Bibliometrics in the context of research evaluation and research policy

Purnell, P.J.

Bibliometrics in the Context of Research Evaluation and Research Policy

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 28 januari 2025
klokke 11:30 uur

door

Philip James Purnell
geboren te Brighton, Groot-Brittannië
in 1970

Promotor:

Prof.dr. L.R. Waltman

Co-promotor:

Prof.dr. A.F.J. van Raan

Promotiecommissie:

Prof.dr. B.A. Barendregt (Decaan Graduate School / voorzitter)

Prof.dr. T.N. van Leeuwen

Prof.dr.ing. J.W. Erisman

Prof.dr. M.J. Spierenburg

Dr. R. Costas Comesana

Prof.dr. W.J.E. Glänzel (KU Leuven, België)

Table of contents

Chapter 1

General introduction

1.1 Overview of the dissertation

Bibliometrics is the quantitative measurement of various aspects of research including productivity, impact, subject categorisation, science networks, and time trends. Such bibliometric measurements are made based on data about scientific publications. Key consumers of bibliometric reports are research policymakers and those conducting research evaluation. Bibliometrics is a useful tool because it is relatively quick, cheap, and easy to use, although research evaluators and policymakers may have other tools available, especially the use of expert opinion, which is more resource-intensive but conducted by peers with subject matter expertise. There is a lot of debate in the research community about how bibliometrics and expert opinion should be combined, or whether one approach is better than the other. The aim of this dissertation is to demonstrate the benefits of bibliometrics to policymakers and other stakeholders, while highlighting and analysing the causes of its limits. Research evaluators and policymakers will benefit from a good understanding of the advantages and limitations of bibliometrics and how expert opinion can enhance the big picture presented by bibliometric analysis.

The following sections of chapter 1 comprise an introduction to research policy and research evaluation, followed by a detailed history and description of bibliometric techniques used in research policy development and research assessment. The chapter concludes with a summary of the motivation and objectives for the dissertation. Chapters 2 – 6 comprise five bibliometric studies that illustrate how quantitative techniques can generate large-scale database comparisons, subject maps, trend and geographical analyses that can inform research policy and evaluation. These studies also highlight some of the limitations of bibliometrics related to data reliability and the need for complementary expert judgement. Across the studies, some common themes emerge that may improve the usefulness of bibliometrics through community-led initiatives and new technologies.

Chapter 7 provides a summary of the findings of each of the studies along with the overall implications. It becomes clear that bibliometrics is not only measuring, but also shaping the research system. Lessons are learned on how bibliometrics can have both desired and unintentional effects on the research system. Key takeaways from the studies on the debate between bibliometrics and expert opinion result in recommendations on finding the ideal balance between these approaches. Positive effects of community-led initiatives already noticeable in bibliometrics are encouraging and suggest there are more ways the community can build on these gains. The future of bibliometrics in research evaluation and policymaking will depend on improvements in data quality and curation, and on ensuring it shapes the research community in a positive way.

1.2 Research policy

Many organisations and regional bodies offer differing definitions of research policy, science policy, and a range of related policies. Science policy can be seen as making public funds ‘more effective and more efficient’ at new knowledge creation (Organisation for Economic Co-operation and Development, 2024). The European Commission’s research & innovation policy aims to address major societal challenges by ‘converting them into opportunities’ for innovation (DG Research and Innovation, 2020). The Commission holds an annual research & innovation policy forum known as the European Research & Innovation Days that aim to convene policymakers, citizens, and other stakeholders to co-design its action plan (European Commission, 2024). The United States science and innovation policy aims to ‘strengthen research and innovation enterprise’ in pursuit of national goals (National Academies, 2024). National or regional science policies often feed into the strategic plans of individual institutions, which may develop their own research policy. Institutional research policies are often concerned with aspects of research that include research ethics, equity, diversity, and inclusivity, open science, and intellectual property, and are closely related to the distribution of research funds (e.g. Cambridge University, 2024; New York University, 2024; University of California Berkeley, 2024; University of Pretoria, 2023; University of Technology Sydney, 2024).

The definitions of science, research, innovation, and other related policies are therefore not clearly distinguished or consistently defined and definitions often overlap, but research policy can generally be seen as a component of a broader science and innovation policy. For the purposes of this dissertation, I consider science policy as a broader term that governs the overall national approach to innovation, while research policy is more concerned with activities conducted at institutional or departmental level. Although the topics discussed here address challenges at a range of levels, I will from here on refer only to research policy for the sake of clarity.

Research policymakers are usually involved in governance or allocation of financial resources to be spent on research. Recipients of research funding may include universities, research institutions, private industry, nonprofit organisations, individual researchers, postgraduate students, or government entities. Typically research grants are used to pay salaries, buy or loan the use of equipment, build new laboratories, cover travel for members of the research team, and pay the costs of publishing research findings. Research policymakers therefore sometimes make use of bibliometric analyses to inform policy development.

In the modern world, scientific knowledge has become a resource upon which societies are built, technological advances based, and for which people compete as they do for other commodities. In this competitive environment, approval of research projects, hiring of academics, and assignation of research funds require justification and are frequently subject to a selective evaluation process. Techniques to assess research productivity, performance, and impact including bibliometrics have therefore been developed to support decision-making and have become a key component of research policy.

1.3 Research assessment

There are two broad approaches to research evaluation, one is to seek expert opinion of researchers familiar with the research being assessed in processes collectively known as peer review. The other is to aggregate quantitative performance indicators such as impact factors and is described as bibliometrics. There are strong arguments for and against both peer review and bibliometrics in research evaluation with research policymakers having to carefully set out their assessment procedures by combining elements of both.

In the next part of this introduction, I will give an overview of the arguments in favour of and against both peer review and bibliometric techniques and then introduce the concept of a middle path which incorporates elements of both methods. Thereafter, I will concentrate on bibliometrics and its relationship with various research policies, which is the subject of this dissertation. Bibliometrics is sometimes seen simply as a tool to support research assessment. I take a broader view that bibliometric techniques go beyond assessment and can be used for other types of studies such as describing the structure of fields, collaboration network and trend analysis, and science mapping studies. I show that bibliometrics can sometimes mislead us and raise questions about the extent to which we are comfortable relying on bibliometric techniques and when we should incorporate a heavier emphasis on peer review and expert judgement more in general.

1.3.1 Peer review

Since 1731, members of the Royal Society of Edinburgh have been asked to review contributions prior to publication, a process followed by the Royal Society of London in 1752 (Spier, 2002). Since then, the academic community has judged research quality among fellow scholars, asking peers to review each other's scientific manuscripts and give their professional opinion on research proposals, applications for academic posts, and many other processes that require specialised knowledge. Peers' expert opinion is intended to provide a unique perspective based on deep understanding of the topic, scientist, or process being assessed.

With this unique perspective comes the risk of prejudiced personal judgement where the reviewer is influenced by preferences, or fear of negative perception and consequences of the review. To encourage candid review, the reviewer's identity is often masked (single-anonymous review) and to mitigate potential reviewer bias the identity of those being reviewed is frequently anonymised (double-anonymous review).

Conducting, writing, and publishing academic research requires knowledge and adherence to specific scientific protocols. Other academics are therefore uniquely placed to judge the quality of another scholar's contribution. The peer review process has been shown to improve manuscripts through suggestions on presentation, readability, and through identifying missing references and scientific or statistical errors (Ware, 2008). Indeed, the very fact that manuscripts are subjected to the peer review process is likely to incentivise authors to invest effort to maximise the chance of their manuscript being favourably judged and accepted for publication (Ware, 2008) and can even improve reviewers' own writing (Lundstrom & Baker, 2009).

Critics of the peer review system point to a number of problems. One common argument against peer review concerns the identity of the reviewer or reviewers. Surely the selection of review committee members will determine the outcome of the review. If alternative members were appointed to the committee, the result would likely be rather different (Bertocchi et al., 2015). Another argument surrounds the risk of bias in which reviewers make prejudiced decisions based on their perceptions of the people or institutions they are evaluating. This introduces the risk of creating self-serving networks that support those scientists who are already in the club and make it difficult for newcomers to gain access or approval of the network. Young people or those newly arrived from other places who have not yet made their mark could inadvertently miss out on positive acclaim in favour of colleagues with established names. Another criticism of peer review centres around the costs involved and the burden on the reviewers, especially in the context of national level research assessments (Geuna & Martin, 2003; Martin & Whitley, 2010).

In summary, the processes behind qualitative evaluation of researchers have been described as resource-intensive, and error-prone (Ioannidis & Maniadis, 2023). The peer review process offers benefits to the scientific process but clearly requires a complementary, objective process to provide balanced research assessment. For a full review of the peer review process, see Lee et al. (2013).

1.3.2 Bibliometric approaches

The existence of large amounts of stored data associated with published scholarly works and the availability of analytical tools means that policy makers can quickly and easily observe the state of certain aspects of science and detect trends that will enable informed decision making. The increase in publication data and associated needs for new mechanisms to manage it has been described as the industrialisation of research (de Solla Price, 1978).

The use of bibliometric analysis in the evaluation of science relies on the assumption that production of new knowledge and the recognition of peers are the pillars of scientific impact (Desrochers et al., 2018). In order to measure new knowledge production, research outputs such as academic papers (and also other outputs including books, patents, and datasets) can be counted, so that a greater number of papers is an indicator of higher scientific productivity. Scientific authorship has been described as the 'undisputed coin in the realm of academia' (Cronin, 2001) and is routinely used to assess productivity of researchers, departments, universities, and countries. The link between the author and institution (and therefore country) is made by the affiliation mentioned on the manuscript submitted to the publisher and facilitates analysis of institutional and international collaboration networks and enables rankings of scientists, universities, and countries.

Authors of studies that quote from, or use the ideas presented in earlier papers are expected to refer to the original source in the cited references at the end of their paper. Citations have therefore been described as 'pellets of peer recognition' (Merton, 1988). In other words, authors cite other scholarly works because the cited work has in some way influenced the author in the writing of their own paper (Merton, 1973). The author acknowledges this influence in the form of a cited reference, which allows the author to use the cited work without committing plagiary (Merton, 1988) and which forms the basis of the incremental nature of scientific discovery. For an in-depth review of the reward system of science, see e.g., Desrochers et al. (2018). Large-scale aggregation of publications and citations forms the basis of bibliometric analysis and serves as a proxy for measuring scientific productivity and impact. Indeed, quantitative bibliometric analysis presents policy makers with a quick, easy, and cost-effective complement to peer review.

Some of the assumptions described above do not always hold true and have been used to question the validity and reliability of bibliometric techniques in research evaluation. Authorship can be complicated by questions about the distinction between contributions that warrant full authorship versus those that merit only acknowledgement (Costas & van Leeuwen, 2012; Desrochers et al., 2017), the hierarchical order in which authors are listed, and the decline in single-author papers (Nabout et al., 2015), among others. The motivation of authors to cite each others' work is also the subject of much debate (Knorr-Cetina, 2013; Nicolaisen, 2007). Problems include the potential for manipulation through citing one's own work or that of colleagues to manipulate citation metrics (Davis, 2012; Fister et al., 2016).

Practical matters also need to be considered when designing or interpreting bibliometric studies. For instance, the usual source of publication and citation data for analysis is a citation database. Any research outputs not indexed in the source database used cannot be included in the analysis. Database coverage therefore becomes one of the defining factors in the outcome of any bibliometric study. We will examine the main bibliometric data sources in section 1.3.

1.3.3 The middle path

As there is no ground truth that defines research quality, we cannot judge whether the peer review or bibliometric approach is more accurate. It may be more prudent to judge the extent to which the two approaches reach consensus, that is to what extent do peer review and bibliometric analysis agree? Some studies have shown broad agreement between the results of peer review and bibliometric findings (Bornmann & Leydesdorff, 2013; Pride & Knoth, 2018; A. Van Raan, 2006), while others have shown weaker correlation (Mryglod et al., 2015b, 2015a; Wilsdon et al., 2015). There appears to be substantial variance in correlation between peer review and bibliometrics depending on the level of aggregation at which the comparison is made (Traag & Waltman, 2019), whether size-dependent or size-independent correlations were used (Mryglod et al., 2013b, 2013a), and when assessing really high-quality research (Bertocchi et al., 2015; CWTS Leiden University, 2007; Rinia et al., 1998).

Although peer review and bibliometrics are often discussed as separate approaches, they are in fact intricately linked and each contains a component of the other. Expert opinion may be influenced by bibliometric indicators and bibliometric studies are often aggregates of expert opinion. For example, several major ranking systems ask academics to vote on the prestige of universities which, when combined with publication and citation counts, determines the rank of universities. On the other hand, when an academic cites another author's papers in their article, this is usually seen as an intellectual endorsement of the paper based on the academic's expert opinion. By counting citations then, we are summing up expert opinions. This shows that there is no clear-cut separation between bibliometric indicators and expert opinion. In practice one will almost always be using a combination of input derived from bibliometric indicators and expert opinion.

Nevertheless, opinions supporting the cases for peer review and bibliometric analysis while eschewing the other are strong, and it can be tempting to see peer review and bibliometric analysis as two opposing camps to choose between. It is easy to take an extreme position that points to the benefits of one camp while emphasising drawbacks of the other. However, embracing the benefits of bibliometric indicators and peer review while taking steps to address their drawbacks is likely to provide the most agreeable way forward in research assessment.

Policy makers and other research stakeholders eagerly seek clarity on the right path to follow; should we embrace, reject or improve bibliometric methods? In this dissertation, I opt for the path of improvement. It is a rocky path, strewn with obstacles and open to criticism from both extremes of the argument. Some of the criticism is valid, but as argued above, thinking in a simplistic way about rejecting bibliometrics in favour of peer review, or the other way around, is counterproductive and misses the crucial point that the two are intricately linked. To find balance, we need a more nuanced perspective, which is what this dissertation aims to offer. Such a perspective is sorely needed to find a middle path as the basis for pragmatic use of bibliometrics by policymakers and evaluators. By identifying weaknesses in the bibliometric system, we can contribute to ways to address these weaknesses and thereby improve the system. This will make the middle path that incorporates elements of both peer review and bibliometric analysis smoother to travel.

1.4 Bibliometric analysis

In this section, I will provide an overview of how bibliometric analysis began and developed into an integral part of systemic research policy development and research evaluation. I will describe the impact of key moments such as the launch of the first citation index, changes brought by the World Wide Web, and the mass meeting of academic minds that insisted on responsible use of metrics. I will also use the section to present a summary of the major data sources used in bibliometric analysis and the important differences between them. These differences are highly relevant because choosing one or another data source can significantly influence the result of any bibliometric analysis. I will also cover the main indicators used in bibliometrics and describe what they are supposed to measure. Due to the popularity (or notoriety) of university rankings and because of their dependence on bibliometric analyses, I will present the major international ranking systems. Finally, I will briefly discuss science mapping. As the amount of data to be analysed has grown so large, creation of science maps has proven a useful way to easily visualise patterns and trends in the networks. I will present a map of the terms that most frequently occur in the five articles of this dissertation to visualise the connections between them.

1.4.1 A brief history of bibliometrics

In one of the first bibliometric studies, F.J. Cole, a professor of zoology and Nellie B. Eales, a museum curator, counted the number of research papers on comparative anatomy published by authors in European countries between 1543 and 1860 (Cole & Eales, 1917). Originally intended to be a study of anatomical museums, the authors realised that the number of museums (537) was too small to draw statistical conclusions. They switched their attention to scholarly literature because research papers are permanent, accessible and it is easy to define who conducted the study and when (Cole & Eales, 1917). At about the same time, there was a shift in higher education from the general cultural education provided by the small community college towards the demand for the expert worker equipped with specialised knowledge acquired from large universities that shifted emphasis towards graduate education (Gross & Gross, 1927). Librarians had to quickly learn how to develop collections of relevant scientific periodicals to cater for their specialist departments and began using citation counts as a method of assessing the relative value of journals when building and maintaining their collections. Librarians at Pomona College in Claremont, California listed the number of references from a single volume of the Journal of the American Chemical Society to other periodicals in successive five-year periods and then ranked the cited journals in order of relative importance to the field (Gross & Gross, 1927).

Following these early works, historians continued using publication analysis to further their understanding of their field and librarians developed techniques to maximise the usefulness of their collections. However, it was the advent of 'big science' that led to the shift from library-based 'statistical bibliography' to the large-scale use of bibliometrics as a key tool for the policy maker (Narin & others, 1976). The need to constitute a coordinated set of measures related to books and documents was first described in French as 'La bibliométrie' (bibliometrics) by Paul Otlet in his *Traitée de Documentation. Le livre sur le Livre. Théorie et Pratique* (Otlet, 1934), and later Alan Pritchard's definition of the academic field of bibliometrics as 'applying mathematical and statistical methods to books and other media of communication' (Lawani, 1980; Pritchard, 1969) was adopted in English (Hood & Wilson, 2001). The closely related term 'scientometrics' is generally accepted to more broadly encompass measurement of any aspect of science and technology and originated from Vassily Nalimov's use of the Russian 'Naukometriya' (Nalimov & Mulchenko, 1969).

A defining moment came in 1955 when Eugene Garfield published his work showing how lists of cited references from scholarly papers could be used to create a network of scholarly publications in specific fields linked by citations (Garfield, 1955). Garfield then founded the Institute of Scientific Information (ISI) in Philadelphia and promptly launched the Science Citation Index. Counting up the citations to articles published in specific journals was considered a proxy for the journal's influence and became a counterweight to the opinions of science leaders. The Journal Impact Factor (JIF) was developed throughout the 1960s (Garfield, 1972), commercialised in 1975, and has played an important role in countless decisions by policymakers (Larivière & Sugimoto, 2019), although Garfield did not personally endorse its use as a performance indicator (Wouters, 2017). One argument against widespread use of the JIF as a performance indicator was the concern that the metrics weren't entirely accurate (Moed & Van Leeuwen, 1995) due to widely varying citation dynamics between journals for certain document types.

In an increasingly data-driven world, various national and multinational bodies began to turn their focus toward measuring and analysing progress of science and technology. The U.S. National Science Board's *Science Indicators* report, first published in 1973 (National Science Board, 1973), presented a system of science indicators to 'describe the state of the scientific endeavour'. According to the report, the indicators were intended to be adapted over time and become the basis for discovering trends and identifying areas of strength and weakness.

Across the Atlantic, the Europeans were keen to investigate potential applications of bibliometric analysis in strategic decision making. In 1980, Leiden University conducted the first detailed bibliometric analysis of 140 research groups in response to a decision by the University Executive Board to allocate funding partially based on research performance. The quantitative analyses were complemented by interviews with the researchers being studied and supported by peer review. Encouraged by this study, the university executive board supported the development of a group working on new and improved bibliometric techniques, which a decade later was formally named the Centre for Science and Technology Studies (CWTS) (van Raan, 2021).

There followed a rush to develop new bibliometric indicators for use in large-scale assessments of scientific communities to support policy making at various levels in many countries. Taking the Netherlands as an example, the Association of Universities in the Netherlands (VSNU) conducted a large-scale evaluation of over 200 academic physics programmes at each of the 13 Dutch universities as part of the Dutch national research quality assessment procedure. Each programme was judged by an international review committee that provided the basis for strategic decision making in research policy management (Rinia et al., 2001; Westerheijden, 1997).

