Institute for Scientific Information

Feburary 2020

Global Research Report
The value of bibliometric databases: Data-intensive studies beyond search and discovery

Jonathan Adams, David Pendlebury and Martin Szomszor

Author biographies

Jonathan Adams is Chief Scientist at the Institute for Scientific Information (ISI). He is also a Visiting Professor at King's College London, Policy Institute. In 2017 he was awarded an Honorary D.Sc. by the University of Exeter, for his work in higher education and research policy.

David Pendlebury is Head of Research Analysis at ISI. Since 1983 he has used *Web of Science* data to study the structure and dynamics of research. He worked for many years with ISI founder Eugene Garfield. With Henry Small, David developed ISI's *Essential Science Indicators*.

Martin Szomszor is Director at the Institute for Scientific Information and has also held the role of Head of Research Analytics at ISI. He was named a 2015 top-50 UK Information Age data leader for his work in creating the REF2014 impact case studies database for the Higher Education Funding Council for England (HEFCE).

Foundational past, visionary future

The Institute for Scientific Information

ISI builds on the work of Dr. Eugene Garfield – the original founder and a pioneer of information science. Named after the company he founded – the forerunner of the Web of Science Group – ISI was re-established in 2018 and serves as a home for analytic expertise, guided by his legacy and adapted to respond to technological advancements.

Our global team of industry-recognized experts focus on the development of existing and new bibliometric and analytical approaches, whilst fostering collaborations with partners and academic colleagues across the global research community.

Today, as the 'university' of the Web of Science Group, ISI both:

- Maintains the foundational knowledge and editorial rigor upon which the *Web of Science* index and its related products and services are built. Our robust evaluation and curation have been informed by research use and objective analysis for almost half a century. Selective, structured and complete data in the *Web of Science* provide rich insights into the contribution and value of the world's most impactful scientific and research journals. These expert insights enable researchers, publishers, editors, librarians and funders to explore the key drivers of a journal's value for diverse audiences, making better use of the wide body of data and metrics available.
- Carries out research to sustain, extend and improve the knowledge base and disseminates that knowledge to our colleagues, partners and all those who deal with research in academia, corporations, funders, publishers and governments via our reports and publications and at events and conferences.

ISBN: 978-1-9160868-7-6

Executive summary

In 2019, around 145,000 researchers from, on average, 139 countries working across a diversity of research disciplines interrogated the *Web of Science* each day to explore research information and discover key literature to inform current research.

In 1981 the *Web of Science* indexed approximately 500,000 papers (substantive academic articles and reviews) from 6,800 journals; this expanded substantially to 2.5 million papers sourced from 21,300 journals in 2019. This is a deep data resource for a wide range of analytic uses.

There are, however, few studies of how the *Web of Science* is used as a bibliographic database other than for the purposes of search and discovery. Our analysis shows that *Web of Science* is the primary source of publication and citation data for the majority of systematic research reviews across a broad range of disciplines and about twice as many research management and evaluation studies as any other source. *Web of Science* is the primary data source for such work in the USA, China, and most of western Europe. Countries where the Scopus database was more frequently acknowledged include Iran and Italy.

A key beneficiary of structured use of *Web of Science* bibliographic records are the biomedical researchers who have an established and structured approach to accessing raw material for reviews that inform the development and current state of research topics that are critical to human health and disease control.

The topical structure of such publications demonstrates the critical relevance of *Web of Science* literature to review health policy targets like cancer, women's health and cardiovascular disease, the management of medical outcomes, and the development of innovative methods and treatments.

Tracking such literature across time reveals the emergence of new fields and helps inform the direction of funding. Identifying frequent authors in a select area reveals the distribution of expertise and supports comparative evaluation of activity and outcomes for national policy and institutional research management.

Our analysis shows that *Web of Science* is the primary source of publication and citation data for the majority of systematic research reviews.

Introduction

The *Web of Science* was interrogated every day in 2019 by around 145,000 researchers across an average of 139 countries, representing the full spectrum of research disciplines in the natural and social sciences and, increasingly, in the humanities.

Researchers use the *Web of Science* to initiate new research plans and help frame the questions they want to answer; and then to search for and discover the key literature that helps to support and inform their current research.

Perhaps surprisingly, there are few studies of the ways in which the bibliographic database is used by the research community for other purposes than for search and discovery. Pringle (2008) discussed the rising use of bibliographic metadata in research evaluation and Schnell (2018) documented the historical place of the *Web of Science* as the first citation index for data analytics and scientometrics. Most recently, Li, Rollins and Yan (2018) confirmed that the quantitative impact of the *Web of Science* had not been rigorously examined by scientific studies. They investigated the ways in which this source was mentioned in a sample of 19,478 papers published between 1997-2017 and analysed the distribution across countries, institutions and domains. This appears to have been the first study to empirically investigate the documentation on the use of the database.

In this report, we extend the studies of Li et al (2018) and Schnell (2018) by focusing on the ways researchers are using *Web of Science* data to learn about the findings reported across the entire scientific and scholarly communication system and how this information can support specific studies, especially systematic reviews, and support better research management.

The numbers of papers in the *Web of Science*, and the spread of research it covers, has expanded hugely over the last few decades. In 1981 it indexed about 500,000 papers (substantive academic articles and reviews) every year from around 6,800 journals. Today, it indexes about 2.5 million papers from about 21,300 journals annually. The index houses around 20 million papers from researchers based in the USA or about 27 million papers sourced from researchers in the European Union: a deep data resource for a wide range of analytic uses.

The numbers of papers in the *Web of Science,* and the spread of research it covers, has expanded hugely over the last few decades.

