-6.9
Art	2,666	74.6	2,386	68.0	-6.6

To provide context to the percentage of complete links observed in the most recent years, in Figure 12 we present the percentage of complete links made between authors and affiliations in each Research Area over the last ten common years in both databases. Substantial changes between years or differences between the databases may require investigation of the cause.

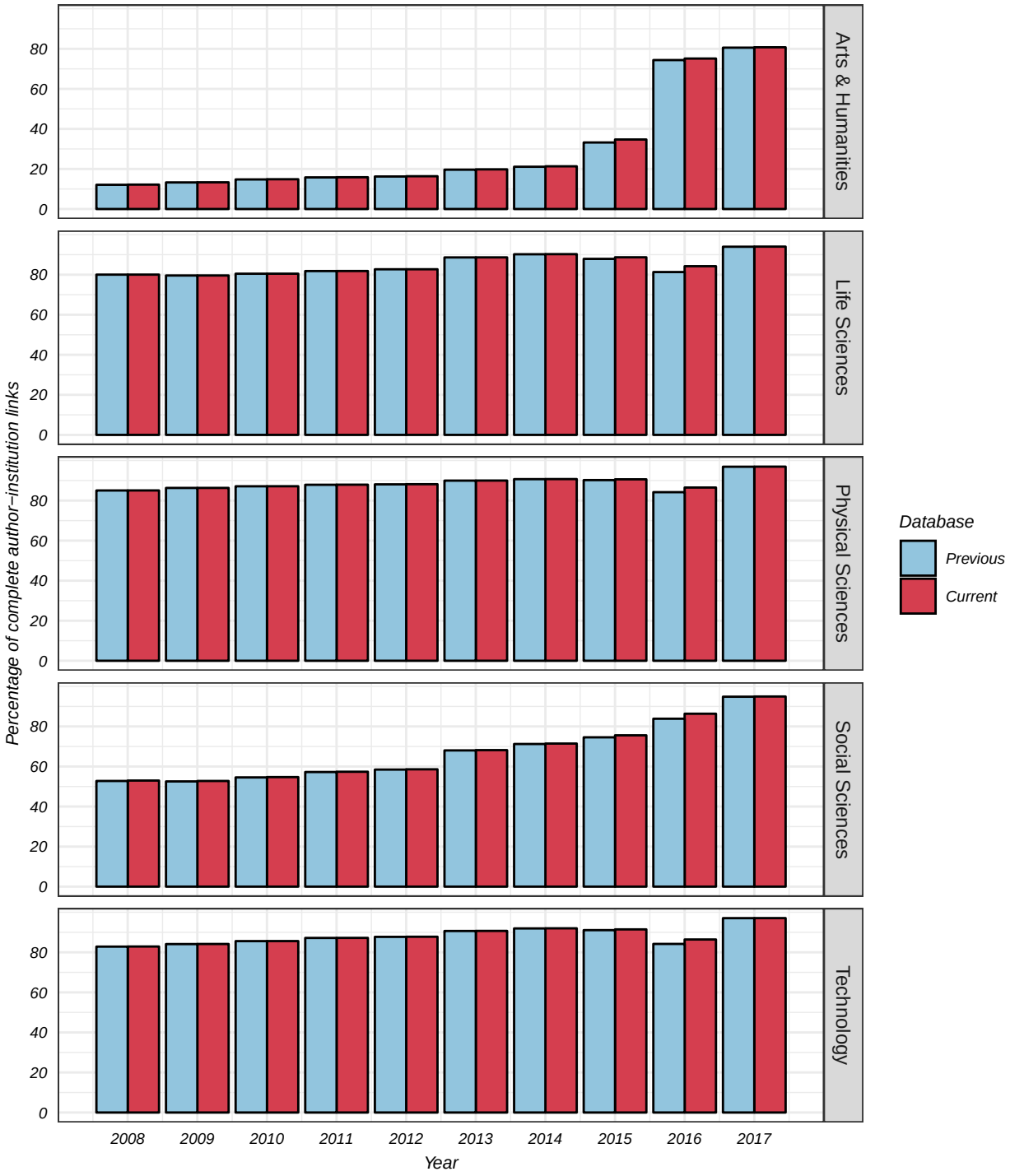

Figure 12: The percentage of complete author-institution links by Research Area and database over time.