In parallel, members of the Foundation for Fundamental Research on Matter (FOM) and CWTS at Leiden University conducted a complementary bibliometric analysis and compared the results with the VSNU expert committees (Rinia et al., 2001). The bibliometric analyses confirmed the results of the quantitative peer assessment to a statistically significant degree (van Raan, 1996). The bibliometric study also demonstrated there was no bias with respect to interdisciplinarity, i.e. peer judgement of physics papers published in non-physics journals show the same level of agreement with bibliometric indicators as physics papers published in journals within the physics field. Interestingly, for papers published by authors in the condensed matter physics programme, regression analysis showed the strongest consensus between peer judgement and article-level citation indicators and only weak agreement between expert opinion and journal-only indicators (Rinia et al., 1998). This was considered important evidence to support the role of bibliometric analysis in research evaluation. Meanwhile, the latter finding supported the notion that journal impact factors alone should not be relied upon to predict the quality of individual papers published in selected journals.

Soon after these pioneering bibliometric studies, a major change in global society occurred that would have far-reaching consequences in the development of bibliometrics – the World Wide Web was invented. The World Wide Web was designed to create a pool of human knowledge (Berners-Lee et al., 1994) and it immediately became clear that it offered huge potential to the scholarly community, especially in navigating forward and backward in time through articles related to each other by cited references.

1.4.2 Impact of the World Wide Web

It was sometimes said at ISI that Garfield's citation index waited 40 years for the World Wide Web to be invented. Indeed, soon after the advent of the World Wide Web, the Web of Science was launched, which comprised the Science Citation Index joined by the Social Sciences Citation Index, the Arts & Humanities Citation Index, and later others. Thousands of universities around the world paid subscription licences so their academics could navigate the citation links and discover papers relevant to their work. Journal editors and publishers recognised the value of being on the inside and still today routinely submit journals for scrutinization against the Web of Science selection criteria. Only those journals that pass the test will be indexed, while others are told they may re-apply in a couple of years. Once a journal is indexed in Web of Science (or in other similar citation indexes, in particular Scopus), its content becomes more easily discoverable to academics who have access to the index. Some publishers see value in this increased visibility and submit their journals to the various selection and inclusion processes set out by the owners of Web of Science and other similar citation indexes. As university ranking systems often use citation indexes as the basis for the bibliometric component of their ranking calculations, some universities incentivise their academics to publish in 'indexed' journals. From the beginning of citation indexing, it was known that journals in different subject fields exhibit different publication and citation behaviour and therefore need to be treated according to the norms in their field (Garfield, 1983). I will now discuss the classification of journals into subject areas.

1.4.3 Science classification systems

Science classification systems developed that revolved around journals as nodes in the system connected by citations between them (Small, 1993; Small et al., 1985; Small & Sweeney, 1985). In the Web of Science, the journals are classified into one or more of approximately 250 subject areas and Scopus has developed a similar system. In both cases, a scholarly paper is not directly assigned to a subject field, rather all the articles in any given journal are assigned to the subject field that the journal has been assigned to. That means the system is not very sensitive because journals are not always strict about the precise field they cover, and many papers address topics that cross boundaries between subjects or are multidisciplinary in their nature.

The inherent differences in publication and citation behaviour between subject fields means that the performance of academics in one field cannot be directly compared with those in another field using simple bibliometric indicators. The top listed journals in the medical sciences have higher Impact Factors than the top journals in the social sciences. Newer journal indicators were developed that take into consideration the context of the subject field in which researchers publish their papers. An example of such an indicator is the Source Normalized Impact per Paper (SNIP) (Moed, 2010).

Alternative classification systems have been developed whereby articles are attributed directly to subject fields based on their contents. Dimensions uses the article-level classification system which cuts out the need to use the journal as the unit of aggregation. In a comparison between journal-level subject classification in Web of Science and the article-level system in Dimensions the authors claimed that the journal-level system in Web of Science was more accurate (Singh et al., 2020).

1.4.4 Bibliometric datasources

In every bibliometric analysis, the data to be analysed has to be defined and collected. Initial studies used journals indexed in the Science Citation Index but following the launch of several other data sources bibliometricians have a choice. The choice of data source is important and can influence the results of any study. Indeed, comparison of bibliometric data sources is a frequently discussed topic within the field of bibliometrics. It is useful therefore to describe the available data sources and I shall use this section to introduce the main ones.

The Web of Science is made up of component databases. The following databases make up the core collection of Web of Science: Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index (Science and Social Science & Humanities editions), Book Citation Index (Science and Social Science & Humanities editions), and the Emerging Sources Citation Index. The databases in the core collection are usually used when conducting bibliometric analyses based on the Web of Science. However, there are also regional databases including the Arabic Citation Index and Chinese Science Citation Database, subject-specific indexes such as the BIOSIS Citation Index, and indexes of non-journal literature such as the Data Citation Index.

In 2004, the publisher Elsevier launched a rival bibliometric database, Scopus, by combining several existing databases including Medline, Compendex, and Embase. This created a competitive market for access to global, multidisciplinary discovery services and bibliometric data that continues to this day. Web of Science and Scopus have each developed a coverage policy that determines which journals, conferences, and books they include in their databases. Web of Science comprises publications from high impact journals and conferences selected by an internal editorial team and focuses on quality (Birkle et al., 2020). Scopus is more inclusive, comprising publications from over 39,000 serial titles and relies on an external content selection and advisory board (Baas et al., 2020). Although there is considerable overlapping content between Web of Science and Scopus, important distinctions remain due to differences in coverage policy, indexing practices, subject field definition, publication date discrepancies, and differences in author name disambiguation, affiliation unification, and document type assignation.

Also in 2004, Google launched its own citation resource known as Google Scholar with almost universal coverage, no date limitation, and the huge advantage that it was freely available. That opened up access to millions of academics all over the world, especially those not working or studying at a university with a Web of Science or Scopus subscription. Google Scholar has become widely used as an Internet discovery tool, but not as a data source for bibliometric studies, mainly because the underlying data is not made available. Google prevents automatic retrieval of its data using technology such as captcha that can only be solved by humans (Else, 2018). Another consequence of Google Scholar's inclusivity is that it may not be sensitive to the differences in output quality as determined by selective databases (Diem & Wolter, 2013) and therefore not an ideal source for evaluation.

In this context, Digital Science launched Dimensions in 2018. Dimensions seeks to harness the best of both worlds (Hook et al., 2018). Largely based on Crossref, Dimensions is more inclusive than Web of Science and Scopus, offers a limited free version, and extends the bibliometric database to cover a greater portion of the research lifecycle. Dimensions comprises data on research inputs such as grants, along with outputs including datasets, publications, and patents, also links outcomes such as social media impact, citations, policy documents, and clinical trials (Herzog et al., 2020).

Crossref is an organisation that generates digital object identifiers (DOIs) and assigns them to academic publications. Scholarly publishers can become Crossref members, which entitles them to upload article metadata from their journals into the Crossref database, whereupon each article is registered to a unique DOI (Collins, 2022). Members are encouraged to link the articles' URL and other metadata with the DOI and update the URL links whenever the article is moved. Regardless of where the document is posted on the Internet, the DOI remains the permanent and unique identifier for the document. Currently, Crossref comprises metadata for more than 100 million records and is adding metadata over 10 million records per year. This metadata is openly available, which makes it an attractive source for bibliometric study (Hendricks et al., 2020).

Following some high-profile cases in which journal selection policy was implicated in producing suspicious university ranking results, Microsoft launched Microsoft Academic Services (MAS) as a fully open scientific knowledge graph (Sinha et al., 2015). MAS collects data from the entire public web and includes studies at all stages of publication to avoid sampling bias and the associated potential impact on bibliometric studies (Wang et al., 2020). MAS does not use DOIs or ORCIDs because of the view that such technical standards are not well used by the scientific community and fail to live up to their promised consistency (Wang et al., 2020). A large-scale comparison between major databases showed MAS had far broader coverage than Web of Science, Scopus, Crossref, and Dimensions (Visser et al., 2021; Wang et al., 2020).

Microsoft Academic and related services were discontinued in January 2022 and OpenAlex was launched to fill the void. OpenAlex is named after the world's first library in Alexandria, Egypt and aims to improve transparency of research evaluation (Piwowar et al., 2022; Priem et al., 2022) by linking scholarly works, authors, venues, institutions, and concepts. The main data sources are the discontinued Microsoft Academic and Crossref along with preprint and data repositories such as arXiv and Zenodo. OpenAlex currently comprises 243 million works (OpenAlex, 2023). While some of the current data sources remain entirely or partially behind a paywall, OpenAlex is a fully open data source with open API and open source code (Priem et al., 2022).

There are many other bibliometric datasources available, and this section cannot mention them all. Some newer global, multidisciplinary sources have not been included in this overview and there are a multitude of region-specific and discipline-specific data sources available. Those mentioned above should be taken as prominent examples of bibliometric data sources in a growing field, rather than an exhaustive review.

1.4.5 Applied bibliometrics – University rankings

Journal metrics such as the Journal Impact Factor have been used to rank journals according to their citation impact leading to competition among publishers to maximise the performance of their journals. In the last 20 years, bibliometric indicators have been used to rank the performance of universities in published league tables or university rankings.

In 2003, the Academic Ranking of World Universities launched the Shanghai Ranking, which quickly stimulated lively debate on the methodology used and its validity (Liu & Cheng, 2005; Van Raan, 2005). A paper entitled *Should you believe in the Shanghai Ranking?* comprises a scathing review of irrelevant criteria, a methodology 'plagued by major problems', and a poorly structured system (Billaut et al., 2010).

The following year, Times Higher Education launched the World University Rankings in cooperation with QS, which eventually split into two competing ranking systems. Each of these ranking systems has developed its own methodology based to varying degrees on bibliometric indicators alongside university-provided metrics such as faculty-student ratio, and surveys that quantify the opinions of the scientific community on the prestige of top institutions. More ranking systems were launched, and fierce competition ensued between bibliometric data providers to become the trusted database partner of the most popular rankings.

Controversy followed (e.g., Gadd, 2020) and moderate critics of university rankings proposed an alternative system in which universities are listed with a range of bibliometric and other indicators and the rank can be changed by the user depending on the criteria they select. A prime example of this is the Leiden Ranking (Centre for Science and Technology Studies, 2022; Moed, 2017; Waltman et al., 2012). In the absence of an overall or definitive ranking, this system can also be known as university profiling.

1.4.6 Community-led initiatives to improve use of bibliometric techniques

The argument about interpretation and use of the Journal Impact Factor raged for many years, until in 2013, following a meeting of the American Society for Cell Biology, 150 scientists and 75 organisations signed a statement calling for an end to misusing metrics to evaluate impact of individual articles and the performance of those who wrote them. At the time of writing this dissertation, the declaration has been translated into 30 languages and signatories number more than 20,000 individuals and 2,850 organisations (*DORA*, 2023). Among the signatories it is likely that a substantial proportion have signed to create the image that they care about responsible use of metrics but continue using the JIF just as before (Torres-Salinas et al., 2023). However, there was no organised resistance to the DORA declaration and no widely communicated counterargument. Even the company that publishes the JIF, Thomson Reuters (now Clarivate) was relieved that the broader scientific community was finally questioning misuse of quantitative indicators and was willing to engage in discussions on newer, more customised impact indicators.

As access to bibliometric databases such as Web of Science and Scopus increased and the owners of these databases launched their corresponding analytical tools, InCites and SciVal, organisations all over the world increased their reliance on bibliometric indicators. The university ranking systems, partially based on publication and citation metrics, gained further popularity to the point where in some parts of the world, ranking methodologies have influenced, or even de facto replaced, university strategy. In 2015, the Leiden Manifesto (Hicks et al., 2015) set out ten principles that describe best practices for metrics-based research assessment.

Similarly, in July 2022, the Coalition for Advancing Research Assessment (CoARA), started by the European University Association representatives along with Science Europe and the European Commission, published the agreement on reforming research assessment (ARRA). Widespread concerns about assessment based on oversimplified metrics have incentivised the European research community to call for an agreement between those being assessed and those conducting the assessment to lay some ground rules (CoARA, 2022b). The agreement envisions assessment of research, researchers and research organisations which is necessarily performed through qualitative peer assessment and supported by responsible use of quantitative indicators.

1.4.7 Network visualisations

The enormous growth in sheer numbers of journals, articles, and academics has created challenges for people tasked with evaluating science and scientists. Faced with large numbers of documents and people, it is sometimes easier to detect links between them when publication networks are presented visually as maps. One bibliometric mapping tool, VOSviewer, is specially designed to enable users to create maps based on network data and to visualise and explore the maps (Waltman & van Eck, 2012). The user can choose which attributes are presented on the map, for example author names, affiliations, or keywords. Chapter 3 of this dissertation compares methodological approaches to identifying research related to the UN sustainable development goals (SDGs) and uses VOSviewer maps to illustrate the difference between clusters as related to climate change using keywords from the title and abstract of the papers.

The opportunities of the Internet combined with the rapidly growing quantity of scholarly articles made network visualisation attractive to bibliometricians. Every research paper can be a node on a map and citations can serve as the links between them. Alternatively, papers can be represented by the keywords mentioned in the title, abstract, or full text and term maps present words more prominently the more frequently they are mentioned. As an example, in Figure 1.1, the most frequently occurring terms in the five articles in this dissertation are presented using VOSviewer. The terms fall roughly into five clusters related to topics studied in the articles and denoted by the different colours. The terms in purple, such as *country*, *country affiliation*, *country name*, and *population,* are all related to geodiversity, while *conference*, *asean country*, and *Indonesian paper* land in the blue cluster because they are related to the work on policy and behaviour. Some of the terms appear to fall between clusters, such as *SDGs*, *action*, and *climate,* which are somewhere between the red and green clusters. The terms indicate a close relationship between the topics of stakeholder collaboration and SDG classification method. Terms at the centre of clusters, such as *affiliation discrepancy* (data quality) and *triple helix* (stakeholder collaboration), are specific to their cluster and are only weakly related to the other topics.

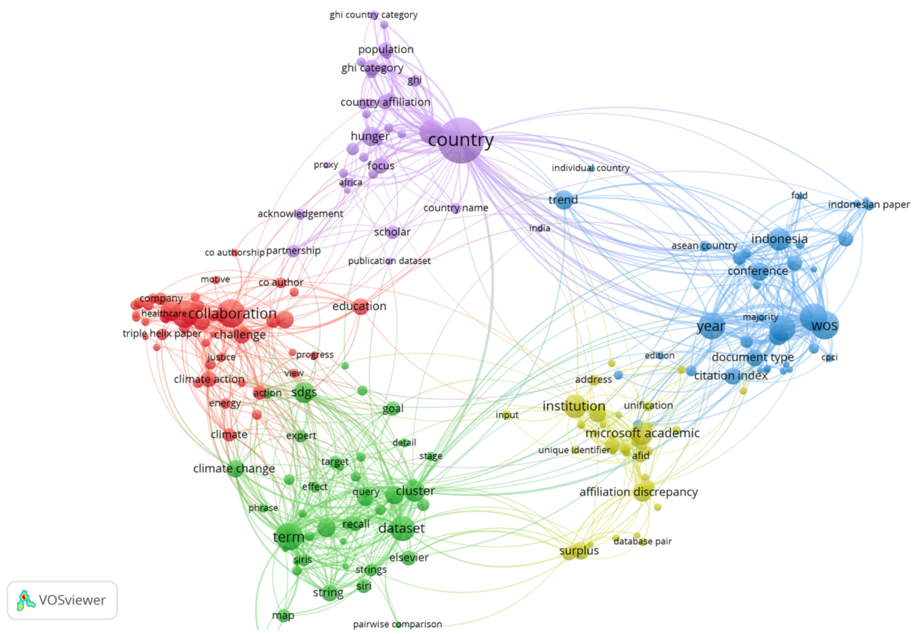


Figure 1.1 VOSviewer term map using the full text of the five chapters of this dissertation. Click here for live navigation

It is easy to see the relationships between concepts described above because we are looking at a map. The analysis of large numbers of terms is simplified through visualisation.

1.5 Motivation and objectives of this dissertation

1.5.1 Motivation

There is a need for clearer guidance on how bibliometrics can be used to support research policymaking and research evaluation. Parts of the scientific community support the use of quantitative indicators in research policymaking and evaluation. Supporters of the bibliometric approach cite its usefulness in relatively quickly and cheaply analysing large amounts of data and reliance on documented methods and indicators. Critics of the quantitative approach point to problems with bibliometric data accuracy and consistency and claim that only fellow academics are qualified to judge research quality. The drawbacks of both peer review and bibliometric approaches require improvement to gain support for use in policymaking and research evaluation.

In reality, neither approach on its own is likely to be ideal for most scenarios and the third option is to use a combination of quantitative bibliometrics and qualitative expert judgement. However, bibliometric data, methods, and approaches are not yet reliable enough to win the overwhelming trust of the broader scientific community. Data quality and curation need to be improved, and the community needs to play a greater role in the development of new methodological approaches and technologies. This dissertation addresses these points through five bibliometric studies that show both the power and the limitations of the quantitative approach. Areas for improvement are identified and solutions proposed that will improve the reliability, trust, involve the community, and provide better guidance to stakeholders on the use of bibliometrics in the context of research policy and research evaluation.

1.5.2 Objective

I have picked five areas that show different ways in which bibliometrics can support research policymaking and research evaluation. The objective is to illustrate a range of scenarios where large-scale, quantitative bibliometric studies offer benefits to evaluators and policymakers. The studies also highlight the limits to bibliometrics and alert stakeholders to potential problems caused by misusing bibliometrics. The overall ambition of this dissertation is to present a series of recommendations for research policymakers on how to employ and interpret bibliometric techniques, and to provide suggestions to the wider scientific community on how bibliometric data, processes, and methods can be improved.

1.6 Structure and research questions

There are many areas in which quantitative research evaluation and research policy interact. The next part of this dissertation comprises five studies that address the objectives described in section 1.5.2. Chapters 2 and 3 present studies that illustrate limitations in data quality in bibliometric databases and subject classification techniques and cautions policymakers against over-reliance on the results of university analyses or rankings. Chapter 4 comprises a large-scale bibliometric study on the academic sector's transdisciplinary collaboration with societal stakeholders. Chapters 5 describes the geodiversity of research and discusses the extent to which bibliometrics can uncover unfair collaboration practices in developing regions of the world. Finally, chapter 6 illustrates a clear case of bibliometric evaluation causing a questionable change in behaviour of a national scientific community. The chapters address the following research questions:

1.6.1 RQ1: How does bibliometric data quality impact research policy?

Chapter 2 addresses the research question on data quality by illustrating wide-ranging discrepancies between four major bibliometric databases in their unification of author affiliations, an important type of bibliometric data. Author affiliations are used to link publications to universities or other research institutions, and it is therefore crucial to unify as many variations of the organisation name as possible. Each of the databases in this comparison uses a different unification system, which means they each produce a different publication set for the same university.