Publication records are structured in journal-based categories (originally designed to make searching easier) but now, are more frequently employed as the basis for comparative analytics. There are 254 detailed *Web of Science* journal categories (as of January 2020) and 21 broader categories in the *Essential Science Indicators*. Citation analysis is often carried out using the framework of *Web of Science* categories, where the close disciplinary relationship between journals assigned to any category means that the average citation count for all articles in that category in a particular year is a meaningful global benchmark.

The focus of research management interest is, however, rarely pitched at this categorical level. It is much more likely to be ‘topical’ at a finer grain, within or between established journal categories. Research is increasingly interdisciplinary, as the focus of research topics often spans traditional disciplinary structures and cutting-edge solutions to major challenges draw on multiple fields. This means that the information needed by both researchers and those tasked with selecting, funding and evaluating the outcomes of research projects is likely to come from targeted searches, analyses and reports, to be of wide interest across a field.

The concentrated mass of the database has many tales to tell. Drawing together related publications in a structured way provides the raw material for reviews that reveal the development and current state of broader research topics. Tracking literature across time reveals the emergence of new fields and may help to direct funding where it can be most usefully spent. Identifying the most frequent authors in a select area reveals the distribution of expertise and supports comparative evaluation of activity and outcomes.

When, what and where?

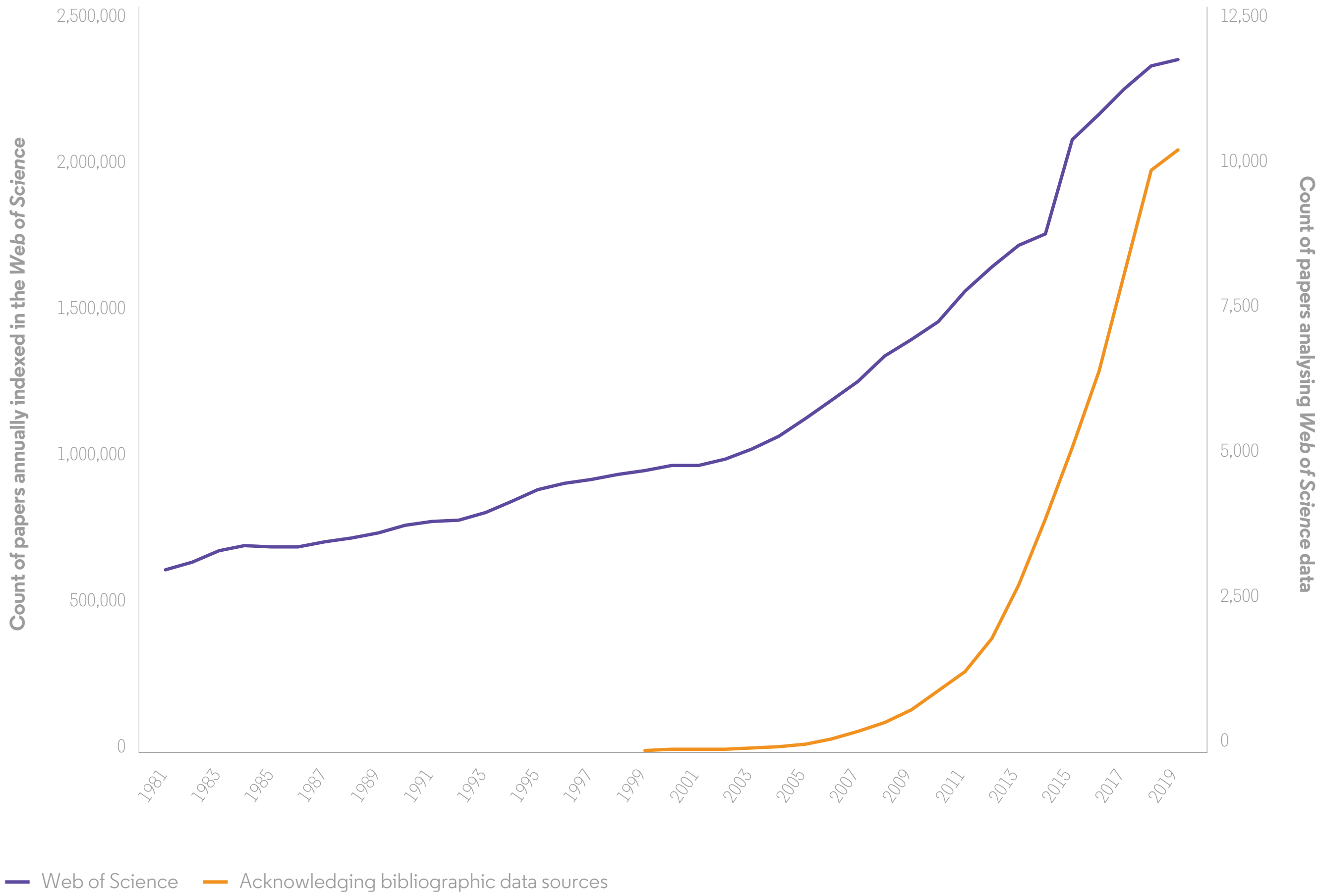
Our first step in this report is to look at the numbers of papers (i.e. both articles and reviews) published each year and indexed in the *Web of Science* that also reference, in their titles, abstracts and key-words, the analysis of data from the *Web of Science* or other widely-used, global publication databases.

Over the period 1999 through to mid-2019, when we extracted these data, we found 51,120 papers that explicitly referenced the use of one or more major bibliographic data sources. We assume that such an acknowledgment implies that there has been more usage than the 'normal' level of academic search and discovery that would be required to produce a research paper.

The data shows how rapid the growth of such studies has been – compared to the overall growth of the underlying database. Whereas acknowledgments of database source were rare in studies using bibliographic data in the 1990s, numbers rose to around 1,000 papers per year by 2010 and growth since then has sky-rocketed. There are now over 10,000 papers per year that use bibliographic data in some form for reviews and analytics (Figure 1).