German institutions: Changes in whole counts of articles and reviews

This section compares changes in the number of articles and reviews published by German institutions between the latest years available in each database. These tables can assist in identifying institutions for which substantial numbers of publications have been added, removed or otherwise changed in the latest database. They can also aid in assessing the degree of change in publication numbers for larger institutions, which may require further examination if considered unusual or excessive.

Table 9 presents potentially new institutions – these had no publications in 2017 in the wos_b_2018 database but more than five publications in 2018 in the wos_b_2019 database. Conversely, Table 10 shows the institutions that had at least five publications in 2017 in the wos_b_2018 database but no publications recorded in 2018 in the wos_b_2019 database. We also highlight in Tables 11 and 12 the larger institutions (with at least 20 publications) that had a change in publication counts of more than 40% between 2017 and 2018 in the wos_b_2018 and wos_b_2019 databases.

Table 9: Institutions with more than 5 publications in the latest year of the current database that had no publications in the last common year in the previous database.

PK_KB_INST	Name	Previous pubs	Current pubs
5348	Deutsches Zentrum fur Lungenforschung	0	341
5368	Translational Lung Research Center Heidelberg	0	44
5352	Psychologische Hochschule Berlin	0	23
1056	Max-Planck-Institut fur Innovation und Wettbewerb	0	19
5371	HI-STEM gGmbH	0	17
5355	CCC Comprehensive Cancer Center	0	12
925	Kriminologisches Forschungsinstitut Niedersachsen e. V.	0	11
5346	ADVA Optical Networking	0	9
5347	Max Planck UCL Centre for Computational Psychiatry and Ageing Research	0	9
5354	Forschergruppe Diabetes e.V.	0	9
708	Institut fur Mikroelektronik Stuttgart	0	8
5369	Piramal Imaging GmbH	0	8
1483	Ford-Werke GmbH	0	7
1012	Max Planck Computing and Data Facility	0	6
5349	DKMS	0	6
5350	DKMS Life Science Lab	0	6

Table 10: Institutions with no publications in the latest year of the current database that had more than 5 publications in the last common year in the previous database.

PK_KB_INST	Name	Previous pubs	Current pubs
250	Europaische Kommission	26	0
1328	Nokia Siemens Networks GmbH & Co. KG	9	0
4697	Deutsches Zentrum fur Herz-Kreislauf-Forschung e. V.	17	0

Table 11: Institutions with more than 20 publications in the last common year in the previous database that had an increase in publication counts of more than 40% in the latest year in the current database.

PK_KB_INST	Name	Previous pubs	Current pubs	No. diff.	Perc. diff.
5210	Berliner Institut fur Gesundheitsforschung	63	231	168	266.7
5290	Deutsches Zentrum fur Infektionsforschung	107	309	202	188.8
1606	Bayer Konzern	86	242	156	181.4
4758	Max-Planck-Institut fur empirische Asthetik	22	47	25	113.6
1637	Zentrum fur Rhinologie und Allergologie	24	46	22	91.7
756	Forschungszentrum caesar	30	55	25	83.3
1766	Universitatsklinikum Gie?en und Marburg GmbH - UGKM	277	480	203	73.3
656	Beuth Hochschule fur Technik Berlin	33	57	24	72.7
575	Hochschule Niederrhein	32	54	22	68.8
28	Leibniz-Institut fur Analytische Wissenschaften - ISAS - e.V.	57	95	38	66.7
48	Leibniz-Institut fur Atmospharenphysik e.V. an der Universitat Rostock (IAP)	25	41	16	64.0
856	Bundesanstalt fur Arbeitsschutz und Arbeitsmedizin	25	39	14	56.0
1133	Fraunhofer-Institut fur Techno- und Wirtschaftsmathematik	43	67	24	55.8
670	Fachhochschule Aachen	38	58	20	52.6
998	Bayerische Akademie der Wissenschaften	42	64	22	52.4
1053	Max-Planck-Institut fur Herz- und Lungenforschung (W. G. Kerckhoff-Institut)	84	128	44	52.4
586	Hochschule Magdeburg-Stendal	23	35	12	52.2