The findings draw attention to the limits in reliability of any university evaluation or ranking system. Understanding of the limitations of bibliometric databases and the discrepancies between them should be a prerequisite for policymakers before they design bibliometric studies of universities or act on their results. This study therefore presents author affiliation discrepancies between databases as an example of how data quality can influence the outcome of policy goals. The chapter concludes by expressing support for a single, community-led affiliation unification initiative such as the Research Organization Registry (ROR). ROR uses a unique identifier for universities in a similar way to the DOI for publications and ORCID for authors. Universal adoption of such a system for universities would channel community efforts into improving data quality and reduce discrepancies between databases.

1.6.2 RQ2: How can bibliometrics help improve the categorisation of sustainability research?

The research question on categorising new research fields is answered in chapter 3 through a comparison of four approaches (Elsevier, Digital Science, STRINGS, and SIRIS) to classifying research publications related to one of the sustainable development goals (SDGs) defined by the United Nations: SDG 13 (climate action). The analysis explores the different steps used by each method and suggests explanations behind the low agreement between the methods.

Sustainability policy has filtered down from the UN to government and university level and pressure is building for entities to report on their contribution to progress against the SDGs. Academic and research institutions have a need to produce bibliometric reports on their publications that show productivity, impact, and collaboration. Sustainability had not been delineated as an academic subject until recently and all the major bibliometric database owners, as well as academic groups, are racing to develop new methods of identifying research papers related to each individual goal.

The SDGs are multidisciplinary, complex, and vague, which means there is no precedent, no pre-existing categories, and no 'ground truth'. Each of these approaches produces different pictures of what an SDG publication dataset should look like, which leaves policymakers wondering which one is right. Chapter 3 contributes to the debate among policymakers on the extent of progress against the SDGs.

1.6.3 RQ3: What can bibliometric analysis tell us about transdisciplinary research collaboration trends?

The research question on transdisciplinary research is addressed in chapter 4 by using bibliometric data to define research output for the academic sector and three major societal stakeholders: government, industry, and nonprofit organisations. The study collects papers published by the academic sector over a 10-year period and presents the changing share of collaborative papers with each of the three societal stakeholders. Given the diverse nature of these actors in different regions of the world, the study then presents the results for the countries with the largest scientific output. This enables us to discuss the findings in the context of transdisciplinary research at national level.

It is noted that different stakeholders often have different motives for collaboration, which could actually inhibit joint research projects. The goal of academic researchers is almost always to publish their results to share knowledge, foment discussion with fellow scholars, contribute to solving problems, and as a means to advancing their careers. Industry on the other hand, is ultimately driven by bringing new products and services to market, which sometimes requires research findings to be kept secret and other times only to make their findings publicly available by patenting their inventions, rather than publishing them in academic journals. These conflicting goals may act as inhibitors to large-scale academia – industry collaboration. Differences between academia and government research focus may produce similar conflicts.

The relationship between stakeholders may therefore be lopsided and limit the amount of collaborative research conducted. Therefore, policies that simply incentivise 'more inter-stakeholder collaboration' might be too simplistic. Research policymakers around the world are calling for closer collaboration and highlighting the perceived benefits of transdisciplinary research, but data on progression is scarce.

Chapter 4 therefore contributes to the current literature on transdisciplinary research by providing feedback to policymakers on the success of their efforts.

1.6.4 RQ4: How can bibliometric techniques be used to describe the geodiversity of research?

The study presented in chapter 5 offers insights into the geodiversity of research through bibliometric analysis of the relationship between research focus and author locations. The study uses SDG 2 (zero hunger) as a case study and uncovers wide variation between geographical topic focus and author location depending on the level of hunger in the country of focus.

Chapter 5 discusses policies emerging among publishers and academic societies that seek to ensure fair and equal partnership between local academics in countries severely affected by hunger and visiting researchers from wealthy countries. Such policies deal with local capacity building, treatment of samples, and the order of co-authors on the resulting publications. This is a case of the publishing community setting policies that could influence academic behaviour and be visible through bibliometric reporting.

Bibliometric studies can help identify the existence of problems through illustration of geographical topic focus and author affiliations using mapping visualisations. Country representation and author position can contribute to our understanding of collaboration dynamics and time series will enable communities to follow progress through trends. Through bibliometric studies, policy makers can identify the existence and extent of problems and monitor progression once steps have been taken to address them.

1.6.5 RQ5: How can bibliometric assessment cause inadvertent behavioural change in the scientific community?

The research question on the effects of bibliometric evaluation on academic behaviour is addressed in chapter 6 through a bibliometric study on published document types in 10 countries in Southeast Asia over a 20-year period. The findings revealed a national change in publishing behaviour specific to Indonesian academics and linked it to the introduction of a specific research policy. The Indonesian government introduced incentives into its national research policy with the intention of stimulating a general increase in academic publishing. However, the academic community responded to the policy by switching from publishing articles in journals to publishing papers in conference proceedings. Conference papers were perceived to be quicker and easier to publish than journal articles and therefore conference publishing was perceived to be a preferable route towards achieving the policy incentive.

This chapter highlights the potential consequences of using a metric as an incentive. The Indonesian study is a classic case of Goodhart's Law, which states that when a metric becomes a target, it ceases to be a useful metric (Goodhart, 1975). The chapter draws parallels between the Indonesian case and the international university ranking systems. Universities in many parts of the world see a rise in the rankings as a mark of success. As academic publications and citations contribute to higher ranking, bibliometric indicators become targets.

Chapter 6 discusses the implication of these findings in that policymakers overlooked the difference between the characteristics of journal articles and conference papers when developing the research policies. The scientific community realised this and accelerated submissions of papers to both national and international conferences whose proceedings would be indexed in bibliometric databases. This is an example of policymakers using bibliometric data to stimulate one change, but inadvertently causing another.

Chapter 2

The prevalence and impact of university affiliation discrepancies between four bibliographic databases – Scopus, Web of Science, Dimensions, and Microsoft Academic[1]

[1] This chapter is based on:

Purnell, P. J. (2022). The prevalence and impact of university affiliation discrepancies between four bibliographic databases—Scopus, Web of Science, Dimensions, and Microsoft Academic. Quantitative Science Studies, 3(1), 99–121. https://doi.org/10.1162/qss_a_00175

Abstract

Research managers benchmarking universities against international peers face the problem of affiliation disambiguation. Different databases have taken separate approaches to this problem and discrepancies exist between them. Bibliometric data sources typically conduct a disambiguation process that unifies variant institutional names and those of its sub-units so that researchers can then search all records from that institution using a single unified name. This study examined affiliation discrepancies between Scopus, Web of Science, Dimensions, and Microsoft Academic for 18 Arab universities over a five-year period. We confirmed that digital object identifiers (DOIs) are suitable for extracting comparable scholarly material across databases and quantified the affiliation discrepancies between them. A substantial share of records assigned to the selected universities in any one database were not assigned to the same university in another. The share of discrepancy was higher in the larger databases, Dimensions and Microsoft Academic. The smaller, more selective databases, Scopus and especially Web of Science tended to agree to a greater degree with affiliations in the other databases. Manual examination of affiliation discrepancies showed they were caused by a mixture of missing affiliations, unification differences, and assignation of records to the wrong institution.

Keywords

Benchmarking, affiliation, disambiguation, unification, bibliometric database, university

2.1 Introduction

2.1.1 The problem of affiliations

The research community understands to varying degrees the importance of getting its university affiliation names right. Individual researchers are now routinely assessed at least in part on their ability to produce published articles and their institutions are at least partially ranked on those very same papers. The single factor tying the paper to its author's employer is the affiliation name given by the author when they submit the manuscript to a journal. There are many ways of acknowledging a university and one can easily confuse a rudimentary database by simply swapping My City University with University of My City. Other common variations involve acronyms, MCU, UMC, or partial acronyms, MC University, Univ MC. The list easily extends to dozens of variants when authors introduce their faculty or department name sometimes at the expense of the university name. Add to that the common practice of incorporating one institution into another or splitting part of a university away from its main organisation, along with larger mergers and creation of international branch campuses and we have a complex problem for those assessing the university's research output.

Indeed, nowadays journal and author names are relatively constant while it is not uncommon for university names to change. Although there have been several initiatives to address the problem by using unique identifiers for research institutions, none have been universally adopted to the same extent as for journals (ISSN), individuals (ORCID), or documents (DOI). These efforts, summarised in Table 2.1 have mainly been made by the major citation indexes such as Scopus (Affiliation Identifier or AFID), Web of Science (Organisation Enhanced), and Dimensions (Global Research Identifier

Database or GRID). In addition, a new community-led collaboration of multiple organisations has been launched and is known as the Research Organization Registry (ROR). It holds promise because it is closely linked to GRID and is to be incorporated into Crossref metadata (Lammey, 2020).

Table 2.1 Disambiguation processes used by bibliographic databases

Database	Disambiguation	Abbr.	Process
Scopus	Affiliation identifier	AFID	Institutions assigned an 8-digit AF-ID and variants linked to main AFID
Web of Science	Organisation enhanced	OE	Unifies the most frequently occurring address variants to preferred names
Dimensions	Global research identifier database	GRID	Freely accessible database of research organisations that are assigned a unique and persistent identifier linked to its variants.
Microsoft Academic	Global research identifier database	GRID	Freely accessible database of research organisations that are assigned a unique and persistent identifier linked to its variants.
Community led registry	Research Organization Registry	ROR	Community-led initiative to supersede GRID and be incorporated into Crossref metadata

Databases used in bibliometric assessments have made strides into resolving the problem of disambiguation using different solutions including manual submission of affiliation variant lists by universities to database owners or automated unification systems. The degree of accuracy is still unquantified and policy makers who rely on bibliometric analysis often overlook an inherent level of error when using these data sources.

2.1.2 The importance of affiliations

Initiatives to identify and list the world's most influential academics based on citations to their work rely on affiliation disambiguation techniques. Clarivate's Highly Cited Researchers list recognises approximately 6,000 scientists who have published papers cited in the top 1% of their field in the preceding decade. The list is published once a year and is searchable by academic institution. The composition and validity of the list is therefore dependent on the affiliation disambiguation performed by Clarivate on its underlying Web of Science database. Similarly, the recently updated 'Stanford' author database of standardised citation metrics (Ioannidis et al., 2020) relies on Scopus affiliation disambiguation. The list of top 2% authors can be downloaded and is sortable by academic institution. As universities seek to recruit scientists who appear on these prestigious lists, an academic's value is increased by virtue of their presence. The ability of the databases to accurately link authors to their affiliations is therefore increasingly valued by recruitment professionals as well as research managers.

Many universities driven by increased external competition (Brankovic et al., 2018; Espeland & Sauder, 2016) seek to maximise their position in the various international ranking tables. The ranking organisations in turn typically assess institutions' performance against a set of criteria that usually include the quantity and impact of research publications (Centre for Science & Technology Studies Leiden University, 2020; QS Intelligence Unit, 2019; Shanghai Ranking Consultancy, 2019; Times Higher Education, 2021; US News & World Report LP, 2019). These ranking systems use either Elsevier's Scopus or Clarivate's Web of Science (Web of Science) to compute the bibliometric component of their tables. Nature Index, a database of author affiliations and institutional relationships also recently used data from Dimensions in an experiment to identify research in the field of Artificial Intelligence (Armitage & Kaindl, 2020).

The level of accuracy of those databases and their ability to assign papers to the correct affiliations consequently becomes one of the limiting factors in an institution's performance (Orduna-Malea et al., 2019). Any 'missing' papers can cost a university valuable places in the ranking table and authors are routinely encouraged to use the official institution name when publishing. Nevertheless, each year universities complain to the ranking systems of missing papers, but the rankers are constrained by the limitations of the citation indexes. Disgruntled universities are usually referred to the database owners to resolve their affiliation-related complaints.

Times Higher Education routinely uses the Scopus affiliation as delivered by Elsevier although it has occasionally worked directly with institutions to ensure the mapping used in the ranking coincides with their organisational structure. QS receives data from Scopus and then groups distinct Scopus AFIDs including medical schools, business schools, hospitals, and technical research institutes into single university entities that match those in its rankings database. In this process, QS relies on input from the institutions to define such relationships (QSIU, 2019). The Leiden Ranking uses its own (CWTS) version of the Web of Science and conducts additional rounds of affiliation disambiguation. Specifically, the CWTS system unifies all address variants that occur at least five times in the Web of Science database, identifies missing university affiliations from departments and city names, attributes papers from hospitals and medical centres to their affiliated universities based on author publication rules (Centre for Science & Technology Studies Leiden University, 2020; Van Raan, 2005; Waltman et al., 2012).

Comparison of database coverage has become easier since most now aim to link their records to the Digital Object Identifier (DOI). The DOI has become established as a persistent, reliable identifier together with a system using that identifier to locate digital services associated with that content (e.g., Gasparyan et al., 2021; Zahedi et al., 2017). In bibliometric studies, the DOI can therefore be used as a common identifier to determine overlapping coverage of records in different databases. The existence of a DOI for a research article depends on the publisher generating the DOI and linking it to the article in the Crossref database. Most publishers now aim to do this routinely but there are plenty of records that do not have a DOI and that limits our ability to use it as a common identifier. Factors associated with lower prevalence of DOIs include non-academic records, arts and humanities fields, document types from books and conference proceedings. This study used DOIs to retrieve papers from selected universities and it was important as a first step to understand what proportion of the actual output we were looking at.

2.1.3 Research design

This paper addresses affiliation disambiguation by attempting to answer the following questions:

- To what extent do discrepancies in author affiliations exist between the major bibliographic databases? Which databases have most discrepancies?
- What types of discrepancies can be identified?
- Are different types of discrepancy more prevalent in different databases?

The answers to these questions will be useful in our understanding of the extent to which research outputs are accurately assigned to universities by the different databases. Since many decisions are based on the outcome of bibliometric studies, comparisons, and university rankings, policy makers will be better informed about the limitations of bibliometrics studies and comparisons. The ranking bodies may take these limitations into account when they publish their league tables. Database owners may incorporate these findings into their development plans and algorithms to improve the accuracy of their products and make them more competitive.

We selected 18 universities for the study, all from the Arab region. Local or regional languages have been found to contribute to mistakes in author affiliations (Bador & Lafouge, 2005; Falahati Qadimi Fumani et al., 2013; Konur, 2013) and to our knowledge, no such study of university affiliations in the Arab region has been published. We selected universities from Gulf countries, the Levant, and North-East Africa because of our familiarity with the region, language, and institutions.

We used a recent five-year time window (2014-18) and extracted records from four major databases, Scopus, Web of Science, Dimensions, and Microsoft Academic.

The first part of the study was to determine the proportion of publications from our selected universities that had DOIs. Records were extracted using DOI and compared with the total number of records for the selected publication years. This was to confirm we could use the DOI to identify a sufficient quantity of scholarly material.

Each database has taken a different approach to affiliation disambiguation, and it is important to know how they compare. Therefore, in the second part of the study we quantified the affiliation discrepancies between databases. We paired the databases and specifically looked at the non-overlapping records i.e., the surplus of records that were retrieved from one database in the pair but not the other. Records could be in the surplus because of discrepancies in affiliation or publication year, or differences in database coverage. We quantified the surplus in each database with respect to the other three for all the universities. We then calculated the proportion of the surplus that was caused by each of the three reasons (affiliation discrepancy, publication year discrepancy, or coverage differences). Discrepancies due to publication year were negligible and grouped together with coverage differences.

The third part of the study concentrated exclusively on those records that were caused by discrepancies in the university affiliation. We manually examined two dozen records from each database pair surplus and attempted to explain the discrepancies in the context of affiliation indexing. This has major implications for any university benchmarking study and particularly the international university rankings. The final section of the study requires close knowledge of university names and we therefore chose to make this a regional study.

2.2 Literature review

Work on author affiliations began in the mid-1980s. For example, the LISBON Institute, the predecessor of CWTS already used affiliation data from the Science Citation Index (SCI) to report the changes in academic collaboration in the Arab region following the geopolitical developments in the 1980s (DeBruin et al., 1991). The problems of author affiliations became more important as bibliometric reports gained popularity and started being offered as a service (Calero-Medina et al., 2020). As in-house citation indexes were developed in the 1990s, care was taken to construct them in a manner that facilitated affiliation disambiguation. The problem of missing author affiliations has been largely addressed but in 2015 the Web of Science indexes still contained sizeable quantities of publications without any affiliations whatsoever (SCIE: 7.6%, SSCI: 6%, and A&HCI: 35%) (Liu et al., 2018).

The owner of Web of Science tried to overcome the affiliation disambiguation problem by introducing its organisation enhanced feature. The disambiguation process involves creating normalised address segments for each record and unifying the most frequently occurring address variants to a list of currently over 14,000 preferred names (Clarivate, 2020b). Smaller or less frequently occurring organisations might not have a preferred name in the system or some of its legitimate address variants might not have been unified. Full details of the process are not available, but one imagines it a significant task to perform and then keep up to date in the light of organisation mergers and divestments. Organisations may request unification or corrections to its address variant list via an online form (Clarivate, 2020a).

The organisation enhanced feature has been selectively used in bibliometric studies to improve accuracy (Baudoin et al., 2018). A recent study (Donner et al., 2020) showed widely varying recall and precision across institutions between the Web of Science organisation enhanced feature, Scopus AFID and a German institution affiliation disambiguation system described as 'near-complete' for German public research organisations. Taking the German system as ground truth, neither Web of Science organisation enhanced nor Scopus AFID provided high recall rates, and both showed widely varying precision across institutions impacting bibliometric indicators. The authors concluded the resulting inconsistencies in publication and citation indicators using the commercial vendor systems should be taken into consideration by policy makers.

A large-scale study comparing database coverage across a multi-institution dataset between Scopus, Web of Science, and Microsoft Academic examined the publication overlap of 15 universities with DOIs serving as the common attribute (Huang et al., 2020). The authors created a Venn diagram for each university showing the proportion of DOIs indexed by all three data sources. The diagram also revealed the extent to which DOIs were covered by only one of the three databases, which we term a surplus. For instance, DOIs found in Web of Science but not in Scopus or Microsoft Academic count as Web of Science surplus. The authors found that Microsoft Academic had the broadest DOI coverage but least complete affiliation metadata. The authors concluded that assessment of any institutional performance will produce different results depending on the source database. They went on to demonstrate that representation in databases varied widely between institutions, which compounds the likelihood of inaccuracies when comparing universities with each other. This introduces potential bias in any bibliometric assessment such as university ranking that relies on a single source. The present study builds on the work by Huang et al. (2020) by further analysing the affiliation assignation of DOIs across the same three databases plus Dimensions for 18 universities in a specific geographical region.