The data shows how rapid the growth of such studies has been, compared to the overall growth of the underlying database.

Figure 1.
Annual numbers of papers (articles and reviews) indexed every year in the *Web of Science* since 1981 and the count of papers that report that their content has drawn on bibliographic data sources for analyses. The data for 2019 cover only a part year (to September).



The *Web of Science* is not the only source for such analyses, and indeed is predated by the Index Medicus (1879) and Biosis (1926), but the Science Citation Index (the forerunner of the *Web of Science* index) is unquestionably the oldest citation database in the sciences. Other important sources exist. These include the Scopus database, created in 2004 and managed by the publishing company Elsevier; and the Google Scholar database, which is not curated or indexed. In order to provide comparability across disciplines, this analysis excludes subject-specific databases such as PubMed, JSTOR and DBLP and does not consider pre-print indexes such as arXiv.

Many prior studies report that they have drawn on more than one of these sources. The overlapping circles of the diagram in Figure 2 shows us that of the 29,079 publications that cited *Web of Science* publication data, there were 7,212 (25%) that also drew on Scopus data and 1,488 (5%) of these also used Google Scholar.

From a different perspective, some 20,511 studies used Scopus data of which about 35% also used *Web of Science* (Figure 2).

This is an interesting outcome. Google Scholar poses problems for bibliographic use, including the need for analysts and other research users to de-duplicate and disambiguate the information. There is no information on the scope of coverage. The benefit of Google Scholar is free access. *Web of Science* and Scopus are commercial and curated sources.

Which types of publications use the Web of Science?

The majority are reviews, but it is also popular with those researching more typical academic articles. Data use varies between these two types of publications. Among reviews, around 20% more chose *Web of Science* over Scopus as a source for their analysis and content.

Among articles, however, about twice as many report using *Web of Science* data compared to the number reporting Scopus data. This is likely linked to the nature of the publication content and analyst's objectives, which we discuss later in this report (Figure 3).

Where do the publications come from?

The long-established research economies in the USA and Western Europe make up a substantial part of the activity, joined by the Anglophone diaspora in Australia and Canada, but by far the biggest single contributor is China. In the great majority of these cases the primary source of information is from the *Web of Science*. In India, however, the majority use Google Scholar, and Google also does well in South Africa. Scopus is the preferred data source in Iran, Italy and Australia (Figure 4).

Figure 2.
The numbers of studies published in academic journals indexed in *Web of Science* (1999-2019) that reported their use of one or more of the three principal bibliographic indexing databases.data sources for analyses. The data for 2019 cover only a part year (to September).

Web of Science 29,079

Scopus 20,511

Google Scholar 14,079

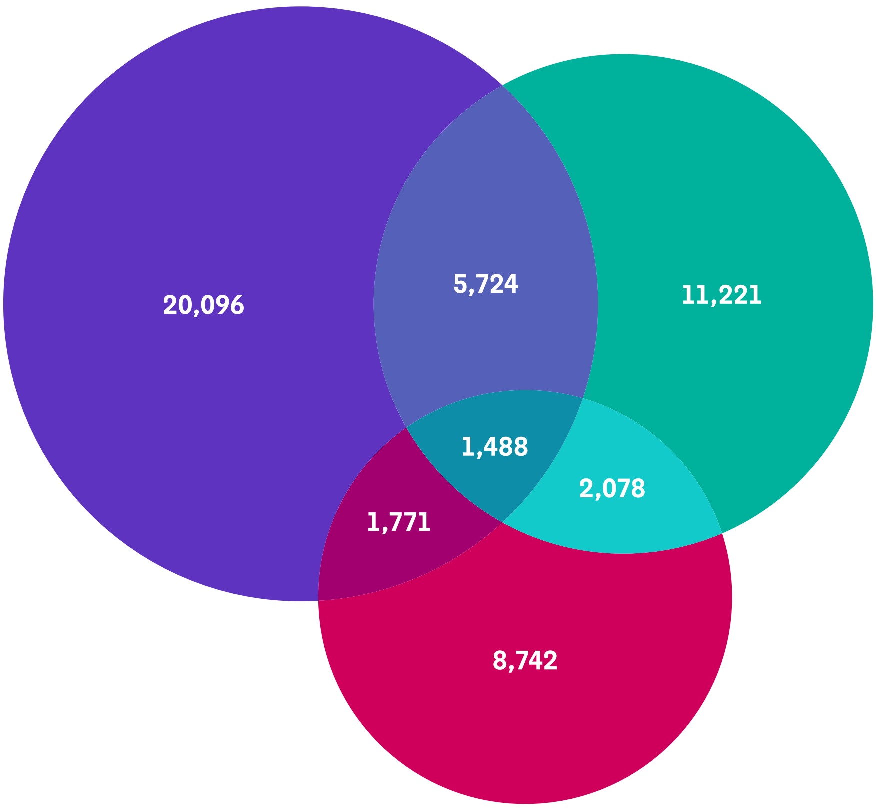


Figure 3.
Types of publications in academic journals indexed in the *Web of Science* (1999-2019) that reported their use of one or more of the three principal bibliographic indexing databases.