PK_KB_INST	Name	Previous pubs	Current pubs	No. diff.	Perc. diff.
1047	Max-Planck-Institut fur Mathematik	72	109	37	51.4
913	Bayerisches Landesamt fur Gesundheit und Lebensmittelsicherheit	61	92	31	50.8
1075	Max-Planck-Institut fur Menschheitsgeschichte	57	85	28	49.1
623	Technische Hochschule Mittelhessen - THM	44	65	21	47.7
477	Klinikum Darmstadt	21	31	10	47.6
30	Leibniz-Institut fur die Padagogik der Naturwissenschaften und Mathematik (IPN)	34	50	16	47.1
4625	Institut fur Molekulare Biologie gGmbH	47	69	22	46.8
523	Klinikum Bayreuth GmbH	33	48	15	45.5
93	Hochschule Reutlingen, Hochschule fur Technik- Wirtschaft-Informatik-Design	31	45	14	45.2
503	Evangelisches Krankenhaus Bielefeld gGmbH	52	75	23	44.2
752	DECHEMA Gesellschaft fur Chemische Technik und Biotechnologie e.V.	28	40	12	42.9
1202	VOLKSWAGEN AG	26	37	11	42.3
4428	Restkategorie Universitaten, Kunst- und MusikhochschulenHochschulen	45	64	19	42.2
654	Fachhochschule Bielefeld	24	34	10	41.7

Table 12: Institutions with more than 20 publications in the last common year in the previous database that had a decrease in publication counts of more than 40% in the latest year in the current database.

PK_KB_INST	Name	Previous pubs	Current pubs	No. diff.	Perc. diff.
259	Marien-Hospital Wesel gGmbH	27	16	-11	-40.7
126	Hochschule fur Musik, Theater und Medien Hannover	28	16	-12	-42.9
821	Deutsches Archaologisches Institut	28	16	-12	-42.9
154	EBS Universitat fur Wirtschaft und Recht	30	17	-13	-43.3
1124	Fraunhofer-Institut fur Zuverlassigkeit und Mikrointegration	23	13	-10	-43.5
629	Hochschule Flensburg	75	42	-33	-44.0
360	Klinikum Leverkusen gGmbH	22	12	-10	-45.5
1180	Fraunhofer-Institut fur Elektronische Nanosysteme ENAS	37	20	-17	-45.9
1226	ThyssenKrupp AG	25	12	-13	-52.0
1575	B. Braun Melsungen AG	25	12	-13	-52.0

PK_KB_INST	Name	Previous pubs	Current pubs	No. diff.	Perc. diff.
392	HSK, Dr. Horst Schmidt Kliniken GmbH	35	16	-19	-54.3
1491	Evonik Industries AG	46	19	-27	-58.7
2153	Bayer HealthCare AG	256	103	-153	-59.8
2015	Zentrum fur Allgemeine Sprachwissenschaft, Typologie und Universalienforschung	23	9	-14	-60.9
474	Stadtisches Klinikum Dessau	30	11	-19	-63.3
4172	Bernstein Fokus: Neurotechnologie (BFNT)	35	9	-26	-74.3

Authors: Mean number of authors by Research Area and discipline

The mean number of authors on a paper can be informative about patterns of collaboration and their potential implications for fractional counting. For instance, increasing levels of inter-sector or international collaboration could result in decreased publication counts for individual sectors or countries when using fractional counting. As such, understanding changes in authorship patterns can provide some insight into potential macro-level changes for entities.

We show in the left panel of Figure 13 the mean number of authors per sc_extended discipline in 2017 in both databases, and in the right panel the mean number of authors per discipline in 2017 in the wos_b_2019 database compared to 2018 in the wos_b_2019 database.

While little change is expected to be seen in the left-hand panel of Figure 13 as the number of authors on a paper is unlikely to change between databases, differences in the right-hand panel indicate potential changes in disciplines' collaboration patterns. Disciplines for which the mean number of authors changed by more than 5%, based on the right-hand panel of Figure 13, are shown in Table 13. Also, to assess trends over a longer time-series, we present the mean number of authors per Research Area over the last ten common years of both databases in Figure 14.