Another large-scale study (Visser et al., 2021) compared coverage and the completeness and accuracy of citation links between five databases, namely Scopus, Web of Science, Dimensions, Microsoft Academic, and Crossref. The authors used a thorough and complex matching technique that compared pairs of records against seven bibliographic attributes and computed a matching score. They found Microsoft Academic had by far the broadest coverage of the scientific literature. In pairwise analysis, Scopus contained large quantities of exclusive publications not indexed in the other databases. Meanwhile, Dimensions/Crossref and Microsoft Academic indexed substantial bodies of records not found in Scopus. The Web of Science surplus compared with Scopus was smaller and limited mainly to meeting abstracts and book reviews which are not indexed in Scopus. Another important finding was that Dimensions and Crossref had almost overlapping coverage, which can be attributed to Dimensions' reliance on Crossref as its core content.

A comparative analysis at institutional level found more institutions had greater coverage in Scopus than in Dimensions, and that up to half the documents indexed in Dimensions have no institutional affiliation (Guerrero-Bote et al., 2021). In the view of the authors, this invalidates Dimensions as a suitable data source for assessing university impact.

2.2.1 Limitations of DOI accuracy

All studies using DOI are susceptible to a number of limitations. First, DOI assignation is not always accurate. Indeed, one group found errors in 38% of the DOIs in the cited references of their sample from Web of Science (Xu et al., 2019). Most (92%) of the errors in the DOI were in the prefix and often included a surplus 'DOI' or a duplication of the DOI. The authors went on to propose an algorithm for cleaning the DOIs in the cited reference database. Another study found 8,841 'illegal' DOIs (defined as those that do not begin with '10') in Scopus (Huang & Liu, 2019) and referred to Elsevier's efforts to clean the Scopus data.

Zhu et al. (2019) created a search string to identify DOIs with errors in Web of Science. They wrote a search strategy aimed at identifying cases in which a numeric digit such as '0' had been replaced with the letter 'o'. Similar errors occurred where the number '6' had been confused with the letter 'b', or the number '1' with the letter 'l' among many other examples identified. In some of these cases, searching both the correct version of the DOI (with the numeric '0') or the erroneous version (with the letter 'o') returned the Web of Science record.

Another problem is duplicate DOIs, this is when a single DOI is linked to multiple papers as reported by Franceschini et al. (2015) or to multiple versions of the same publication (Valderrama-Zurián et al., 2015). Databases have taken different approaches to this problem and Elsevier has recently invested in improving the Scopus data completeness and accuracy (Baas et al., 2020).

The DOI was introduced in 2000 although there have been a number of other unique identifiers for published research papers. With the advent of electronic publishing, the Uniform Resource Locator (URL) was an early candidate. A study of more than 10,000 MEDLINE abstracts in 2008 looked at the presence of decay rate of URLs between 1994 and 2006 (Ducut et al., 2008). The results showed that most (81%) of the URLs were available but that only 78% of the available URLs contained the actual information mentioned in the MEDLINE record, and one in six (16%) of the total were "dead" URLs. A study comparing multiple identifier systems found the DOI to be among the best following evaluation against seven criteria including identifier features, digital coverage, and comprehensiveness of scope (Khedmatgozar & Alipour Hafezi, 2015).

Over time, publishers incorporated DOIs into their online metadata. An examination of Web of Science and Scopus records revealed that by 2014, most (90%) of 'citable items' defined as journal articles, reviews and conference proceedings papers were being assigned DOIs in the sciences and social sciences (Gorraiz et al., 2016). The figures were lower for all document types and much lower in the arts & humanities, 50% for journal citable items and just 20% for books and book chapters. Articles published in regional publications may also be less associated with DOIs, a sample of scholars from Brazil showed DOIs among less than half of their journal papers and less than a tenth of their conference papers (Rubim & Braganholo, 2017). Meanwhile, Mugnaini et al. (2021) found the presence of international co-authors increased the proportion of DOIs.

2.3 Data and methods

This study examines the overlap of indexed DOIs for 18 universities (Table 2.2) of all document types, in four international, multidisciplinary bibliometric databases often used in bibliometric studies, namely Scopus, Web of Science, Dimensions, and Microsoft Academic.

From the Web of Science, we used five citation indexes; the Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, the Conference Proceedings Citation Index-Science, and the Conference Proceedings Citation Index-Social Sciences & Humanities. We used neither the Book Citation Index, nor the Emerging Sources Citation Index because we do not have access to them. From Dimensions, we extracted only publications because these are comparable with the documents in the other databases, but not grants, patents, datasets, clinical trials, or policy documents. All data were retrieved from the Centre for Science and Technology Studies (CWTS) database system at Leiden University. The data was received in March 2020 (Web of Science), April 2020 (Dimensions), July 2020 (Microsoft Academic), and April 2021 (Scopus).

We extracted records based on the highest level of affiliation disambiguation available for the selected universities in each database. In Scopus, we used the Affiliation identifier (AFID) which is a unique identifier for the institution to which records are tagged. In many cases, Scopus includes one AFID for the overall organisation and multiple additional AFIDs for partner institutions and component units to reflect the organisational structure of the university. This enables the user to search for records from the entire organisation or for its sub-units individually. This process appears to have been conducted more rigorously for some universities than others. The result is that for some universities there is only one AFID while for others there are many and we therefore used only the main AFID for each university. From Web of Science, we used the organisation enhanced feature which is a preferred institutional name searchable in the database and to which records from that organisation and its component parts are unified. The organisation enhanced unification process is performed by the database owner with voluntary input from the institutions themselves and suffers from the same inconsistencies as the Scopus AFIDs. Therefore, we used only the high-level organisation enhanced name and no additional variants. Dimensions and Microsoft Academic each use GRID which was developed by Digital Science to describe both parent-child relationships between institutions and external related organisations. GRID disambiguates affiliation names for approximately 100,000 organisations and we therefore used the GRID record linked to the generally accepted name for each of the 18 university names.

Table 2.2 The universities used along with their abbreviations

Abbreviated name	Full institutional name	Country
Ain Shams	Ain Shams University	Egypt
Alexandria	Alexandria University	Egypt
Assiut	Assiut University	Egypt
AUB	American University of Beirut	Lebanon
Babylon	University of Babylon	Iraq
Baghdad	University of Baghdad	Iraq
Bahrain	University of Bahrain	Bahrain
Carthage	University of Carthage	Tunisia
Jordan	University of Jordan	Jordan
Khalifa	Khalifa University	United Arab Emirates
King Abdulaziz	King Abdulaziz University	Saudi Arabia
King Saud	King Saud University	Saudi Arabia
Kuwait	Kuwait University	Kuwait
Qatar	Qatar University	Qatar
Lebanese	Lebanese University	Lebanon
Sfax	University of Sfax	Tunisia
Sultan Qaboos	Sultan Qaboos University	Oman
UAEU	United Arab Emirates University	United Arab Emirates

In the first part of the study, we determined the proportion of records with a DOI in each database. We extracted the unique DOIs present in each database between the years 2014 and 2018 inclusive and calculated the proportion of total records in the database with a DOI. The overall share is reported as well as a share for each of the studied universities.

The second section of the study quantifies the share of the surplus caused by affiliation discrepancy and separately, the share of the surplus caused by database coverage. For this we used only those records with DOIs in the same five-year period and for the 18 selected universities. Any records whose publication year differed between databases meaning they were included in the five-year window in one database but not another, were counted under coverage discrepancy. Preparatory work for the study demonstrated the number of records in this category was negligible. Because of the complexities of database comparison, we compared coverage between the databases in pairs. That is to say we extracted DOIs using university affiliation names and the publication time window in one of the databases ('Primary'). We then searched for those same DOIs in the second database ('Comparator'), restricting the search to the same affiliation and the same publication time window. Those DOIs found in both databases constitute overlapping coverage and were not studied further. Those DOIs not found in the comparator database are termed the 'surplus' and the reasons for their absence are investigated. To establish these reasons, we systematically repeated the comparison removing elements of the search in the comparator database for affiliation and publication time period.

In the final part of the study, we manually analysed a sample of 24 affiliation discrepancies for each of the 12 database pairs. Examples were selected at random with a maximum three publications in each comparison from the same publisher. These examples served to illustrate the presence and types of affiliation discrepancies between databases. For each database pair surplus, we selected the university with the highest proportion of affiliation discrepancies. To have as much diversity as possible in the universities studied, each university was selected at most once. (If the university with the highest proportion of affiliation discrepancies had already been selected for another database pair, we moved on to the university with the next highest proportion.) Examination was performed by manually searching for the records on the web interface versions of each database and checking the published PDF documents as the ground truth.

2.4 Results

2.4.1 Number and proportion of DOIs

The total number of records along with the number of unique DOIs retrieved for each of the four databases is shown in Table 2.3.

Scopus and Web of Science each had DOIs for more than three quarters of the publications. Both these databases incorporate a selective expert assessment of journals, books, and conference series before they are admitted to the database. This process ensures that most indexed articles go through the established scholarly publishing route and are therefore increasingly likely to have a DOI. The proportion of records with DOIs for the selected universities in this study (89.9% Scopus and 82.3% Web of Science records) seem somewhat higher than the proportions reported by Gorraiz et al. (2016). However, that study used earlier data up to 2014 and demonstrated an upward trajectory for documents with a DOI that would roughly coincide with the figures presented here. Records with a DOI comprise 96% of the Dimensions database, which is the highest share among the four. As Dimensions uses Crossref as the key pillar of its content along with PubMed, the high proportion is not especially surprising. All Crossref records have a DOI and 90% of PubMed articles already had DOIs by 2015 (Boudry & Chartron, 2017).

There is some variance in DOI prevalence between the universities. The variance might be due to the over representation of certain document types that are less frequently assigned DOIs. e.g., books, book chapters, and conference proceedings. Equally, universities that publish more papers in subject fields such as the arts and humanities could also contribute to a lower proportion of DOIs among their publications (Gorraiz et al., 2016).

For studies concerning detailed comparison of database coverage, additional metadata such as article title and author names should be used to maximise identification of overlapping records. The focus of the current study however was not coverage, but the prevalence of affiliation discrepancies between databases. Therefore, we could exclude records that do not have DOIs without negatively impacting the study results. It was more important to identify records with a common identifier so we could compare their affiliations in the different databases. This approach was largely inspired by the work of (Huang et al., 2020) who also took this approach.

Table 2.3 Number and proportion of records with DOI by university

	Scopus			Web of Science			Dimensions			Microsoft Academic		
Institution	DOI	Total	% DOI	DOI	Total	% DOI	DOI	Total	% DOI	DOI	Total	% DOI
Ain Shams	7,785	8,567	90.9%	6,621	8,139	81.3%	9,741	9,809	99.3%	10,834	11,230	96.5%
Alexandria	6,875	7,524	91.4%	5,277	6,584	80.1%	7,218	7,280	99.1%	7,198	7,528	95.6%
Assiut	4,608	5,027	91.7%	3,930	4,762	82.5%	4,439	4,511	98.4%	4,538	4,829	94.0%
AUB	4,364	4,733	92.2%	4,420	5,569	79.4%	4,080	4,115	99.1%	5,620	5,961	94.3%
Babylon	1,269	1,978	64.2%	389	472	82.4%	618	621	99.5%	701	878	79.8%
Baghdad	2,687	3,827	70.2%	1,380	1,629	84.7%	1,930	1,959	98.5%	2,214	3,003	73.7%
Bahrain	811	952	85.2%	506	566	89.4%	722	723	99.9%	722	778	92.8%
Carthage	4,595	4,957	92.7%	3,985	5,263	75.7%	4,103	4,113	99.8%	3,257	3,433	94.9%
Jordan	3,725	4,334	85.9%	2,599	3,132	83.0%	3,277	3,306	99.1%	3,758	4,441	84.6%
Khalifa	5,781	6,252	92.5%	3,936	5,176	76.0%	4,991	4,991	100.0%	3,746	4,031	92.9%

King Abdulaziz	22,038	24,045	91.7%	19,287	21,502	89.7%	19,080	19,221	99.3%	18,966	19,889	95.4%
King Saud	21,620	23,870	90.6%	18,246	21,457	85.0%	19,470	19,788	98.4%	19,165	20,305	94.4%
Kuwait	3,204	3,617	88.6%	2,406	2,976	80.8%	3,002	3,021	99.4%	3,040	3,280	92.7%
Lebanese	2,880	3,107	92.7%	1,590	2,198	72.3%	2,933	2,967	98.9%	2,884	3,148	91.6%
Qatar	6,080	6,686	90.9%	4,182	5,393	77.5%	5,888	5,895	99.9%	5,980	6,373	93.8%
Sfax	7,339	7,863	93.3%	6,296	8,155	77.2%	6,793	6,803	99.9%	6,055	6,340	95.5%
Sultan Qaboos	3,727	4,229	88.1%	2,650	3,241	81.8%	3,485	3,539	98.5%	4,471	4,898	91.3%
UAEU	3,699	4,090	90.4%	2,540	3,172	80.1%	3,393	3,419	99.2%	3,327	3,626	91.8%
Total 18 universities	107,578	119,720	89.9%	85,069	103,363	82.3%	100,511	101,390	99.1%	102,145	109,503	93.3%
Whole database	13,046,237	15,495,969	84.2%	9,833,766	12,975,857	75.8%	21,691,146	22,589,839	96.0%	21,042,708	55,168,811	38.1%

2.4.2 Prevalence of affiliation discrepancies

In the next part of the study, we examined the prevalence of affiliation discrepancies in four citation indexes, Scopus, Web of Science, Dimensions, and Microsoft Academic for 18 selected Arab universities.

In Figure 2.1 we present a stacked bar for each database pair with the primary database on the left and the comparator database on the right. The first portion of the bar reading from the left represents the surplus records only found in the primary database and not present in the comparator. For instance, the top bar shows 30,060 Scopus DOIs that are not included in Web of Science. The second portion includes those DOIs that are present in both databases but are not assigned to the relevant university affiliation in the comparator database. In the first bar that means 1,831 Scopus DOIs are not found to have the same affiliation in Web of Science, even though they are present in Web of Science. The central portion of the bar shows the number of DOIs found in both databases with the same affiliation. The penultimate portion of the bar shows that 5,924 Web of Science DOIs are also found in Scopus but not with the relevant affiliation. In the final portion of the bar, there are 3,114 Web of Science DOIs that are not included in the Scopus database.

Those DOIs in the second and fourth portions of the bar therefore represent DOIs where there is a discrepancy between the affiliation assigned in the two databases. It is not possible at this point to say which, if either is wrong. To make a decision about the accuracy of the assigned affiliation, we need to look at the individual cases and usually check with the published PDF document. We report this in Section 2.4.4 of the study. Therefore, we do not refer to affiliation errors but prefer the term discrepancies. There will also be cases in which the affiliation is missing or has failed to be assigned to the university in both databases and those records would not appear at all in the results. We are therefore not presenting a comprehensive list of errors, but rather an indication of the proportion of discrepancies between databases pairs.

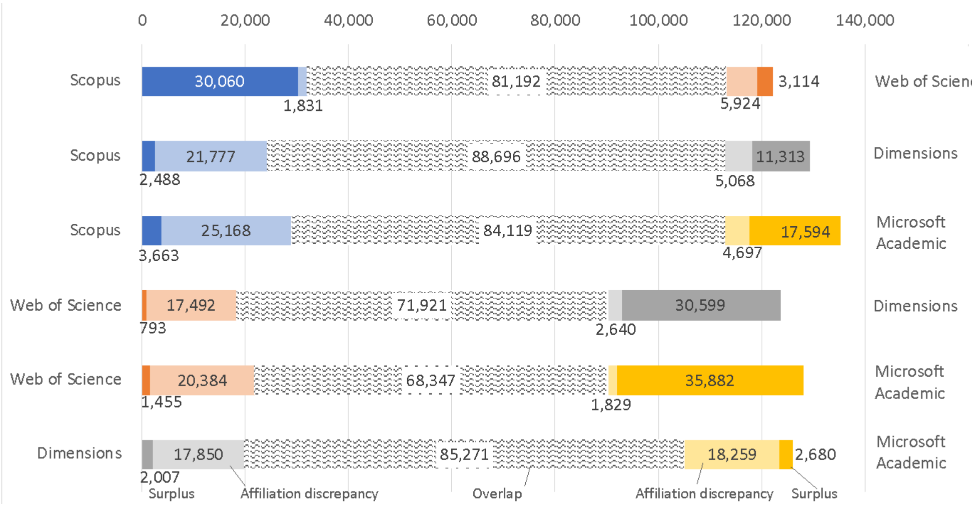


Figure 2.1 Affiliation discrepancy for the 18 universities by database pair

We can now analyse the relative affiliation discrepancies between each of the database pairs presented in Figure 2.1. For ease of discussion, we have organised the database pairs into four sections with the primary database as the heading. In each comparison, we only used those records associated with a DOI.

Scopus

The 1,831 Scopus records from the 18 selected universities that are present in Web of Science but not linked to the same affiliation represent about 2% of the actual overlap between the two databases. The next two bars show far more sizeable proportions of Scopus records not linked to the same affiliation in Dimensions and Microsoft Academic. We can interpret these results by suggesting the Scopus affiliation is more likely to agree with that assigned in Web of Science than it is with those assigned in either Dimensions or Microsoft Academic.

On the other hand, a small share (4% – 7%) of publications has been assigned to the 18 universities in Scopus, but not in Web of Science (5,924), Dimensions (5,068), and Microsoft Academic (4,697).

Web of Science

Records linked to the selected universities in Web of Science but not assigned to the same affiliations in each of the other three databases appeared to constitute about 7% of the overlapping Scopus records, and about a fifth of the overlapping DOIs in Dimensions (17,492) and Microsoft Academic (20,384). This means the Web of Science affiliation does not concur with the other three databases in a substantial proportion of cases.

Comparing in the other direction, the penultimate section of the Web of Science bars (1,831 records for Scopus, 2,640 for Dimensions, and 1,829 for Microsoft Academic) is relatively small in each case. That means there are comparatively few records in the three comparator databases in which the affiliations fail to concur with their corresponding record in Web of Science.

Dimensions

Dimensions records were clearly more likely to coincide with the affiliations assigned in Web of Science and Scopus than they were for affiliations assigned by Microsoft Academic. Dimensions affiliation discrepancies with those assigned in Web of Science (2,640 records) and Scopus (5,068 records) accounted for only 2% and 4% of the overlapping coverage respectively. Meanwhile, 15% Dimensions records with overlapping coverage in Microsoft Academic showed affiliation discrepancies with their corresponding records in Microsoft Academic.

The opposite comparison shows that a substantial number of publications have not been assigned to the 18 universities in Dimensions, while they have been assigned to these universities in the other three databases. This suggests that Dimensions may incorrectly not have assigned these publications to the 18 universities.

Microsoft Academic

Similarly, there was a relatively small number of Microsoft Academic records that did not agree with affiliations assigned in Web of Science (1,829) and Scopus (4,697). These discrepancies represented 2% and 4% of the overlapping records respectively. A more sizeable proportion of around 15% Microsoft Academic records failed to match the affiliations assigned in Dimensions.

From the other direction, a substantial number of publications have not been assigned to the 18 universities in Microsoft Academic, while they have been assigned to these universities in the other three databases. This suggests that Microsoft Academic may incorrectly not have assigned these publications to the 18 universities.