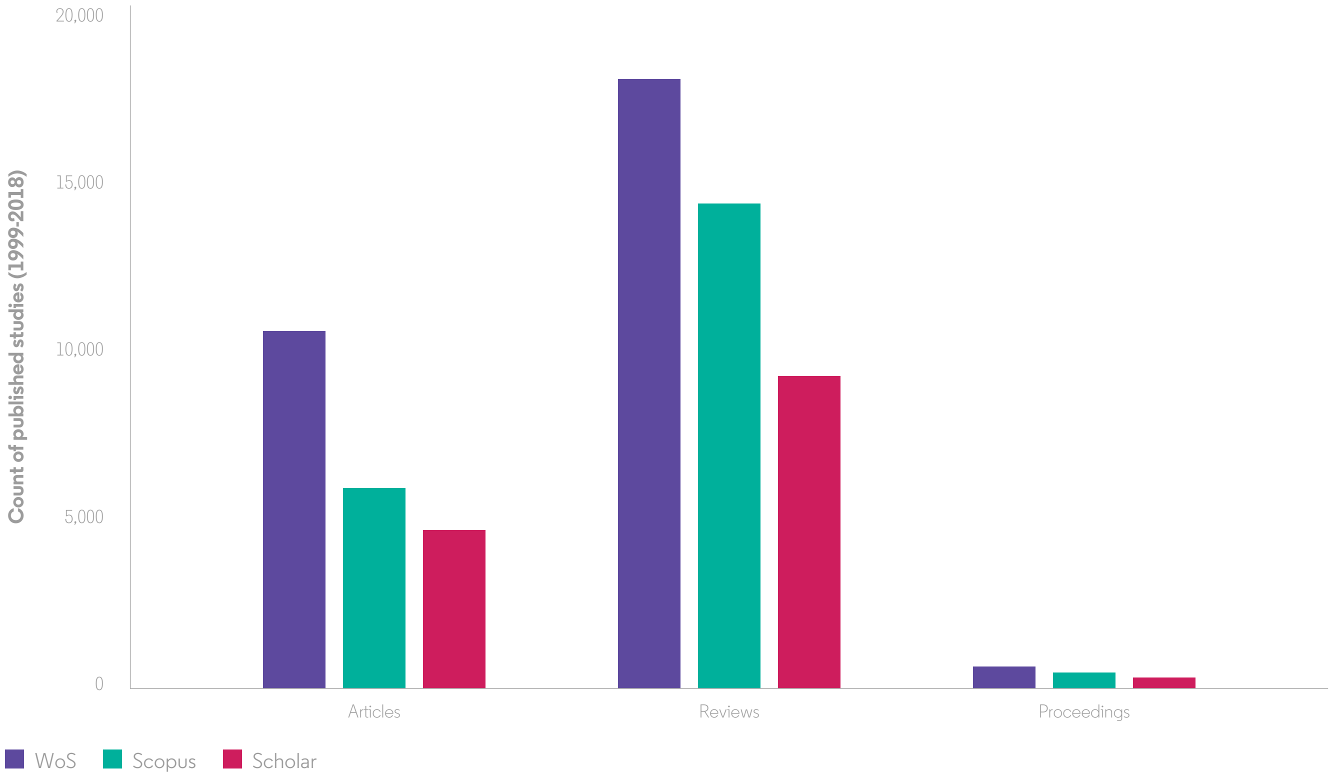
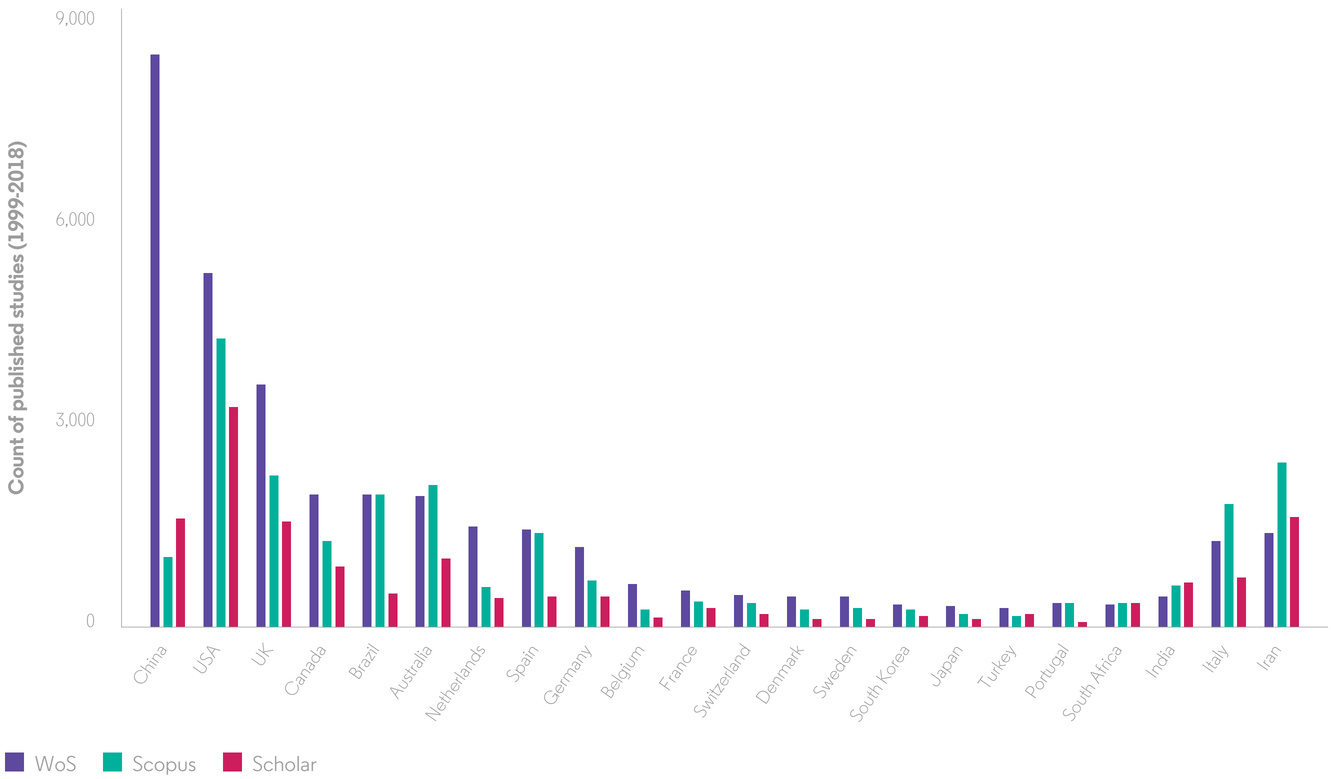


Figure 4.
Regional distribution of publications in academic journals indexed in the *Web of Science* (1999-2019) that reported their use of one or more of the three principal bibliographic indexing databases.