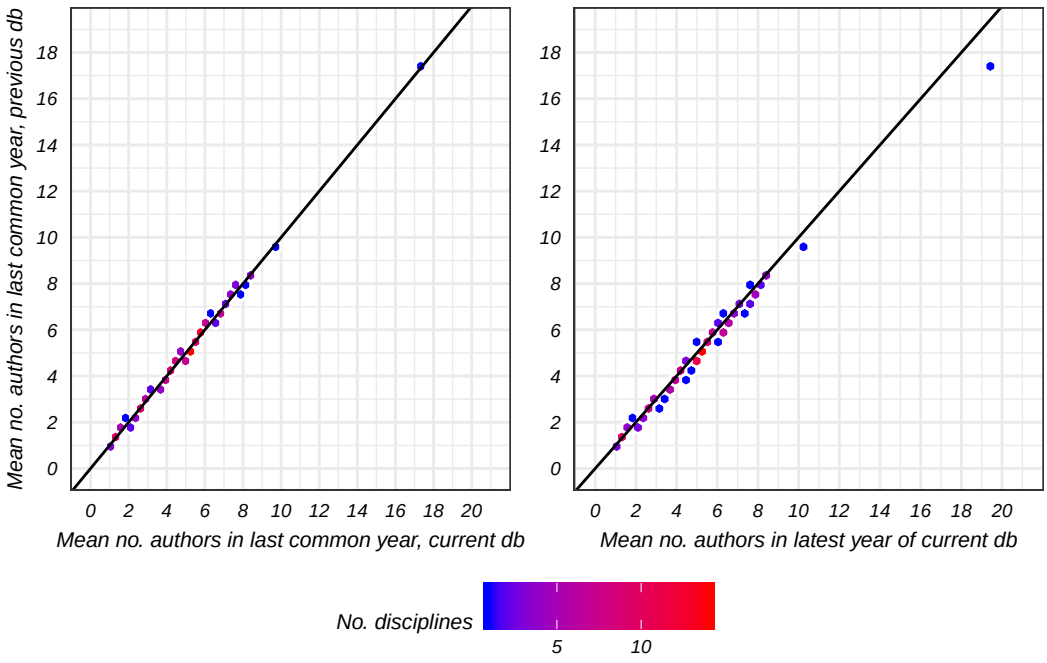

Figure 13: Mean number of authors per discipline (sc_extended) between databases, where colour denotes the number of disciplines with this combination of mean authors.