Overall summary of the results:

When Web of Science or Scopus assign an affiliation to a publication, the same affiliation is usually assigned by the other databases. The same is not the case the other way round. The largest share of discrepancies occurred with affiliations assigned by Microsoft Academic when compared with the other three databases. Dimensions assigned affiliations also found sizeable shares of discrepancies when their records were compared in the other databases.

The database coverage played a role with many records from the larger databases (Dimensions and Microsoft Academic) not found indexed in the smaller, more selective databases (Scopus and especially Web of Science).

2.4.3 Database surplus by university

Figures 2.2 – 2.7 show the extent of differences in coverage and affiliation between databases for each of the 18 selected universities. The additional level of data can help us interpret the differences between the way the databases approach affiliation disambiguation.

There were some patterns that emerged from this analysis. For example, there was a consistently higher proportion of records assigned to the Lebanese University than for the other universities found in Scopus, Dimensions, and Microsoft Academic that were not assigned to this university in Web of

Science. We should recall that for this study, we defined Web of Science records as belonging to the universities only if they were retrieved using the Organisation enhanced unification tool. If records contained the correct address but was not unified to its affiliation in the Organisation enhanced tool, then they would not have been retrieved.

Similarly, records assigned to Khalifa University in Scopus, Web of Science, and Dimensions were frequently not found assigned to that university in Microsoft Academic. Khalifa University is the result of a merger between three institutions that took place in 2017. Some of these records will be examined in Section 2.4.4 to determine how different databases approached the resulting affiliations.

In the case of AUB many publications have been assigned to the university in Web of Science and Microsoft Academic, while they have not been assigned to the university in Dimensions. As we will show in the next section, this was likely due to differing treatment of records from the American University of Beirut Medical Centre.

Babylon had a higher proportion of Scopus records than other universities that were not assigned to that university in Dimensions and Microsoft Academic. It should be noted that Scopus uses one affiliation identifier (AFID) for the main institution and assigns other AFIDs for the sub-units of the university based on the organisation hierarchy. However, this process has been conducted more vigorously for some institutions than it has for others, and that difference might influence the level of affiliation discrepancy found. For example, Babylon has only one Scopus AFID and the lowest or nearly lowest share of affiliation discrepancies in all the database pairings where Scopus is the comparator. Conversely, Carthage has one Scopus AFID for the main institution and 31 AFIDs for the sub-units which add 78% more records to the university total when they are all included in the search. As discussed in Section 2.3, we used only the main AFID to identify the publications of a university in Scopus. Hence, in the case of Carthage, the large number of AFIDs for sub-units of the university probably explains why the university has the highest or second highest share of affiliation discrepancies in all comparisons where Scopus was the comparator.

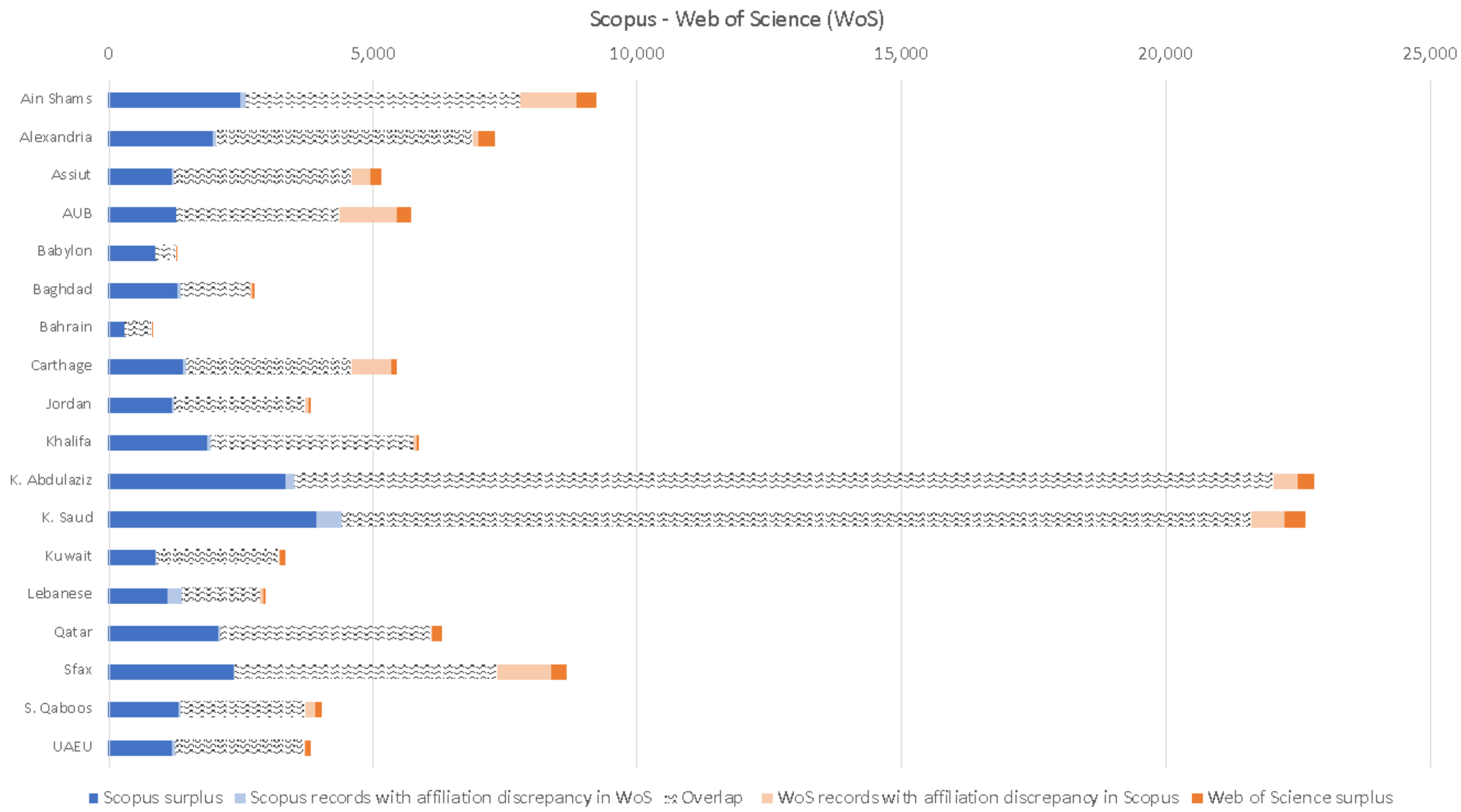


Figure 2.2 Scopus – Web of Science differences in coverage and affiliation by university

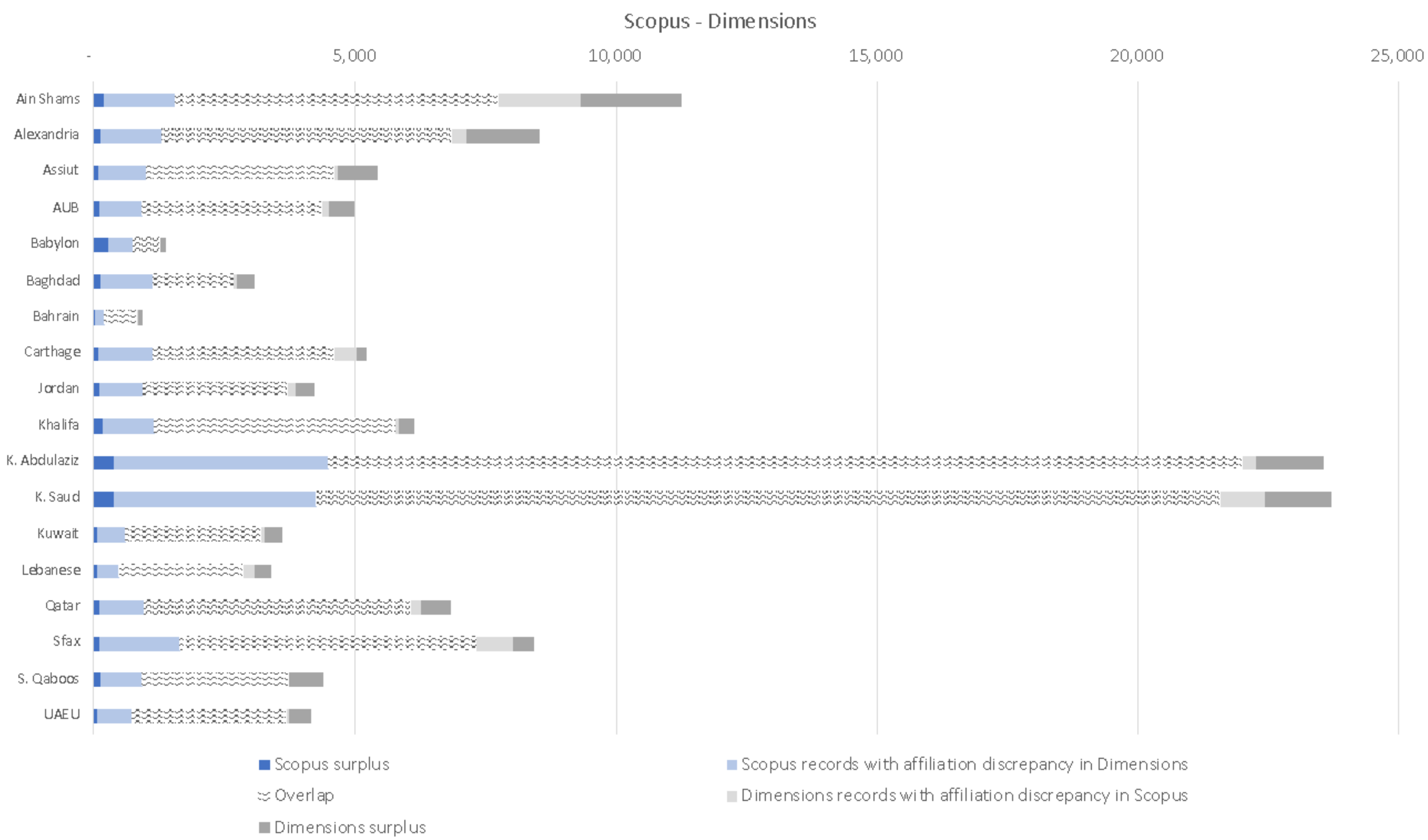


Figure 2.3 Scopus - Dimensions differences in coverage and affiliation by university

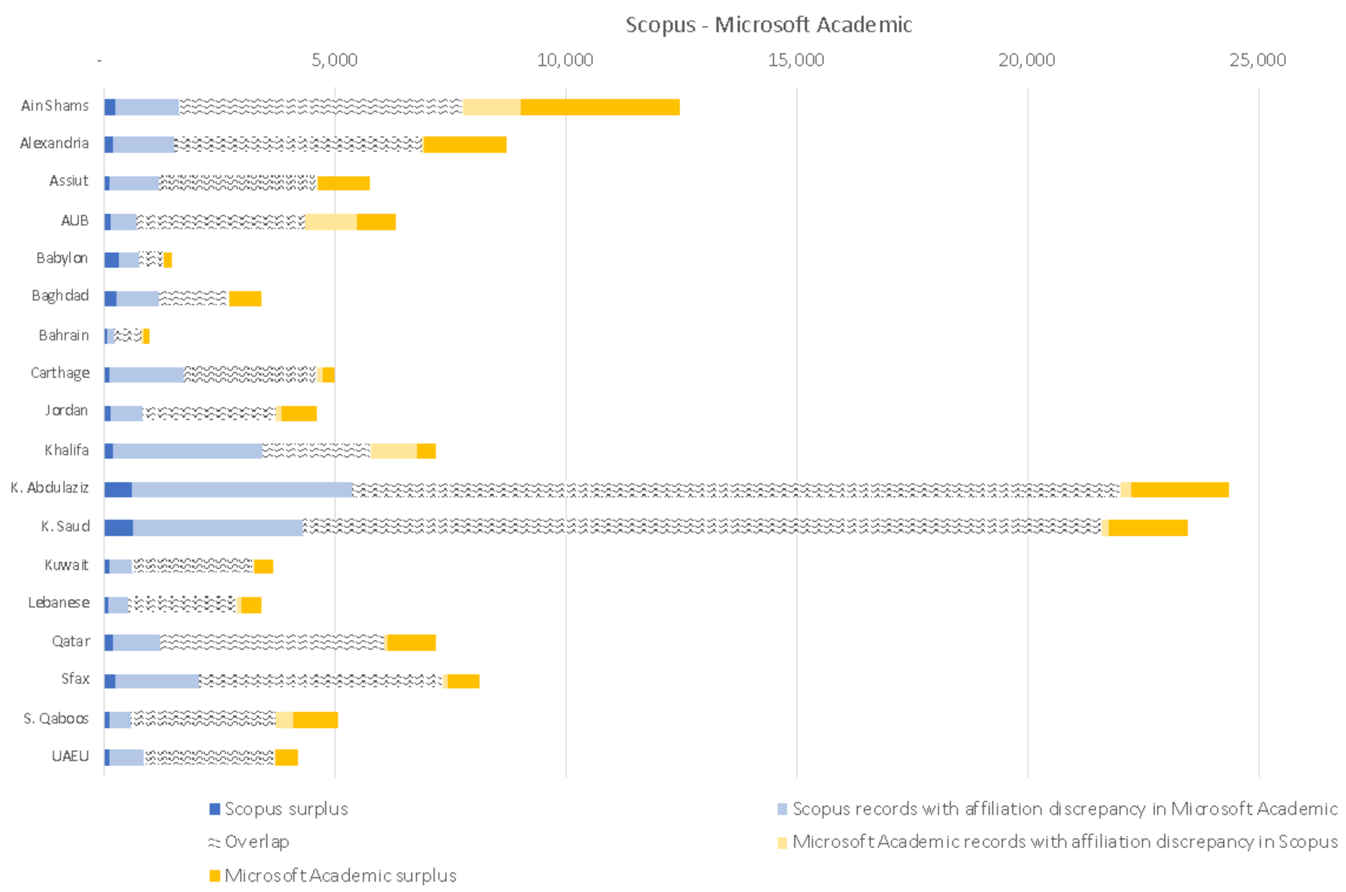


Figure 2.4 Scopus – Microsoft Academic differences in coverage and affiliation by university

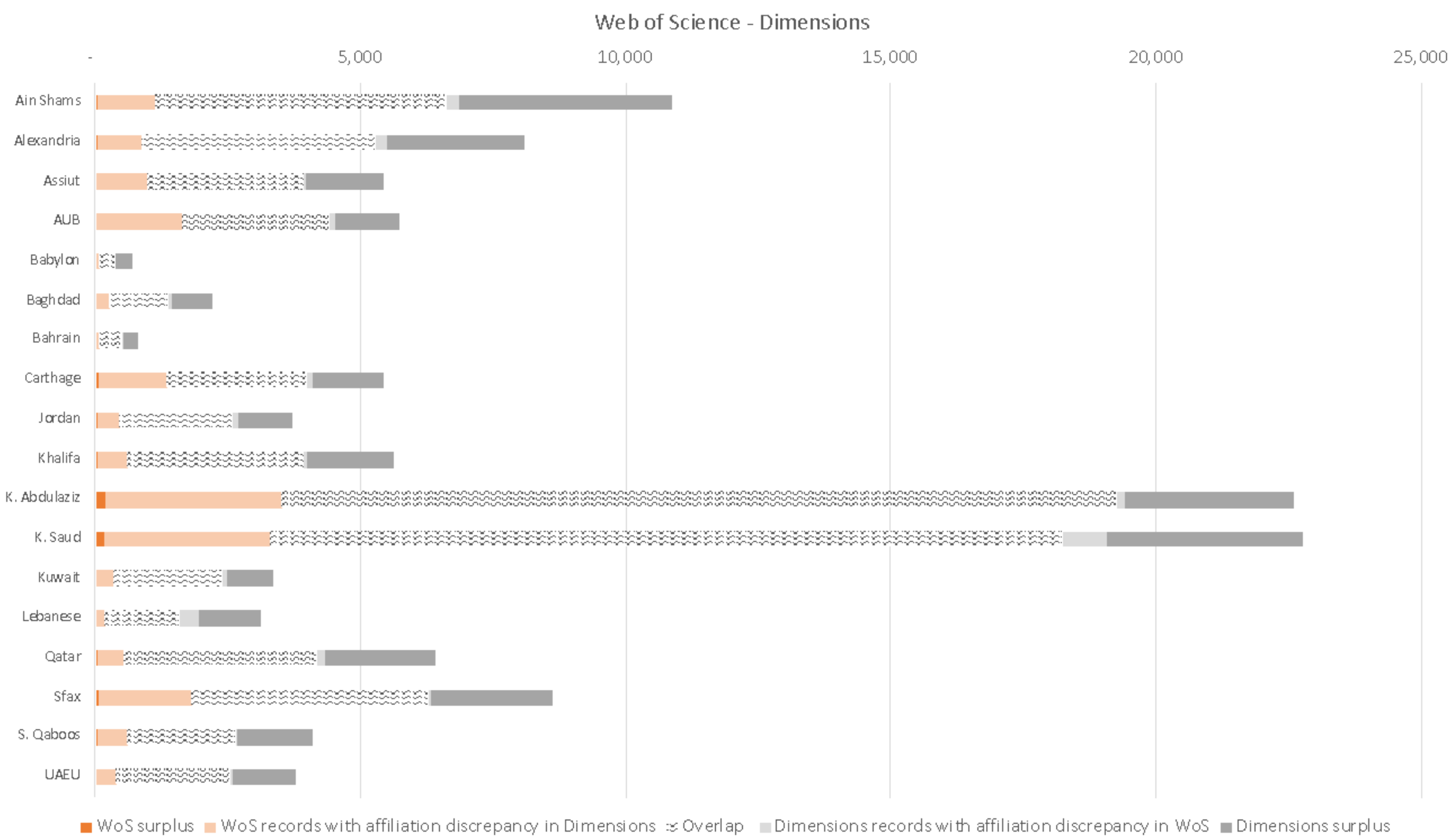


Figure 2.5 Web of Science – Dimensions differences in coverage and affiliation by university