What topics are included?

The rise in papers acknowledging *Web of Science* data and other bibliographic databases as a critical source tells us that the utility of bibliographic data for analysis, as well as for search and discovery, has become evident and valuable to many researchers.

However, we need to recognise that the rise in acknowledgments may also reflect a change in researcher behaviour. People turned to the *Web of Science* as the primary and natural source for any analysis of the literature for many years but, since the appearance of Google Scholar (beta release in 2004) and then Scopus (2004), they have recognized a need to include a variety of sources - often to assure others of the provenance and quality of their information.

What are the characteristics of bibliographic records that deliver those benefits? Two primary use cases stand out: systematic review and research evaluation. There are variant approaches to both.

As in any study, from any field, the identification and description of the data sources used is paramount.

Reviews

There are many approaches to reviews of existing literature, typically for a select topic within a discipline. Reviews have long been identified as one of the most important parts of the entire research corpus and some review journals are amongst the most frequently cited serials in their field. Eugene Garfield, the founder of ISI, recognized the important role of reviews and reviewers in the scientific literature. He noted that statements in reviews were valuable for indexing and that nearly every statement was referenced so the citations were, by extension, also indexing 'statements' (Garfield 1976, 1982).

Some reviews comprise of an annotated list and precis of the recent literature - but others are truly systematic. A good review summarises the results of studies already refereed, accepted and published in the corpus and provides a high level of evidence on the outcomes of prior work and the state of existing knowledge. It will include judgements about the evidence and inform recommendations for further work. Such a complex review may pool data through meta-analyses to arrive at a better overview as the authors assess the supportive and contradictory evidence.

To enable such a significant contribution, there is no better starting point than a curated, deeply structured, bibliographic database that not only contains much of the material required for a review but also offers indexing and data enhancement - enabling the reviewer to rapidly filter, sift and prioritize the material at hand.

In practice, it is difficult to see how an effective review could be implemented without an appropriately comprehensive database. The reviewer has a range of search tools available and should be confident that the publications discovered come from authoritative sources. Those items come from journals that have to pass stringent editorial standards. It is therefore no surprise that the *Web of Science* appears as the most frequent source in review publications.

Research evaluation

Garfield noted in Science in 1955 that citation counts might reflect the impact a publication had on other researchers (Garfield, 1955). Since then, the field of scientometrics has expanded significantly and many academic groups in North America, Europe and – more recently – Australasia and Asia now contribute to its development through studies of the data and the development of sophisticated descriptive and comparative indicators. These have been taken up by research managers and policy analysts in national reports and funding programme evaluation, for which they are well suited.

As in any study, from any field, the identification and description of the data sources used is paramount. There will consequently be many papers in disciplinarily specific journals and a spread across journals relevant to other disciplines where the costs and benefits of research assessment is a topic of interest. Because of the increase in research investment globally, the policy attention paid to methods and to comparative outcomes means that the field has become regionally pervasive.

Methodology

To determine the topics that are the targets for analysis, we examined the text in the titles, abstracts and keywords of 51,120 publications extracted, from the *Web of Science* over the period 1999 to mid-2019. A standard unigram topic-modelling pipeline was used, filtering out terms that appear in more than 50% of the corpus and those that appeared in fewer than three publications. All terms are converted to lowercase, resulting in a dictionary containing 36,095 words. The weighting for each word (i.e. the number of times it appeared in the title, abstract, and keywords) was used as input to the Non-negative matrix factorization (NMF) algorithm to produce topic models for a specified number of topics. The process was executed in Python using standard packages from Scikit learn, a free software machine learning library for the Python programming language (Pedregosa et al., 2011).

We generated a range of models, each producing a different number of topics (ranging from 10 to 50) and evaluated them qualitatively, checking for coherence of the output (i.e., did the model contain topics with words that are clearly related?) and granularity (i.e., how specific in terminology are they?). Although it is possible to produce topic models with very few, or a very large number of topics, the nature of the corpus typically dictates what are sensible choices for a given analysis. We found that a target set of topics that produced a useful result, in the sense that the clusters were of a relatively similar size, contained clearly related terms and were sufficiently specific to partition the publication dataset into identifiable clusters.

At the development stage we did not seek to test the relative diversity of topic content, nor the degree to which they drew on a few or many *Web of Science* journal categories.

We also avoided seeking for balance across years within each topic. In practice, each topic is likely to be strongly influenced by the content of the much more abundant data records of the last few years (see Figure 1).

We found that a target set of about 30 topics produced a relatively balanced set of clusters. A topic is an arbitrary partition in the data records using thresholds of similarity. In this case we used the frequency with which terms are shared. The topics themselves share these terms so we can build up a family tree that relates the topics to one another. This is called a dendrogram and it successively clusters the topics in pairs, groups and families. Topics that cluster together are more similar in their use of terminology than topics that are far apart in the dendrogram.

We found that a target set of about 30 topics produced a relatively balanced set of clusters.

Figures 5 and 6 display and show the relationship between the topics we identified. Each topic has a label that is determined from an analysis of the most frequent terms used by the set of papers in that topic. Labelling is indicative and proper interpretation requires informed review of the actual papers that form the topic. Labelling can never be absolute: expert views differ as to the precise nature of the material in any group created in this or any way; such views also tend to evolve as analysis proceeds.