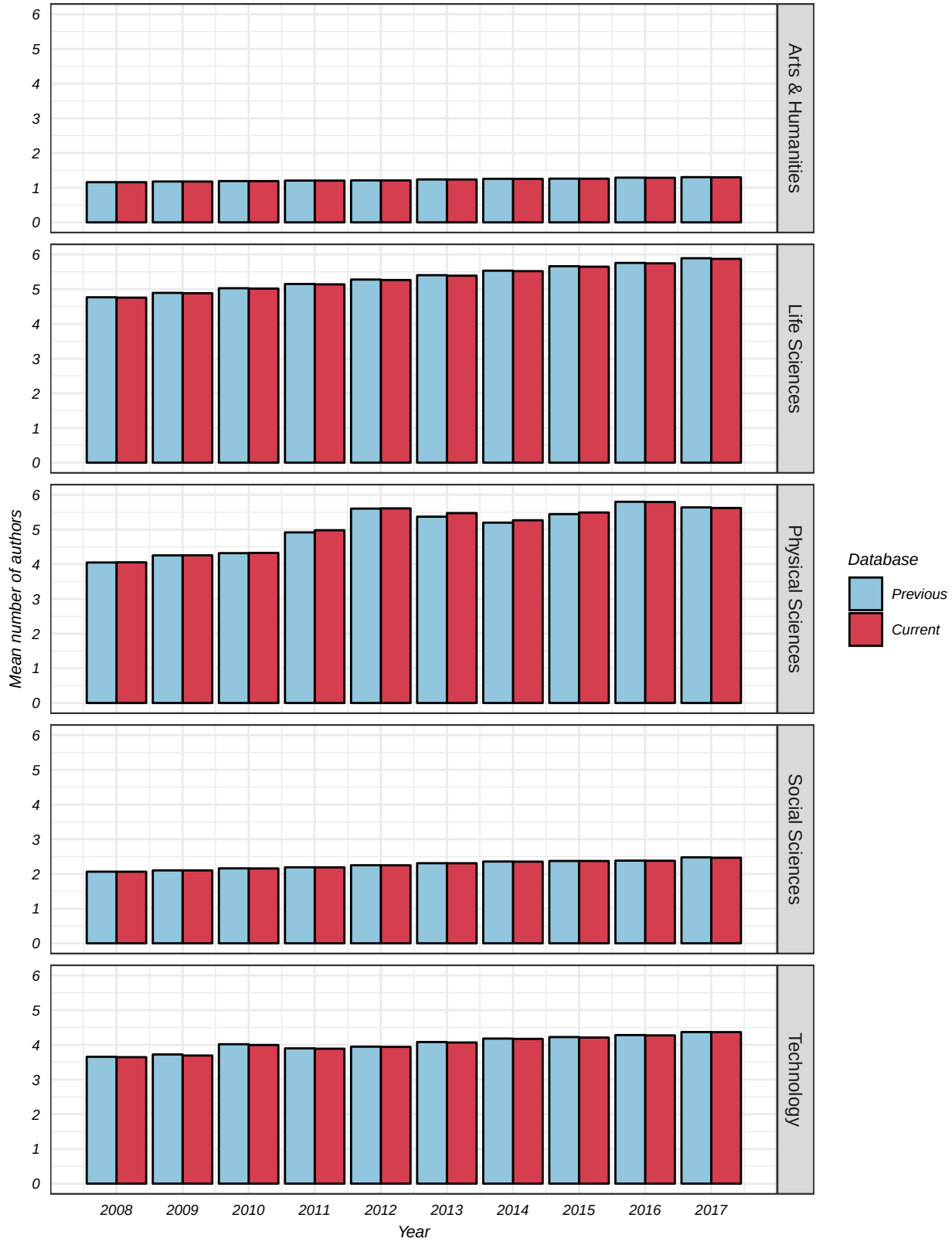

Figure 14: Mean number of authors by Research Area and database over time.

Table 13: Disciplines (sc_extended) where the mean number of authors changed by more than 5% between the last common year in the previous database and the latest year in the current database.

Discipline	Previous mean authors	Current mean authors	No. diff.	Perc. diff.
Medical Ethics	2.5	2.9	0.4	15.1
Astronomy & Astrophysics	17.5	19.5	2.0	11.3
Social Work	3.0	3.3	0.3	11.3
Biomedical Social Sciences	4.1	4.5	0.4	10.7
Dance	1.1	1.2	0.1	9.4
Food Science & Technology	4.9	5.3	0.4	8.2
International Relations	1.8	1.9	0.1	8.0
Physics	9.6	10.3	0.7	7.2
Public Administration	2.3	2.5	0.2	6.6
Biotechnology & Applied Microbiology	5.6	6.0	0.4	6.4
Education & Educational Research	2.9	3.1	0.2	6.2
Family Studies	3.4	3.6	0.2	6.2
Ophthalmology	5.4	5.7	0.3	5.6
Physiology	5.8	6.1	0.3	5.5
Social Sciences - Other Topics	2.4	2.6	0.1	5.5
Fisheries	5.0	5.3	0.3	5.3
Remote Sensing	4.5	4.7	0.2	5.3
Crystallography	4.9	5.1	0.3	5.2
General & Internal Medicine	6.6	6.9	0.3	5.0
Research & Experimental Medicine	6.8	7.2	0.3	5.0

Source items: Percentage by Research Area and discipline

Source items refer to whether the publications on the reference list of an indexed publication are also indexed in the database, as opposed to not indexed and therefore non-source. Only source items are included in citation counts and so understanding the percentage of items cited that are also source can give an indication of the depth of WoS' coverage of a discipline. That is, if a large number of indexed items' sources are not indexed, the reverse is also likely true and a large number of citations of indexed items are also missing, which has the effect of reducing citation counts for disciplines with lower coverage, such as the arts and humanities.