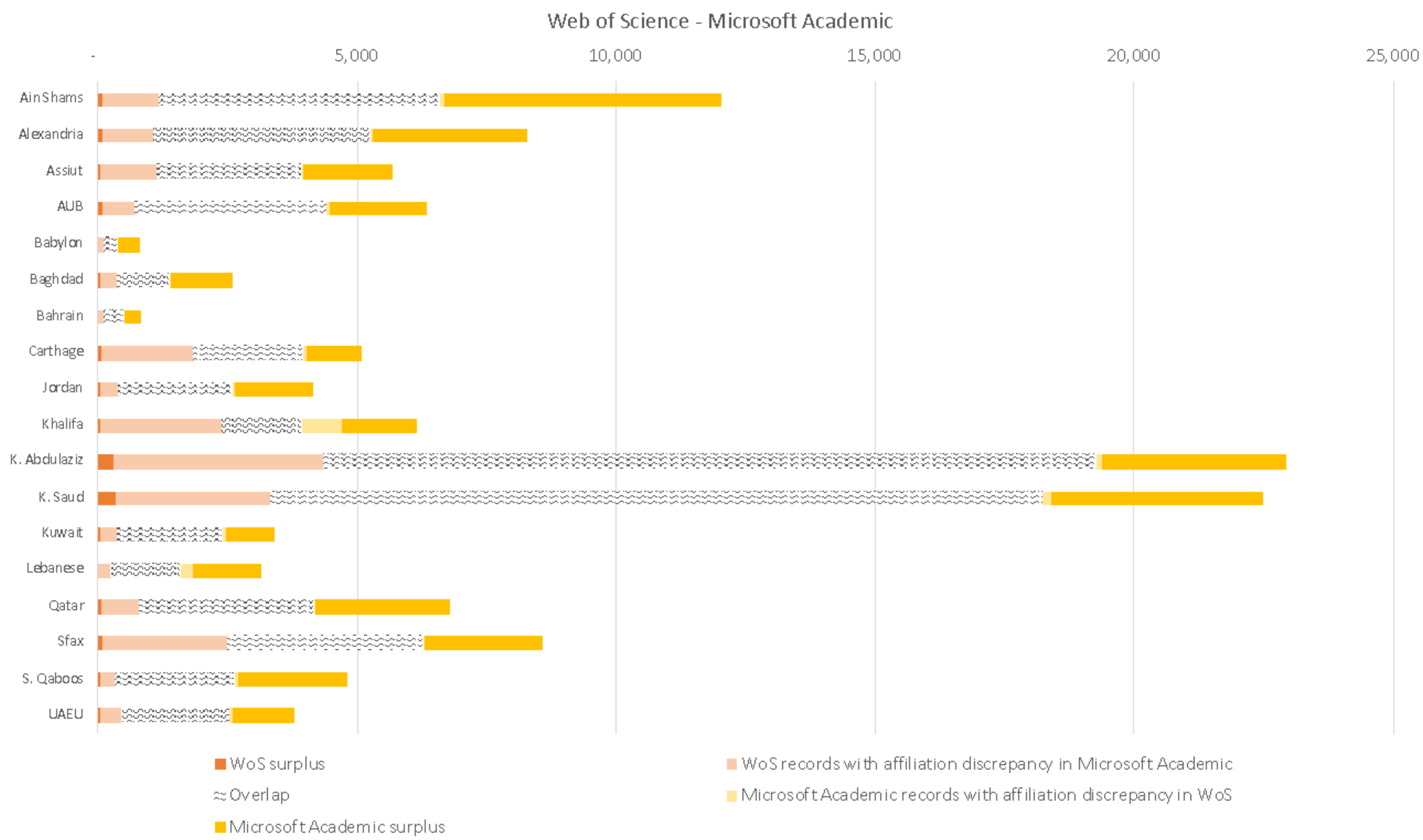


Figure 2.6 Web of Science – Microsoft Academic differences in coverage and affiliation by university

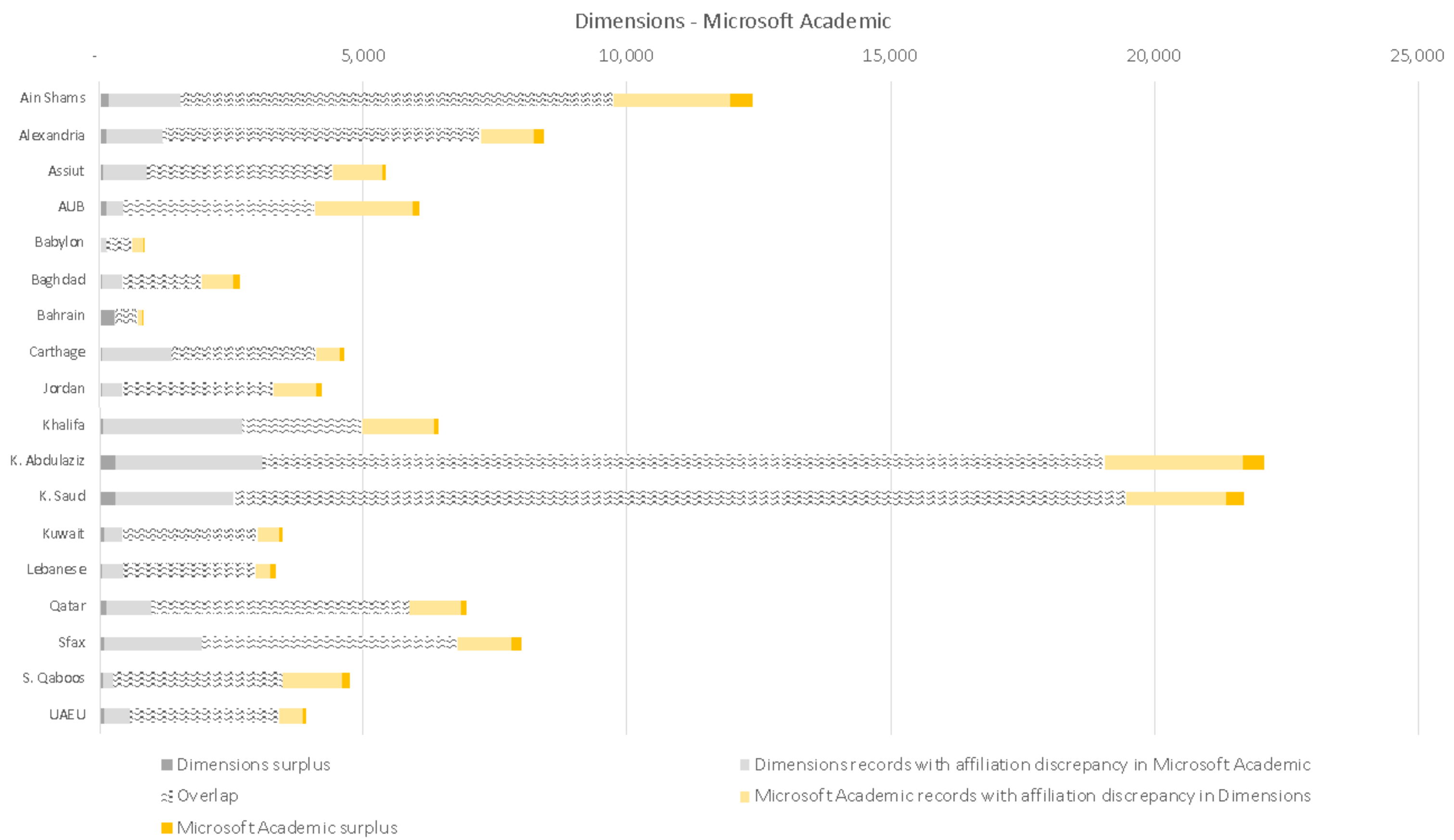


Figure 2.7 Dimensions – Microsoft Academic differences in coverage and affiliation by university

2.4.4 Types of affiliation discrepancy

The discrepancies in affiliations are the main focus of this study and vary widely in their prevalence and some interesting examples are described in this section. These highlight the challenges faced by database providers and the various ways they have responded to them. We found we could organise the affiliation discrepancies into four main groups as shown in Table 2.4.

Table 2.4 Types of affiliation discrepancy between databases

Affiliation discrepancy type	Definition	Example	Database
Missing affiliation	Author's affiliation is missing	10.4018/jdm.2016100102	Scopus
Missing second affiliation	Author's first affiliation is present, but their second affiliation is missing	10.1166/asl.2017.7424	Dimensions
Unification	Affiliation mentioned in some form but not linked to unified record	10.2174/1386207319666161214111822	Web of Science
Assigned to wrong institution	Author address linked to a different institution than that intended	10.1016/j.compfluid.2014.07.013	Microsoft Academic

We manually examined two dozen sample records at random from each of the database pairs using the web interface for each database. For each of these examples, we attempted to discover the reason for the affiliation discrepancy for one of the universities between the databases. The main reasons are summarised in Table 2.5. These sample analyses serve to illustrate that discrepancies exist and to shed light on the possible reasons behind them. However, a larger study would be needed to provide a more robust comparison.

Table 2.5 Affiliation errors

Database pair	Institution	Missing affiliation	Missing second affiliation	Assigned to wrong institution	Unification	Inconclusive
Scopus - Web of Science	Lebanese			1	22	1
Scopus-Dimensions	Sfax	11			11	2
Scopus-Microsoft Academic	Khalifa	2			20	2
Web of Science - Scopus	Saud	6	1	13		4
Web of Science - Dimensions	AUB	6	1	1	14	2
Web of Science - Microsoft Academic	Assiut	6	12	5	1	
Dimensions - Scopus	Carthage	3			21	
Dimensions - Web of Science	Bahrain	1		19	4	
Dimensions - Microsoft Academic	Babylon	9	6	7	2	
Microsoft Academic - Scopus	Kuwait					24
Microsoft Academic - Web of Science	UAEU	4	1	3	15	1
Microsoft Academic - Dimensions	Qatar	20	1	1	2	

Missing affiliation

The author affiliation has not been captured by the database and therefore a search for the affiliation name did not retrieve the record. We found almost all Qatar University records in the Microsoft Academic – Dimensions surplus were caused by missing affiliations in Dimensions. In most of these cases all author affiliations were missing, and the papers were conference proceedings or book chapters. Similarly, we found where Web of Science has missed author affiliations, the majority were meeting abstracts.

Missing second affiliation

The author's first affiliation has been listed but not the second. This is similar to the above category but worth separating as it appears that distinct groups of papers are indexed in some databases for which the additional affiliation is missed while the first is captured. As an example, on 10.1166/asl.2017.7424 the PDF shows University of Babylon as second affiliation for one author that is omitted from the record in Dimensions and Microsoft Academic, but included in Scopus and Web of Science. Sometimes we can speculate on the reason for this. In cases such as 10.1016/j.asoc.2016.06.019, the author's second affiliation is listed separately from the first on the PDF under categories such as 'Author's current address' or in this case, 'Correspondence address'. Scopus and Web of Science included this as second affiliation, while Dimensions and Microsoft Academic did not.

Assigned to the wrong institution

We found records from the College of Information Technology, University of Babylon and from the College of Information Technology, UAEU that had been erroneously assigned to the College of Information Technology, an institution registered in Pakistan in Microsoft Academic, e.g., 10.1007/s00500-018-3414-4. In other databases, these records had been correctly assigned to their respective universities. The term 'College of Information Technology' is a common university sub-unit, and it appears these words have triggered unification in Microsoft Academic to the stand-alone institution with the same name. Using these data in a bibliometric study would therefore produce a lower-than-expected count of papers from the affected institutions and an artificially high result for the real College of Information Technology in Pakistan. The same phenomenon occurs for Information Technology University in Pakistan, and the University College of Engineering in India, each of which are assigned additional papers in Microsoft Academic.

Other examples showed records assigned to UAEU in Web of Science organisation enhanced were in fact published by authors at a Moroccan institution called Université Abdelmalek Essaadi, locally abbreviated to 'UAE University' and mistakenly unified to the wrong institution. Similarly, authors from LaSTRe Laboratory in Tripoli, Northern Lebanon, which is affiliated to the Lebanese University, had been erroneously affiliated to the University of Benghazi in Tripoli, Libya in the Web of Science organisation enhanced. This might have occurred because of the appearance of the city name, Tripoli which is the capital of Libya but also a city in Lebanon. Two further records from the same university were assigned to the United States of America due to confusion over the town of Lebanon in Grafton County, New Hampshire. It appears therefore that the presence of a city name or country name might sometimes trigger unification to the wrong organisation enhanced name in the Web of Science.

Unification

An affiliation is listed but it has not been unified to the main or correct university. For example, most of the Scopus records that we did not find in Web of Science under the Lebanese University affiliation did actually mention the university in the address field but were not unified to that university in the organisation enhanced field. This is a plausible explanation for the large proportion of Lebanese university affiliation discrepancies with Web of Science described in section 2.4.3. Similarly, we discovered several examples of records attributed to either the Masdar Institute or Petroleum Institute in Abu Dhabi which have both been part of Khalifa University since a merger in 2017. Microsoft Academic has kept the original affiliation while the other databases have unified records to Khalifa University even from before the merger. This explains the high proportion of records unified to Khalifa in Scopus, Web of Science, and Dimensions but not found with that affiliation in Microsoft Academic. It also highlights the difficulties faced by database owners with treating records from organisations following their mergers or separations.

Many papers attributed to authors from the American University of Beirut Medical Centre have been unified to AUB in Web of Science, but not in Dimensions which treats it as a separate institution. That explains the notable proportion of affiliation discrepancies found with Dimensions in Figures 2.5 and 2.7. Similarly, many authors in Tunisia have acknowledged their institution as Faculty of Sciences at

Sfax. We found that Scopus unified these papers to the University of Sfax while Dimensions treated it as a separate organisation. In another similar case also in Tunisia, we found several records affiliated to the National Institute of the Applied Sciences and Technology and several other research centres which Web of Science and Dimensions unified to the University of Carthage while Scopus did not. In most of these cases, the university was not mentioned on the PDF document, but Scopus has made the link through its disambiguation process. The Scopus interface offers users the option of searching the 'whole institutions' that includes all affiliated institutes or 'affiliation only' which is the main identifier for the university.

Inconclusive

We classified cases as inconclusive where there was no obvious reason for the DOI not being retrieved, a human indexing decision involved, or access to PDF proved impossible. An example of a human indexing decision is a book preface with a DOI (10.1016/B978-0-12-800887-4.00034-1) but no authors or affiliations. While Scopus had assigned the book editors as authors, Dimensions had not. Another interesting case was a letter to the editor published by three authors with a long list of additional signatories at the end. Scopus had counted all the signatories as authors of the letter while Dimensions limited authorship to the three at the top of the paper. Neither of these cases is clear cut, if one accepts the book editors in the first case, and letter signatories in the second should be named as authors then their affiliations are missing in Dimensions. If they should not be named, then they are phantom affiliations in Scopus.

2.5 Discussion and conclusion

Our results showed the proportion of Scopus and Web of Science records with DOIs was in line with previous studies and that Dimensions showed near universal DOI coverage. Microsoft Academic included a large proportion of content including patents and other non-academic document types that do not have DOIs. There was some variance in DOI coverage between universities probably due to prevalence of certain document types or subject fields which vary in their DOI assignation. Overall, bibliometric studies like the one presented in this paper can use the presence of DOIs to limit their datasets to comparable scholarly material.

We analysed overlapping coverage between databases in pairs and organised the results based on whether or not the author affiliations matched. Some records were assigned to the same university in both databases in the paired comparison. A substantial share of records were assigned to a university in one database but not in the other. This study concentrated on the affiliation discrepancies between the databases for 18 selected Arab universities.

We found evidence that up to one in five publications can have discrepancies in author affiliations between the major bibliographic databases. We found the discrepancies more frequently in the larger databases, Dimensions and Microsoft Academic. The highest incidences of discrepancy were Web of Science records not found to have the same affiliation in Microsoft Academic and then Dimensions. The next highest were Scopus records not found to have the same affiliation in Microsoft Academic and then Dimensions. Meanwhile, when publications were assigned affiliations in any of the databases, the same affiliation was usually assigned in Scopus, and especially in Web of Science. These two databases are smaller, more selective, and crucially, frequently engage with institutions to improve the unification of affiliation variants. Our Web of Science results might be more favourable than the real situation because our version excludes book chapters and a lot of geographically diverse journals often from university presses that might less rigorously assign author affiliations.

Our results revealed different reasons behind the discrepancies. These included problems of unifying address variants to the main institution, publications with missing author affiliations, and cases of records clearly being assigned to the wrong institution. One common source of difficulty was the naming of institutions like the College of Information Technology which is also the name of the sub-unit of many universities around the world. Another arises from cities of the same name in different countries such as Tripoli that exists in both Libya and Lebanon.

In our sample, we found that discrepancies in Web of Science were most frequently due to problems with unifying variants and in some cases, confusion clearly led to assigning records to the wrong institution. Discrepancies among Dimensions records were more likely found due to missing affiliations but there were also some issues with unification. In Scopus and Microsoft Academic there was no clear pattern and causes of discrepancy were mixed.

The Scopus affiliation identifier (AFID) and the Web of Science organisation enhanced feature each depend on engagement with the institutions to link affiliation variants to the main organisation. Where Scopus has assigned multiple AFIDs to sub-units of universities, such as the 31 AFIDs assigned to sub-units of the University of Carthage there is more chance of discrepancy with other databases. When comparing institutions, bibliometricians should resist the assumption the disambiguation process has been performed to the same level for all institutions in the analysis. Our results show this is not always the case and that some comparisons will produce misleading results.

Manual examination of individual records revealed examples of publication records where databases have made different choices about how to unify author affiliations. The correct answer is often not clear and using one or another database will incorporate the impact of the choices of database owners into the results of any bibliometric analysis or benchmarking exercise based upon them.

University rankings providers usually rely on the disambiguation used in the source databases. The major exception is the Leiden Ranking which disambiguates all affiliations from its proprietary database (Calero-Medina et al., 2020), while both QS (QSIU, 2019) and Times Higher Education have begun supplementary work on Scopus unification in special cases. This practice is welcomed and providers of products that derive from bibliometric data sources should be encouraged to increasingly participate in the analytical process and assume a share of responsibility for the accuracy of the resulting publication.

These results support the conclusions of Huang et al. (2020) who encourages university ranking publishers to employ multiple bibliographic data sources. While Visser et al. (2021) discuss the merits of more selective databases for university rankings, the results in this paper show the presence of author affiliation disambiguation in Scopus and Web of Science still poses a significant limitation to accuracy.

A number of limiting factors should be considered when discussing the results presented here. This study only used publications assigned a DOI in the comparisons between databases. As shown in Table 2.3 that excludes up to a fifth of the records depending on the database used which might also contain affiliation discrepancies and influence the results. In addition, as discussed in Section 2.2.1, there are potential errors in assigning DOIs including duplicate DOIs for the same publications, multiple publications, or versions of one publication with the same DOI, and various types of error within the

DOI rendering it unable to link to its assigned publication. The DOI is the most appropriate identifier we found for use in this study, but we acknowledge its limitations.

The universities selected for use in the study were all from a specific geographical region and a broader comparison including institutions from different regions from around the world would show whether our findings generalise to other regions. Some databases have engaged with the institutions in the study to varying degrees as evidenced by the range in the number of affiliation variants listed among the Scopus AFIDs and Web of Science Organisation enhanced names. The engagement with universities will clearly improve the ability of the database to accurately identify and assign publications and therefore introduce a variable in our results.

The examination of types of affiliation discrepancy summarised in Table 2.5 relied on manual work and introduced an element of human judgement when comparing the published PDF document with its corresponding record in the bibliographic databases. This combined with the limited sample size for each comparison requires the reader to interpret the results as illustrations of the type of discrepancy and give an idea of the size of the problem. However, we do not interpret these data as statistically representative of the full, global databases.

Another considerable limitation to the present and most previous studies on affiliation disambiguation is the fact that university names change over time. Changes result from a number of factors including mergers and splits but also for reasons of government naming conventions, changes of country leaders, city names, and other factors. Once an organisational sub-unit is unified to an affiliation, all its prior papers will be found when searching the new unified affiliation. Studies are therefore a time-frozen shot of the current unification and do not take account of unification dynamics over time.