Topics of interest

Some topics clearly relate to current policy priorities in health; others are about methodology; and others call out underpinning research. Some are tricky to interpret: Topics 2 and 5 are both evidently about genetics, particularly polymorphism and its relationship to patient-specific remedies, but an expert is needed to understand why the algorithm recognises two topics rather than one. Topic 26 refers to social media, but it should be interpreted in the context of media as a source of information and misinformation, particularly about health.

The picture is dominated by papers (which are mainly reviews) related to clinical medicine and health, and to areas of health research including both healthcare, treatment and innovative research. This is unsurprising. The concept of the 'systematic review' is well established in biomedical research. First, because such research has life-critical outcomes so best practice is a key requirement in enabling rapid and efficient progress and this state of knowledge is constantly under review. Secondly, because a very large share of public and commercial research funds are invested in these areas, both the management of current research and of future investment depends on regular and effective monitoring of outcomes.

The Cochrane database of systematic reviews (https://www.cochrane.org/about-us) is particularly important in this regard and has been globally influential in setting standards and demonstrating value and utility. The use of multiple databases to inform such reviews has been studied: Bramer et al (2017) concluded: *"Optimal searches in systematic reviews should search at least Embase, MEDLINE, Web of Science, and Google Scholar as a minimum requirement to guarantee adequate and efficient coverage."*

The upper part of the tree covers global medical priorities (diabetes, obesity, children & pregnancy), disease control and diagnosis (Topic 19 indicates the rising significance of plant-related pharmacology strategies), and a set of topics related to the increasingly important areas of mental health and management. Linked to this are three Topics (3, 21 and 22) which are about aspects of healthcare promotion and management. Health research, data and analysis form a separate cluster (Topics 0, 7, 24 and 11). Standing next to the main cluster of medical and health topics is a separate cluster on earlier stages of research in genetics and molecular biology associated with future health solutions.

The lower part of the diagram shows a quite distinct cluster of topics related to the education and research ecosystem. Topic 26 relates to social media and scholarly communication while Topic 18 is about learning and skills in the context of healthcare training. Finally, publications that describe the use of *Web of Science* data for research and innovation management (Topic 1) and research evaluation (Topics 12 and 27) are the largest individual categories, each with well over 2,000 associated papers. These three topics also differ from the rest of the dataset because the papers are mostly articles, not reviews. Overall, they account for only a minority of the data use cases. This may be a surprise to researchers who believe that they are increasingly accountable for and monitored in their research activity. Of course, these papers are from academics involved in scientometric research, describing case studies that develop or illustrate analytical methodology, rather than from those carrying out evaluations, which are less often published.

A topic and document map

We can also visualise the topics on a map of the publication landscape, rather than as the family tree of Figure 5. The map includes each publication in our dataset as an individual point, groups them by their textual similarity, and pictures the relationship between topics and the clusters, which the dendrogram showed us diagrammatically. It gives us further insights as to the relationship between these groups of documents.

Once we have located the topics among the publication points in the map, we can add other information to reveal further patterns that help with interpretation. In this instance we have added a second version of the map where the publications are colored differently to identify those that are informetric (colored red, where titles or abstracts contain keywords such as 'scientometric', 'bibliometric', 'informetric'), or reviews (colored green, when titles or abstracts contain phrases such as 'systematic review', 'meta-analysis', and 'literature search'). This immediately reveals the differentiation between scientometric topics and the rest of the literature. A total of 4,852 papers were classified as informetric, and a total of 36,957 were classified as reviews.

It is of interest to note that 435 were classified as both informetric and a review. These point to a growing trend to apply bibliometric techniques to enhance the systematic review methodology, for example, to identify emerging research areas within disciplines, to perform topic analysis of the literature using the citation graphs or keyword analysis, and to inspect trends in funding. These enhance understanding of how the research works, providing useful quantitative indicators to inform research strategy.

Figure 5.
Dendrogram showing the similarity between the topics identified among 51,120 papers in the *Web of Science* that use a bibliographic database (publications from 1999-2019). The numbers of papers in each topic is shown to the right of the topic label. Colors indicate topics that cluster together with relatively high degrees of similarity.