The percentage of references that are source items is expected to increase over time as Clarivate continues to index journals and makes efforts to improve coverage of journals from disciplines with known low coverage. The percentage is not likely to ever reach 100% however, as authors will continue to cite items outside of the scope or coverage of WoS.

We show in the left-hand panel of Figure 15 the percentage of references that are source items per sc_extended discipline in 2017 in both databases, and in the right-hand panel the percentage of references that are source items per discipline in 2017 in the wos_b_2019 database compared to 2018 in the wos_b_2019 database.

It is in the right-hand panel that the effect of recently indexed journals may become apparent, where an increase in the percentage of source items may be seen if the journal is often cited within a discipline. The disciplines with a change in the percentage of indexed references of more than five percentage points between databases, based on the right-hand panel of Figure 15, are shown in Table 14. Longer term trends can be seen in Figure 16 where we present the percentage of reference that are source items per Research Area over the last ten common years of both databases.

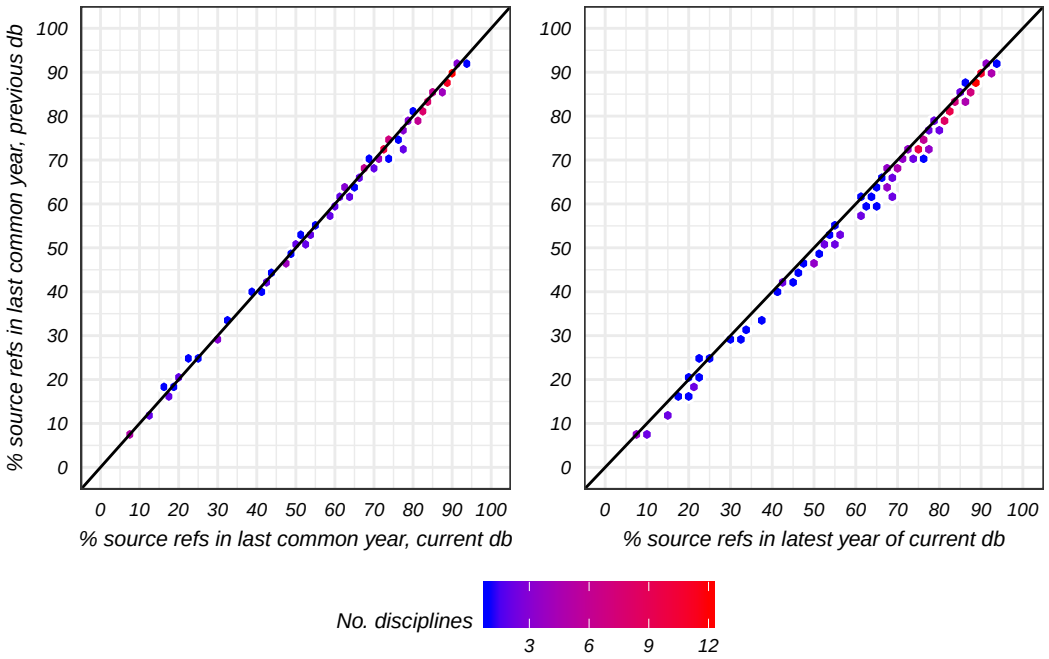

Figure 15: The percentage of cited items that are source items per sc_extended discipline by database, where colour denotes the number of disciplines with this combination of source references.