There is a need for a universal unique identifier for academic institutions that should reflect the current and historical organisation relationship tree. The ideal indicator will be supported by input from the institutions themselves in the same way that researchers maintain their own ORCID records. That way the accuracy, maintenance, and historical record will be maximised. There will be scope for non-maintenance or misuse especially where an institution can benefit from a certain interpretation of its organisation, but these will be outweighed by the benefits. In addition, in the case of an open infrastructure, any misuse will be publicly visible which will act as a disincentive. Universities and their stakeholders should still decide their own names and they are still the most appropriate managers of the public record of their relationships with sub-units and external entities.

This study has demonstrated the scope for improvement in four bibliographic databases and highlighted many problems faced by those attempting the task of disambiguation. Many of those cases can potentially be resolved by incorporating a global, universally accepted identifier for use by the worldwide research community and supported with input from universities.

Acknowledgements

Ton van Raan and Ludo Waltman for expert guidance throughout the study, Martijn Visser for helping improve the data accuracy, Christian Herzog (Digital Science) and Martin Szomszor (Clarivate) for their useful comments on earlier versions of the article, two anonymous reviewers whose suggestions helped improve the manuscript.

Data availability

The Scopus and Dimensions data used in this paper have been made freely available to CWTS for research purposes. The WoS data have been made available to CWTS under a paid license. For Microsoft Academic, we made use of an openly available data dump. We are not allowed to redistribute the Scopus, WoS, and Dimensions data used in this paper. The statistics presented in the figures in this paper are made available in the accompanying supplementary material.

Conflict of interest/competing interests

The author was previously affiliated with Thomson Reuters erstwhile owner of Web of Science, now owned by Clarivate.

Chapter 3

A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case study of SDG 13 – Climate Action[2]

[2] This chapter is based on:

Purnell, P. J. (2022). A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case Study of SDG 13: Climate Action. Quantitative Science Studies, 1–56. https://doi.org/10.1162/qss_a_00215

Abstract

As sustainability becomes an increasing priority throughout global society, academic and research institutions are assessed on their contribution to relevant research publications. This study compares four methods of identifying research publications related to United Nations Sustainable Development Goal 13 – climate action. The four methods, Elsevier, STRINGS, SIRIS, and Dimensions have each developed search strings with the help of subject matter experts which are then enhanced through distinct methods to produce a final set of publications. Our analysis showed that the methods produced comparable quantities of publications but with little overlap between them. We visualised some difference in topic focus between the methods and drew links with the search strategies used. Differences between publications retrieved are likely to come from subjective interpretation of the goals, keyword selection, operationalising search strategies, AI enhancements, and selection of bibliographic database. Each of the elements warrants deeper investigation to understand their role in identifying SDG-related research. Before choosing any method to assess the research contribution to SDGs, end users of SDG data should carefully consider their interpretation of the goal and determine which of the available methods produces the closest dataset. Meanwhile data providers might customise their methods for varying interpretations of the SDGs.

Keywords

Sustainable development goal – Climate action – bibliometrics – artificial intelligence – machine learning

3.1 Introduction

3.1.1 UN Sustainable Development Goals

The United Nations described a set of sustainable development goals (SDGs) within its 2030 sustainable development agenda. These goals were launched on 1st January 2016 and will be in place until 2030. The agenda includes 17 SDGs that are associated with 169 targets and progress is to be measured using 232 indicators (United Nations, 2017). The goals urge political, scientific, economic, and societal change to address global challenges and ensure sustainable development of the planet and all its inhabitants. To achieve these goals, all sectors of society are expected to participate including higher education institutions and research centres.

Table 3.1 United Nations SDG 13 goals and targets

SDG	13
Short name	Climate action
Long name	Take urgent action to combat climate change and its impacts
Targets	13.1 Strengthen resilience and adaptive capacity to climate-related hazards and natural disasters in all countries 13.2 Integrate climate change measures into national policies, strategies and planning 13.3 Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning 13.A Implement the commitment undertaken by developed-country parties to the United Nations Framework Convention on Climate Change to a goal of mobilizing jointly $100 billion annually by 2020 from all sources to address the needs of developing countries in the context of meaningful mitigation actions and transparency on implementation and fully operationalize the Green Climate Fund through its capitalization as soon as possible 13.B Promote mechanisms for raising capacity for effective climate change-related planning and management in least developed countries and small island developing States, including focusing on women, youth and local and marginalized communities

Source: United Nations resolution A/RES/71/313

One key step in assessing progress of the academic community against the SDGs is to identify the relevant research outputs. These are usually articles published in scholarly journals and books, or presentations at conferences. Research publications are indexed in large databases which can be searched using strings of keywords. If the search terms match words in the article title or abstract, then that article is included in the search results. In this paper, we compare different methods of identifying research publications related to SDGs. Our focus is on SDG 13: Climate action, whose goals and targets are shown in Table 3.1. We chose SDG 13 because it affects the entire global population and environment and is strongly dependent on scholarly research.

As research into the SDGs develops, the number of efforts to create search strings grows, each different from the others. Current methods of which we are aware are summarised in Table 3.2.

Table 3.2 Current methods of defining SDG-related research

Group	Data source	Method
Elsevier 2020	Scopus	Boolean search strings
Elsevier 2021	Scopus	ML-enhanced
Bergen	Web of Science	Boolean search strings
Aurora (Elsevier 2019)	Scopus	Boolean search strings
Clarivate ISI	Web of Science	Citation enhanced
STRINGS	Web of Science	Citation clustering-enhanced
SIRIS Academic	Various	ML-enhanced
Digital Science	Dimensions	ML-enhanced

3.1.2 Study aims

This study aims to quantify the different datasets produced by following four methods and shed light on the underlying causes of those differences. Each of the methods selected for this study has made some attempt to enhance their datasets through algorithms. These were done in different ways and sometimes at different stages of the process. Although the machine learning element was not scrutinised in this paper, it is worth highlighting the differences because they are likely to influence the resulting datasets. The following methods were chosen for the study because they cover all SDGs, and we have access to the search terms used and resulting publications:

- Elsevier (2021) – used to calculate part of the 2021 Impact Rankings (Times Higher Education, 2021)
- STRINGS – Steering Research and Innovation for Global Goals (Confraria et al., 2021)
- SIRIS Academic – a European consulting firm (SIRIS Academic, 2020)
- Dimensions – developed by Digital Science (Wastl et al., 2020)

It is important to point out that we did not attempt to evaluate the accuracy of the methods or to pick a winner. We deliberately chose not to develop our own method because of the multitude of questions raised when defining a ground truth (Gläser et al., 2017). Our intention was to shed light on the discrepancies produced when applying different perspectives to the same question.

We did not use other published methods because (e.g.):

- Bergen: We could not run the complex search in our version of Scopus and there were too many records to export from the Scopus database
- Aurora: The method was not fully developed for global analysis
- Clarivate ISI: The publications are not assigned to individual SDGs

Specifically, we analyse the search strings (inputs) used by each method, and sub-classify them into general terms, policy-related terms, and technical terms. The use of subject matter experts will surely influence the type of search terms used and consequently determine the set of research publications identified as related to SDG 13.

We then compare the size of the resulting sets of publications identified by each of the four methods. We perform quantitative comparison of the overlap and surplus of each of the methods. We then discuss the influence of the type of keywords used in the search strings in determining the final dataset.

Finally, we compare the articles identified by the different methods (outputs) using VOSviewer maps. These maps help visualise the nuances of each of the methods and show the links with the corresponding search strategies.

Research questions:

1. To what extent do different search strategies produce different sets of SDG13-related publications?
2. What is the impact of including different types of search terms in the search strings?
3. What is the impact of using larger, more inclusive data sources over smaller, more selective ones?

3.1.3 Assessing university impact

This study is important because universities are increasingly asked to demonstrate their 'impact' on society in areas such as sustainability, so there is an increasing need to expand the definition of university performance to encompass the area of societal impact. University contribution towards the SDGs is therefore both welcomed and expected by their stakeholders and society in general. This expectation is accompanied by efforts to measure progress against the SDGs using performance indicators appropriate for universities.

Academic publications are frequently used in research evaluation and universities are routinely assessed on their article output for internal performance review and international benchmarking. Research articles related to the SDG goals and targets are therefore an appropriate unit upon which to base such assessments. Indeed, there is now a global ranking of universities based on their progress against the SDGs, about a quarter of which is based on their research publications related to the SDGs (Times Higher Education, 2021). The first two editions of this ranking assessed universities on their Scopus-indexed publications retrieved via a series of search strings. The 2021 edition further extended the publication datasets through a process of machine learning.

Research publications are typically analysed using large international multidisciplinary databases comprised of scholarly research papers. Bibliographic databases such as Scopus, Web of Science, and Dimensions include journal articles, conference proceedings, and research published in books. However, they each have their own selection and coverage policy that results in differences between the publications included. Therefore, the choice of bibliographic database will determine the resulting dataset depending on the selection and coverage policy of the database. Running the same search in different databases will yield different results.

3.1.4 Search strategies

Even using the same data source does not make SDG-related publication datasets comparable, indeed it has been shown (Armitage et al., 2020a) that differences in search strategies make a big difference in outcomes. For SDG 13: Climate action, only about one-third of articles were found by two different approaches.

In bibliographic searching, the search strategy is of key importance in determining the final set of publications. The search for SDG-related research is in its infancy and we aim to advance current understanding of the relationship between different search strategies and the resulting publication datasets. The UN described the goals, targets and indicators using specific terms, and the UN and other bodies have published related documents and reports also using subject-specific language. As these reports were written by people with close subject knowledge, they can be used as sources of search terms in a bibliographic database. Subject matter experts can then refine the searches to improve their recall and precision.

In terms of information retrieval, recall is the number of relevant publications retrieved as a share of all the relevant publications. To maximise recall, one would make the search strategy as broad as possible and not be concerned with the prospect of finding false positives among the results. Meanwhile, precision is the number of relevant publications retrieved as a share of all retrieved publications. To maximise precision, the search strategy should be as narrow as possible to exclude any irrelevant publications. The trick in identifying SDG-related research is for the search strings to be both effective at recall while remaining precise.

In order to assess recall, it is first necessary to define a precise, yet representative, set of reference publications. The way a method operationalises its interpretation of the SDGs, through a reference dataset used to measure recall, will largely influence the results and overlap with other methods.

3.2 Literature review

3.2.1 Current methods of defining SDG research

One of the earliest insights into sustainability science was developed by the publisher Elsevier in collaboration with SciDev.Net (Elsevier & SciDev.Net, 2015). This was the first in a series of reports that aimed to describe the research landscape in areas related to the SDGs. In order to identify research papers related to the SDGs, Elsevier worked with field experts to design sets of Boolean queries that were applied to the Scopus Advanced search (Jayabalasingham et al., 2019). The keywords were related to research themes linked to the six Essential Elements (Dignity, People, Prosperity, Planet, Justice, and Partnership) described by the United Nations (2014). The 17 SDGs are grouped around these six Essential Elements. The experts identified key phrases from the titles and abstracts of relevant reports and those keywords were then used to search for scholarly articles indexed in the Scopus database. The advantage of this method was its ability to retrieve papers that use specific terms related to various aspects of one or more of the SDGs, without having to explicitly use the term ‘sustainable development goal’.

A team of library and information specialists from the University of Bergen in Norway set out to discover the degree to which the design of the search strings affected the resulting set of publications (Armitage et al., 2020a). They developed a set of complex Boolean search strings (Armitage et al., 2020b) for selected SDGs and queried them in Web of Science Advanced Topic Search. They then translated the 2019 Elsevier Boolean queries (Jayabalasingham et al., 2019) into Web of Science search strings and compared the results with their own 'Bergen' set. Comparison showed only a quarter of records were returned by both the 2019 Elsevier and the Bergen search strings with large quantities of publications not retrieved when using one or the other set of terms. The authors concluded that even mildly modifying the search strings used for specific SDGs will significantly change the resulting set of academic papers found. The Bergen group also addressed the question about what constitutes SDG-related research by creating two sets of search terms for each SDG; those relevant to the topic of e.g., clean water and sanitation, known as the Bergen Topic Approach (BTA), and papers on efforts to actually combat the challenges described by the SDGs, known as the Bergen Action Approach (BAA). The topic approach retrieved larger datasets than the action related searches, although the degree of overlap of these two approaches with the Elsevier dataset varied by SDG.

The SDG targets can be vague, weak, or non-essential (International Science Council, 2015) which makes it unclear which words or phrases in a target should be used in a search. Even search strings using the same initial terms will produce different results depending on how they are refined. The Bergen group reported that their queries tended to use more combinations of terms requiring each to be included in a paper for it to be returned in the results. For example, the Bergen group required the term 'climate change' to be combined with other terms found in the SDG 13 targets such as 'adaptation' or 'mitigation', whereas the 2019 Elsevier search would return results due to the simple appearance of the term 'climate change'. On the other hand, the 2019 Elsevier strategy refined its final dataset by excluding any papers that contain the term 'drug' or 'geomorphology'. Such papers relate to medicine and changes in earth layers related to prehistoric climate changes rather than those related to modern day climate action (Jayabalasingham et al., 2019). Both these methods aim to refine the dataset but will obviously lead to differences in results.

Since the comparison with the Bergen method, the complexity of Elsevier's search strategy has increased considerably. The breadth of Boolean queries has expanded to capture a wider range of related concepts such as carbon capture/mitigation, CO_2 in combination with global warming, or environmental impacts. The sole appearance of the term 'climate change' is no longer sufficient to retrieve publications. Similarly, the exclusion criteria in the 2021 method have been refined to over 30 specific terms, replacing the two in the 2019 method. Elsevier has published a full description of their methods along with the search strategies (Rivest et al., 2021).

Science-Metrix, now part of Elsevier, has described how analysts who are familiar with the SDG targets have defined sets of seed keywords for each SDG (Provençal et al., 2021; Rivest et al., 2021). In this scenario, the preference for precision over recall is emphasised meaning that the dataset is expected to contain publications with high relevance to the SDG targets, even at the expense of missing some.

Several of the methods including Bergen, SIRIS and Dimensions aimed to capture phrases used in context rather than only in their exact form by using proximity searches, so that 'climate impact'~3 would also capture phrases such as 'climate change impact' and 'changing climate and its impact on health'. Again, these methods used different levels of proximity and on different search terms so the effect would of course compound differences in the publications retrieved.

Other groups have also developed Boolean search queries to describe bodies of SDG-related research. For example, Jetten et al. (2019) aimed to discover the extent to which Wageningen University's work to improve food security through innovative technologies influenced media and policy documents. Similarly, Körfgen et al. (2018) developed a detailed keyword catalogue which found nearly a fifth of Austrian universities' research output was related to the SDGs. An attempt to assess Spanish public universities' contribution to the SDGs (Blasco et al., 2021) used a composite indicator that included the Times Higher Education Impact rankings which is in turn partially based on Elsevier's 2021 keyword search string.

The Aurora Network of universities created an initial classification model to enable them to identify which research publications were related to each SDG and whether these influenced government policy (Vanderfeesten & Otten, 2017). They began with a strict version that was limited to keywords found in UN policy documents that described the goals, targets, and indicators. A number of subsequent versions of these search terms gradually added more keywords including synonyms, new terms from updated UN documents, keyword combinations, and terms retrieved though survey data (Vanderfeesten et al., 2020), Elsevier (Jayabalasingham et al., 2019; Rivest et al., 2021), and SIRIS Academic (Duran-Silva et al., 2019). The Aurora bibliometric tool queries the Scopus database and has been used by the Association of Dutch Universities (VSNU) to create a sustainability impact dashboard (Association of Dutch Universities, 2019).

Clarivate has used a technique known as bibliographic coupling to approach the problem (Nakamura et al., 2019). The Clarivate method identifies any paper in Web of Science that has used the term 'sustainable development goal' in the title, abstract, or keywords and defines them as 'core' papers. They then add to these any paper that has cited one or more core papers. The citing papers plus the core papers makes up the SDG-related dataset.

The Science Policy & Research Unit at Sussex University and the United Nations Development Programme are leading a collaboration of several research centres known as STRINGS – Steering Research and Innovation for Global Goals. This collaboration has taken a novel approach in order to discover whether SDG research priorities in certain countries match those in which the related socio-economic challenges are greatest. They selected seed terms from a broad range of policy, technical, and scientific reports along with web forums and official UN documents using a combination of algorithms and expert opinion (Confraria et al., 2021) in an attempt to capture terms used by a broad section of society. They also compared their search strings with those used by Bergen and SIRIS Academic to remove false negatives from their results (Rafols et al., 2021). The resulting combinations of terms associated with each SDG were then searched in Web of Science and used to identify clusters of SDG-related publications. If a certain proportion of publications were retrieved from one cluster, then the whole cluster was added to the dataset. If the threshold was not reached, the cluster was not added. These clusters group publications that are related by citation links and this offers a way to identify not only SDG-related publications that use specific search terms but also SDG-related publications that do not use these search terms but that have citation links to publications that do use the search terms.

SIRIS Academic, a European consulting firm has looked through a broader set of document types including R&D projects hosted on the Community Research and Development Information Service (CORDIS) (SIRIS Academic, 2020). This repository comprises primary results from European Union funded Framework Programme projects ranging from FP1 to Horizon 2020 (European Commission, 2020).

Digital Science has developed an approach that queries the Dimensions database. The results were analysed by country and the proportion of national output calculated for each of the SDGs. A similar proportion in each SDG was considered a well-rounded footprint, while diverse emphasis was considered a skewed profile. Digital Science has also attempted to establish the extent of international collaboration for each SDG and to map the SDGs onto established scientific fields. The Dimensions SDG data have been used in a study by the Nature Index of leading science cities (Nature, 2021).

3.2.2 Algorithmic enhancements

An emerging trend is to employ machine learning to enrich datasets of SDG-related publications. In this model, manually selected keywords are used to identify a set of seed papers from a bibliographic data source. An AI algorithm then learns from these seed papers to recognise other relevant publications.

In 2021, the Elsevier team enriched its 2020 dataset through machine learning adding approximately 10% to its dataset by improving recall. They used the title, keywords, key descriptor terms, journal subject area, and abstract from around 1 million publications related to SDGs to create a computer algorithm that elicited records relevant to each of the SDGs through a machine learning model (Rivest et al., 2021). Times Higher Education used these results in the calculation of the 2021 Impact Rankings (Times Higher Education, 2021).

SIRIS has created a controlled vocabulary for each SDG defining its 'semantic breadth' through a manual process of reading reports and identifying seed keywords (SIRIS Academic, 2020). As a second step, they have used deep learning to train a neural network model to find synonyms with the seed keywords and create an ontology. The ontology is then matched with terms logically linked with the seed keywords in the CORDIS repository. A final quality check comprised human revision of results generated by the automated method for relevance to the original definition of the SDGs.

Digital Science's machine learning approach (Wastl et al., 2020) involved generating 17 training sets and using natural language processing to create an SDG classification scheme searchable in Dimensions (Wastl et al., 2020).