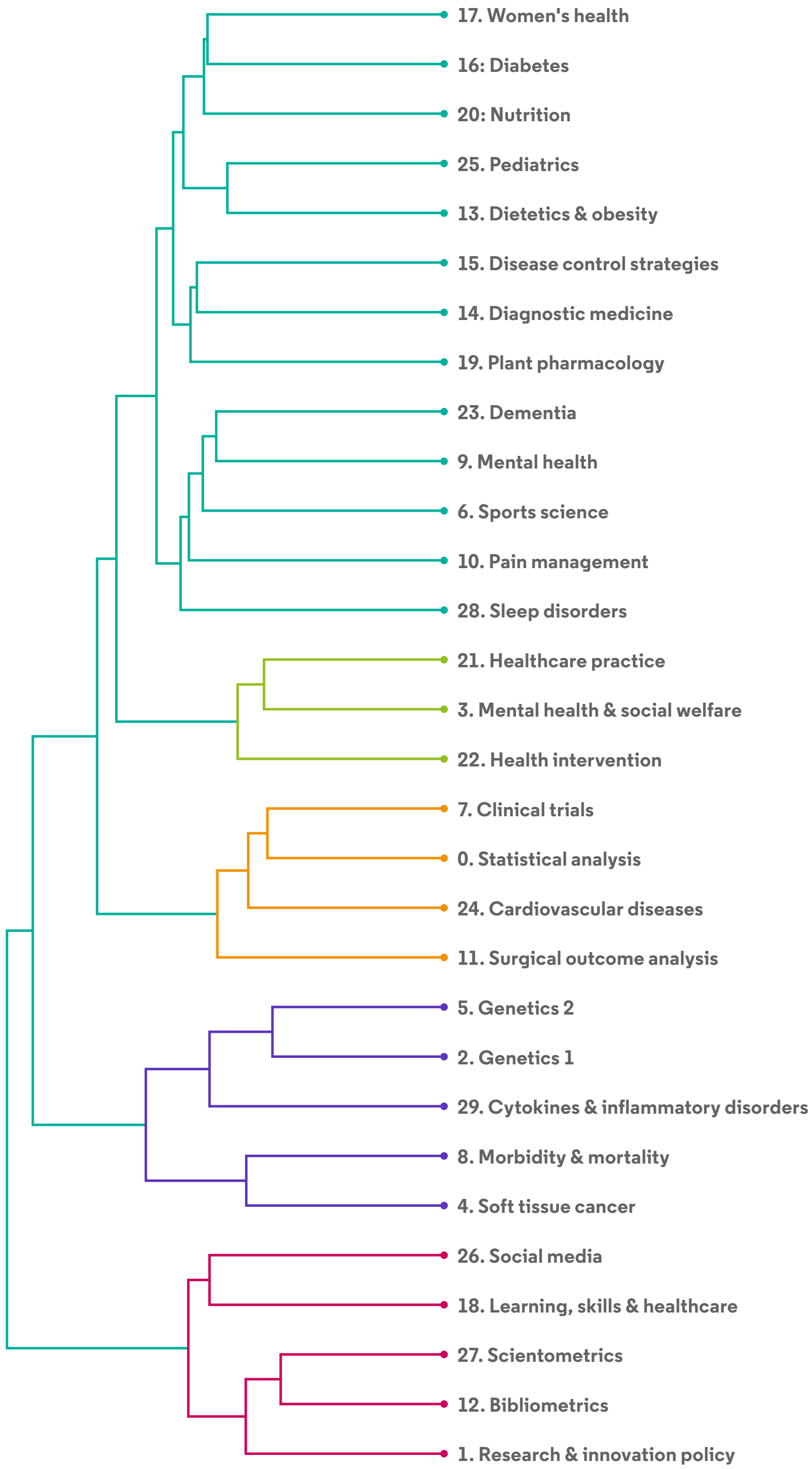
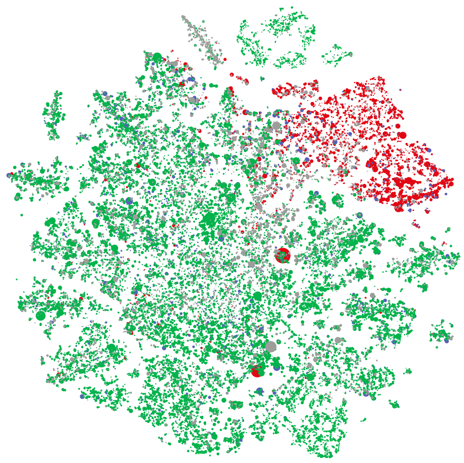


Figure 6.
A topic map of 51,120 papers (1999-2019) that acknowledge the particular use of a bibliographic data source.



Commentary:
The map reveals the structure and makeup of the various research topics that make use of bibliographic databases. The individual topics are clustered using colors corresponding to those shown in the dendrogram (see Figure 5). The crimson cluster contains topics 18, 26, 1, 27 and 12: those that feature numerical approaches and statistical analysis. The green area (topics 21, 3, and 22) are those that relate to social welfare, often practitioner oriented, and are close to specific disease focused topics that form that larger teal cluster. The split purple cluster (topics 2,5 4 and 8) are spread over difference regions, with those in genetics appearing closer to the crimson cluster due to the analytical methodologies used.

Topic growth and category spread

The growth of specific topics over time broadly reflects the general pattern shown in Figure 1. It rises markedly in the years after 2008 and much more steeply in the last few years.

The only topics that stand out as being different, (and then only by degree rather than fundamental profile) are the scientometric topics (12 and 27) where researchers began to more frequently acknowledge their use of specific databases earlier than other topics. The subsequent growth trajectory was then linear rather than tracking the exponential profile seen in biomedical topics.

The topics created from the text in the dataset of 51,120 papers' titles, abstracts, and keywords are a high-level view of the activity. These papers were published in a wide diversity of journals and any one topic may be made up of components not only from the discipline areas that would be expected by a subject-matter expert, but also from other areas related to supporting research and technologies.

There are 254 *Web of Science* journal categories - but more than half of the 51,120 papers discussed in this report were published in just 27 of them (i.e. 11% of journal categories).

Information Science accounts for 1,378 (2.7%) of the total papers that acknowledge their use of bibliographic information. Computer Science (889) and Education Research (476) are also among the more frequently interrogated categories. The rest are biomedical and health orientated with Oncology (2,104) and Public Health (2,020) standing out at the top of the table (Table 1).

There is an expected pattern of inclusion between specific *Web of Science* categories and the indicative labelling of topics. For example, Topic 24 was identified as being primarily about cardiovascular risk management and it is no surprise that 572 (28%) of the 2,008 papers linked to this topic can be found in journals in the 'Cardiac & cardiovascular systems' *Web of Science* category, accounting for close to half of all the Cardiac journal papers (see Table 1), so the remainder are relevant to and spread across other topics.

The other three-quarters of the papers linked to Topic 24 were drawn from a huge diversity of journals. Though most were grouped in expected areas of medical and related research, the spread of other journal categories including chemistry and engineering is an indicator of the increasingly cross-disciplinary nature of the cutting-edge of relevant research captured in an effective review. The completeness and accuracy of the network of citation links in a properly curated bibliographic database becomes an essential part of the art for a topical review author. Tacit, expert knowledge is no longer enough.

Table 1.