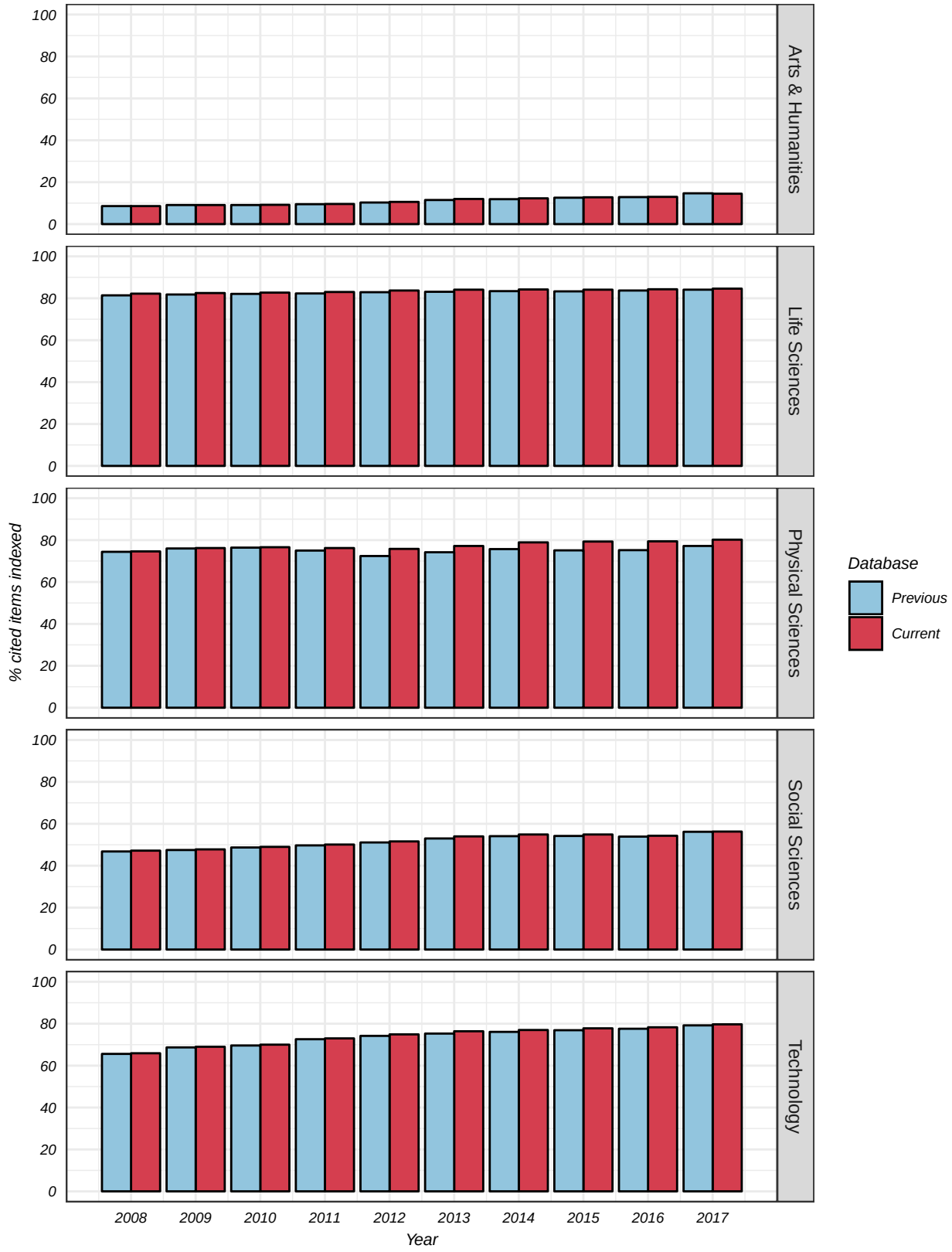

Figure 16: The percentage of cited items that are also indexed in WoS by Research Area and database over time.

Table 14: Disciplines (sc_extended) where the percentage of indexed references changed by more than 5 percentage points between the last common year in the previous database and the latest year in the current database.

Discipline	Previous no. refs.	Prvs % source	Current no. refs.	Crrnt % source	Change
Robotics	396,130	62.3	442,083	68.3	6.0
Physics	84,146,213	72.5	104,611,868	78.3	5.8
Instruments & Instrumentation	4,557,511	70.1	5,670,642	75.7	5.6
Telecommunications	2,607,182	62.2	3,716,239	67.6	5.4

References

[1] S. Stahlschmidt, D. Stephen and S. Hinze. “Performance and Structures of the German Science System”. In: Studien zum deutschen Innovationssystem. Expertenkommission Forschung und Innovation (EFI), 2019. Chap. Studie 5-2019.

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