Web of Science journal category	Before 2009	2009-2013	2014	2015	2016	2017	2018	2019 (part)	Total
Grand total	1,166	7552	3,860	5,047	6,302	7,943	9,654	9,596	51,120
Oncology	28	330	283	205	263	266	370	359	2,104
Public, environmental & occupational health	40	286	160	193	237	331	390	383	2,020
Information science & library science	93	323	111	139	133	197	240	142	1,378
Dentistry, oral surgery & medicine	41	167	88	116	168	203	285	297	1,365
Pharmacology & pharmacy	45	256	92	110	171	188	246	243	1,351
Surgery	19	162	74	135	186	204	277	281	1,338
Nursing	24	162	80	110	159	197	226	252	1,210
Gastroenterology & hepatology	30	176	115	138	154	167	181	185	1,146
Endocrinology & metabolism	26	162	89	93	118	169	219	254	1,130
Cardiac & cardiovascular systems	27	164	91	112	132	168	209	161	1,064

Future use

As evidence grows, so researchers can link back into the citation network of the Web of Science to collate the most impactful material.

The clusters of topics in Figure 5 are, as noted, dominated by medicine/health (13 topics related to disease, three to disease management, four to medical data and five to basic biology research). The remaining five cover education and research management.

The clustered documents in all but the two topics in research management could be largely classified as reviews, whereas most of the documents that acknowledge *Web of Science* and other bibliographic databases and that would be classified as articles are in the research management area.

Research evaluation and policy, as well as management, is a key part of the support of scientometric research by *Web of Science* publication records and associated data. In this area the *Web of Science* data are the primary data source and used far more often than any other. Eugene Garfield was the key figure in origin, early evolution and underpinning ideas of scientometrics, and his SCI database, the fore-runner of *Web of Science*, gave birth to the field. His legacy continues to inform the field globally today.

The reviews in the biomedical topics draw on a diversity of the *Web of Science* categories (Table 1). This is a reminder of the contribution made by a multiplicity of disciplines, but what is missing in the topic map is any substantive component related to the physical sciences, technology and the social sciences (Figure 6).

This outcome reflects two things: first, the relative maturity of literature review as a tool underpinning research progress in biomedicine; second, the relative frequency and pervasiveness of reviews as a part of biomedical research management. This has been driven by both the scale of investment and the policy priority underpinning that investment.

In biomedicine, the review not only considers the literature but also considers the detail of the research outcomes. The data source becomes an essential part of the audit trail where demand for best practices is also an effort to optimize results in regard to money spent.

Although reviews are also well established and of recognised significance and value in other subjects (e.g. Annual Review of Condensed Matter Physics, Annual Review of Materials Research) they are milestones rather than a part of research management. The raw material captured in the *Web of Science* is also digested and communicated through other routes, often in the world of 'gray literature' which includes government and agency reports; and professional publications. An innovative model is the web-based rolling review, employed by climate change scientists in the ScienceBrief website sciencebrief.org/about where analyses of individual papers are written by scientists as a, "transparent, continuous, and rapid system for reviewing current knowledge". As evidence grows, so researchers can link back into the citation network of the *Web of Science* to collate the most impactful material.

References and background reading

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J. and Franco, O.H. (2017). Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews*, 6, article number 245.

Garfield, E. (1955). Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, 122, 108-111.

Garfield, E. (1976). Significant journals of science. *Nature*, 264, 609-615.

Garfield, E. (1982). ISI's "new" *Index to Scientific Reviews* (ISR): applying research front specialty searching to the retrieval of the review literature. *Current Contents*, 39, 5-12.

Garfield, E. (1987a). Reviewing review literature: Part 1 - definitions and uses of review. *Current Contents*, 18, 5-8.

Garfield, E. (1987b). Reviewing review literature: Part 2 – the place of reviews in the scientific literature. *Current Contents*, 19, 5-10

Gurney, K. A. and Pendlebury, D. (2012). *Funding acknowledgements in the journal literature: uses and benefits for funders, recipients, and analysts*. Philadelphia, PA: Thomson Reuters.

Li, K., Rollins, J. and Yan, E. (2018). *Web of Science* use in published research and review papers 1997–2017: a selective, dynamic, cross-domain, content-based analysis. *Scientometrics*, 115, 1–20.

Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pringle, J. (2008). Trends in the use of ISI citation databases for evaluation. *Learned Publishing*, 21, 85-9.

Schnell, J. D. (2018). *Web of Science*: the first citation index for data analytics and scientometrics. In *Research Analytics: Boosting University Productivity and Competitiveness through Scientometrics*. Edited by Francisco J. Cantú-Ortiz. Boca Raton, FL: CRC Press, pp. 15-29. ISBN-13: 978-1498785426

About the Global Research Report series from the Institute for Scientific Information (ISI)

Our Global Research Reports draw on our unique industry insights to offer insights, analysis, ideas and commentary to enlighten and stimulate debate. Each one in the series demonstrates the huge potential of research data to inform management issues in research assessment and research policy and to accelerate development of the global research base.

Download here:
www.webofsciencegroup.com/isi

About the Web of Science Group

The Web of Science Group, a Clarivate Analytics company, organizes the world's research information to enable academia, corporations, publishers and governments to accelerate the pace of research. It is powered by the *Web of Science* – the world's largest publisher-neutral citation index and research intelligence platform. Its many well-known brands also include *Converis, EndNote, Kopernio, Publons, ScholarOne* and the Institute for Scientific Information (ISI).

The 'university' of the Web of Science Group, ISI maintains the knowledge corpus upon which the index and related information and analytical content and services are built; it disseminates that knowledge externally through events, conferences and publications and it carries out research to sustain, extend and improve the knowledge base. ISI constantly reviews existing standards and works with public and private sector partners to explore publication and citation data to create innovative indicators of value to researchers and research managers.

For more information, please visit webofsciencegroup.com

e: ISI@clarivate.com

@webofscience

webofsciencegroup.com/isi

WS679443270 / 01