3.2.3 Bibliographic data sources

Since the 1970s, Web of Science and its components have been routinely used for evaluating journal impact e.g., Garfield (1972), university benchmarking e.g., van Raan (1999), national research impact assessment (Adams, 1998), the contribution of individual researchers (Hirsch, 2005) and the development of advanced bibliometric indicators (e.g., Waltman et al., 2011).

Elsevier launched Scopus, its global abstract and citation database of research papers from scholarly books, scientific conferences, and academic journals in 2004. Scopus has gradually become a key data source used for bibliometric studies of research output (Archambault et al., 2009; Baas et al., 2020; Schotten et al., 2017). Recently, Digital Science's Dimensions has also become an interesting data source for bibliometric studies (Herzog et al., 2020; Hook et al., 2018; Thelwall, 2018).

Each of these databases is built in a different way and has its unique selection criteria, indexing process, and therefore content. Web of Science is traditionally the most selective (Clarivate, 2022) and aims to concentrate on the highest impact academic journals. Scopus has broader coverage than Web of Science (Huang et al., 2020; Schotten et al., 2017), and Dimensions is the broadest of the three (Harzing, 2019; Visser et al., 2021). Therefore, searching the same terms will produce a different result depending on which data source is searched.

3.3 Data sources and methods

3.3.1 Creating the datasets

We created four datasets of SDG 13-related research using the methods described by the distinct research groups as follows. In each case, we used all document types and limited records to the five-year time window 2015-19

- Elsevier – We used the Elsevier 2021 method which is the result of a two-step process. First, Scopus records were extracted using the search string defined as SDG 13 in the fourth update (Rivest et al., 2021). The resulting set of articles were then fed into an algorithm described by Rivest et al. (2021) that uses machine learning methods to enhance the original list.
- STRINGS – We used the search terms elicited through the methods described by (Confraria et al., 2021) to query the titles, abstracts, and keywords of publications in Web of Science. The resulting publications were searched in approximately 4,000 clusters based on an article-level citation clustering described by Waltman & van Eck (2012). If a minimum 15% of any cluster contained our SDG13 related publications, then all publications in that cluster were included in our dataset. If the 15% threshold was not reached, then none of the publications in the cluster were included.
- SIRIS – We used the search strategy that combines keywords as described in the visual essay by SIRIS Academic (2020) to query the titles, abstracts, and keywords of publications in Web of Science.
- Dimensions – We used the SDG methods including the machine learning enhancements described by Wastl et al. (2020).

We ran the queries described above against versions of Web of Science and Dimensions housed in the database system of the Centre for Science and Technology Studies (CWTS) at Leiden University.

Our version of Web of Science includes the following five indexes: Science Citation Index – Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, and both editions of the Conference Proceedings Citation Index). Neither the Book Citation Index nor the Emerging Sources Citation Index were used because we do not have access to them.

Elsevier's International Centre for the Study of Research (ICSR) Lab kindly made the 2021 dataset available for the purpose of this study, which we used in combination with the CWTS version of Scopus.

3.3.2 Search term classification

We collected the search terms used by each of the methods and organised them into three groups using our own general knowledge of the field. We classified the search terms for each method according to the following criteria:

- General: Terms used in society, e.g., “temperature rise”. The general public would use these terms.
- Policy: Terms that require knowledge of policy contents, e.g., “emissions trading”. They do not include mere mention of policy, e.g., “Kyoto protocol” which would count as a general term.
- Technical: Terms that are technical in nature, so subject matter experts would use them. They either refer to a technology, e.g. "thermal energy storage", or require technical knowledge, e.g. "radiative forcing". Standard technologies, e.g. "solar panel", do not count – these would be considered general terms.

3.3.3 DOI analysis

First, we calculated the total number of publications for the five-year period 2015 to 2019 from each of the four datasets related to SDG 13. We then determined which of these records had a DOI. We subsequently used the DOI as the unique identifier when comparing the datasets. This means only records with DOIs were included in the comparisons.

3.3.4 Pairwise coverage comparisons

We performed pair-wise comparisons to examine the overlap between the four datasets. So, each dataset was compared with the other three, thereby making a total of six pairwise comparisons.

Because the four methods use different data sources, some of the surplus is due to differences in coverage between those data sources. For example, STRINGS uses Web of Science while Elsevier uses Scopus. Therefore, we sub-divided each surplus into two portions. One portion of the surplus was due to differences between the search strategies described by each of the methods, while the other portion was due to coverage differences between the data sources. We termed these ‘surplus (method)’ and ‘surplus (coverage)’ respectively.

3.3.5 Visualisation of the outputs

For each pairwise comparison, we then presented the results in the form of VOSviewer term maps. These maps visualise terms found in the titles and abstracts of the articles in two datasets and show the terms in different shades of colour depending on the frequency with which they occur in each dataset. Each term needed to appear a minimum of 70 times in publications retrieved from a pairwise combined set of search terms. The term maps offer an easy way to see which topics are over- or underrepresented in one dataset compared to the other.

3.4 Results

3.4.1 Publications with a DOI

Table 3.3 shows the total number of records identified by each of the four methods, and those that are associated with a DOI.

Table 3.3 SDG 13 records and DOIs for each selected method (2015-19)

Method	SDG13 Publications	Publications with DOI	Share of publications with DOI
Elsevier 2021	214,369	195,734	91.3%
STRINGS	166,528	156,010	83.7%
SIRIS	177,154	164,800	93.0%
Dimensions	205,190	203,447	99.2%

The number of publications related to SDG 13 was relatively similar for each of the four methods chosen for the study. The largest set of records was found through the Elsevier 2021 method, although only about 5% larger than the Dimensions dataset. The SIRIS and especially STRINGS datasets were smaller, although the STRINGS dataset was still over three-quarters the size of the Elsevier 2021 total.

Each method had DOIs for at least 91% of its SDG 13 related publications. That meant we had a comparable set of publications associated with DOIs from the four methods for comparative study.

3.4.2 Comparison based on classification of search terms

The Elsevier 2021 method used an expansive list of keywords covering a range of topics related to climate change such as greenhouse gas emissions, and global warming. The list also extended into terms describing actions taken to address the problem such as policies and laws, but also addressed developments of resilient foods and agricultural methods. The term "legum* breed*" AND ("climate" or "drought" or "flood") is one of many technical terms related to food and agriculture in the context of climate action. These highly specific terms are designed to maximise recall while maintaining a high level of precision. There is also a considerable set of exclusion terms that use the AND NOT command, e.g., "Prehistoric Climate" and "blood". These are intended to exclude publications captured by the initial search terms, but that are not related to the current challenges surrounding climate action. The exclusions therefore improve precision of the dataset.

STRINGS used a lot of broad, simple terms, for example, “climate change”. STRINGS but not SIRIS used the term “carbon economy”, while SIRIS instead used more specific terms not employed by STRINGS, e.g., “carbon accounting”, “carbon audit”, “carbon credit”, “carbon dividend”, “carbon fee”, and “carbon finance”. STRINGS extracted search terms from lay documents including web forums and grey literature as well as policy documents and scientific publications to capture terms used by a broad section of society. The STRINGS surpluses due to method were all higher than those of other methods. We speculate that is the effect of the enhancement step that is based on the grouping of Web of Science into 4,000 clusters of publications related by citation links even in the absence of explicit use of keywords. The enhancement makes a decision about whether to include or exclude each of the Web of Science 4,000 topic clusters in the final dataset. If the cluster is selected, then all publications in that cluster are added. As the threshold for inclusion was set at 15%, it means that all publications in any cluster in which 15% of the records contain the seed keywords are included. However, the 15% inclusion threshold makes it possible that up to 85% of the records in a selected cluster did not in fact contain the keywords. This record 10.1080/09540091.2017.1279126 is exclusively retrieved by the STRINGS method, although it contains none of the STRINGS keywords. The explanation must be inclusion in a topic cluster selected because of the existence of other publications bearing the search terms. The citation-based grouping seems to have been more inclusive than the other methods in the study and emphasise recall over precision. Conversely, any topic cluster in which fewer than 15% records contain the seed keywords is excluded along with all its publications, even those that did contain the keywords. An example is 10.1371/journal.pone.0137275 which contains the term ‘climate change’ in its abstract. This term is included in the STRINGS seed keywords but the publication is not included in the final dataset. It must therefore have been excluded from STRINGS due to existence in a topic cluster mainly populated with less relevant papers. Therefore, even where the different methods used the same keywords, this enhancement step has produced discrepancies compared with the other methods.

Overall SIRIS used more than twice as many search terms as STRINGS, many of them technical. There were 54 ‘technical’ search terms compared with only four in STRINGS. For example, “ocean acidification” and “radiative forcing” found thousands of records in SIRIS that did not appear in STRINGS. Sometimes SIRIS was restrictive, for example requiring the term “climate change” to be combined with others, i.e., "climate change" and ("policies" OR "education" OR "impact" OR "reduction" OR "warning" OR "planning" OR "strategy" OR "mitigation"). Conversely, the simple mention of “greenhouse gas” qualified publications for inclusion in SIRIS, while STRINGS required the same term to be combined with another term like “emission”, “reduction”, or “changing climate”. The technical terms used by SIRIS contributed to large numbers of publications in the SIRIS surpluses against all the other methods.

Dimensions used only 45 search terms, most of them general. However, these were searched against a larger database. The Dimensions method also employed the proximity search in almost all the search terms so that phrases that included certain words in close proximity would be found. For instance, ‘Climate related hazards’~3 will also find articles that contain ‘hazards related to climate change’ in their titles or abstracts. The advantage is that publications that include phrases used in the context of climate action could be returned rather than only finding an exact phrase.

The number of search terms used by each method is shown by type in Table 3.4. The Elsevier 2021 method used mainly general and technical terms plus about 14% policy related terms. The STRINGS method used a high proportion of general terms, but the remainder were almost all policy related with very few technical terms. The SIRIS method was far more specific with about a quarter of the search terms policy related and a quarter technical in nature.

Table 3.4 Search term classification

Method	General	Policy	Technical	Total
Elsevier 2021	210 (46%)	62 (14%)	186 (41%)	458
STRINGS	70 (71%)	24 (24%)	4 (4%)	98
SIRIS	119 (52%)	55 (24%)	54 (24%)	228
Dimensions	34 (76%)	9 (20%)	2 (4%)	45

The full list of terms along with their classification is available in Zenodo (Purnell, 2022a).

3.4.3 Comparison based on overlap of publications

In each pairwise comparison the set of overlapping records is shown in the central portion of the Venn diagram. Only publications with a DOI are used in order to make these comparisons. The records found in one dataset but not the other can be termed surplus. In the sample diagram (Figure 3.1), the two portions to the left are each included in dataset A, but not in dataset B and therefore comprise the dataset A surplus. As the methods use different bibliographic databases (Table 3.2), the surplus can be subdivided into the portion of the surplus due to the differences in method, and the portion due to differences in coverage.

The reader may consider these comparisons as a Venn diagram flattened into a stacked horizontal bar as shown in Figure 3.1.

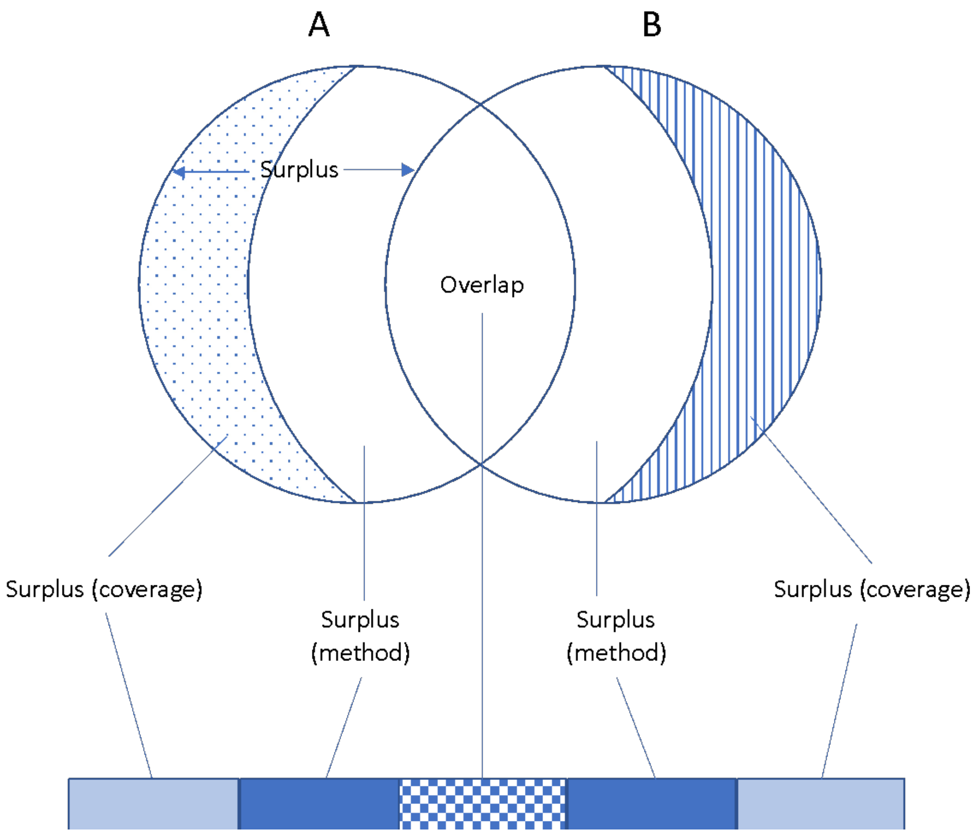


Figure 3.1 Key to overlap and surplus

Figure 3.2 shows pair-wise coverage comparisons of SDG 13 related publications between the four different methods. Each bar is labelled with the two datasets compared. The number of records in each portion of the pairwise comparison is shown in Table 3.5.

The first comparison shows that 56,043 publications were found by both Elsevier 2021 and STRINGS, and these are represented by the central portion of the bar. Immediately to the left of the overlap are 79,613 records found in the Elsevier 2021 dataset, but not STRINGS because of the difference between the two SDG 13 search strategies (surplus due to method). The far-left portion of the bar represents a further 36,544 records in the Elsevier 2021 dataset, but not in the STRINGS set because these records are not found in Web of Science (surplus due to coverage).

Likewise, the other end of the bar shows 2,949 STRINGS publications that were not found through the Elsevier 2021 method because they are not indexed in Scopus. The remaining 97,018 STRINGS publications were not found in the Elsevier 2021 dataset due to differences between the two search strategies.

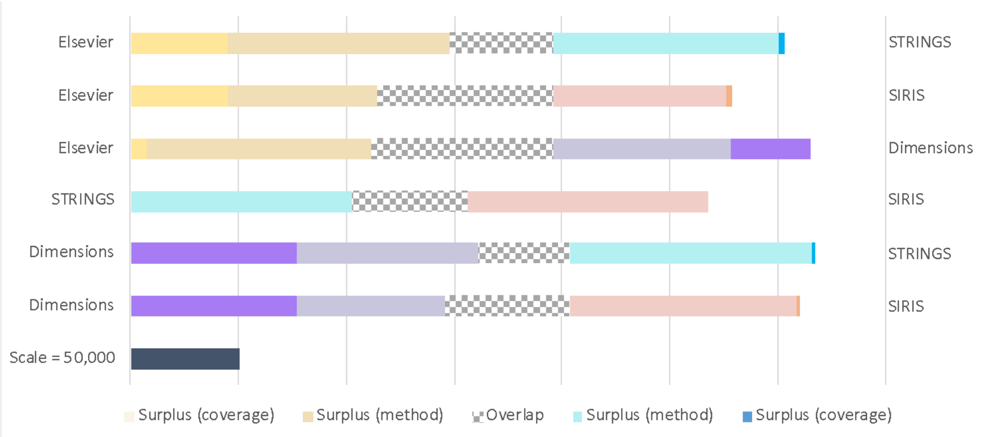


Figure 3.2 Number of overlapping and surplus publications between methods

The largest overlap (29.2%) was between the Elsevier 2021 and SIRIS methods (Table 3.5), while the lowest (13.4%) was between Dimensions and STRINGS. Overlap means that both methods in the comparison retrieved the same publications. The range of agreement is surprisingly low, indeed no two methods compared show a high degree of overlap.

Surplus due to method ranged from 22.2% to 41.6%. These are publications that were found by one method but not the other, where the discrepancy was attributed to the method of identifying the publications. The high level of surplus due to method demonstrates the large disagreements between all four methods.

The surplus due to database coverage was very low (maximum 1%) for both methods that used Web of Science (STRINGS and SIRIS), confirming the selective coverage of Web of Science. Conversely, Dimensions showed in one case (vs. SIRIS) that almost a quarter (24.6%) of the combined records in the pair were in its surplus due coverage. These are publications found by one method but not the other where the discrepancy is attributable to the coverage of the data source. There was no surplus due to coverage for the STRINGS vs SIRIS comparison because both used the Web of Science database.

Table 3.5 Number and share of overlapping and surplus publications

Method A	Surplus (coverage)	Surplus (method)	Overlap	Surplus (method)	Surplus (coverage)	Method B
Elsevier	44,764	102,702	48,269	104,792	2,949	STRINGS
	14.8%	33.8%	15.9%	34.5%	1.0%	
Elsevier	44,764	69,502	81,469	80,421	2,910	SIRIS
	16.0%	24.9%	29.2%	28.8%	1.0%	
Elsevier	7,103	104,613	84,019	82,587	36,831	Dimensions
	2.3%	33.2%	26.7%	26.2%	11.7%	
STRINGS	0	102,564	53,446	111,354	0	SIRIS
	0.0%	38.4%	20.0%	41.6%	0.0%	
Dimensions	76,389	84,629	42,429	112,522	1,059	STRINGS
	24.1%	26.7%	13.4%	35.5%	0.3%	
Dimensions	76,389	68,933	58,125	105,494	1,181	SIRIS
	24.6%	22.2%	18.7%	34.0%	0.4%	

Sample DOIs for each group are available via a link in Zenodo (Purnell, 2022a).

3.4.4 Comparison based on topical focus of publications

Analysis of the resulting publications visualised through VOSviewer showed terms extracted from the titles and abstracts of publications and grouped by co-occurrences in publications. Each comparison shows two maps. The first map allows us to assign broad descriptive phrases like ‘energy problem’ to clusters of papers with the most relevant and frequently occurring terms represented by different colours. In Figure 3.3, the first map groups terms into three distinct, colour coded fields related to energy problem (green), CO2 problem (blue), and climate change (red).

The second map (also figure 3.3) shows for each term which of the two methods of identifying SDG 13 related research captured more publications that use the term. Each bubble represents a term, and the colour of the bubble reflects the score of the term. Terms that occurred more frequently in the publications identified by the first method have a negative score and appear over a blue bubble, while terms that occurred more frequently in publications identified by the second method have a positive score and appear over red bubbles. Terms that appear over the faded colour bubbles occurred evenly in publications identified by both methods. In Figure 3.5, the second map shows us that the terms related to the CO2 problem tend to occur more frequently in the Elsevier dataset. Meanwhile, the terms related to the energy problem appeared more frequently in the Dimensions dataset.

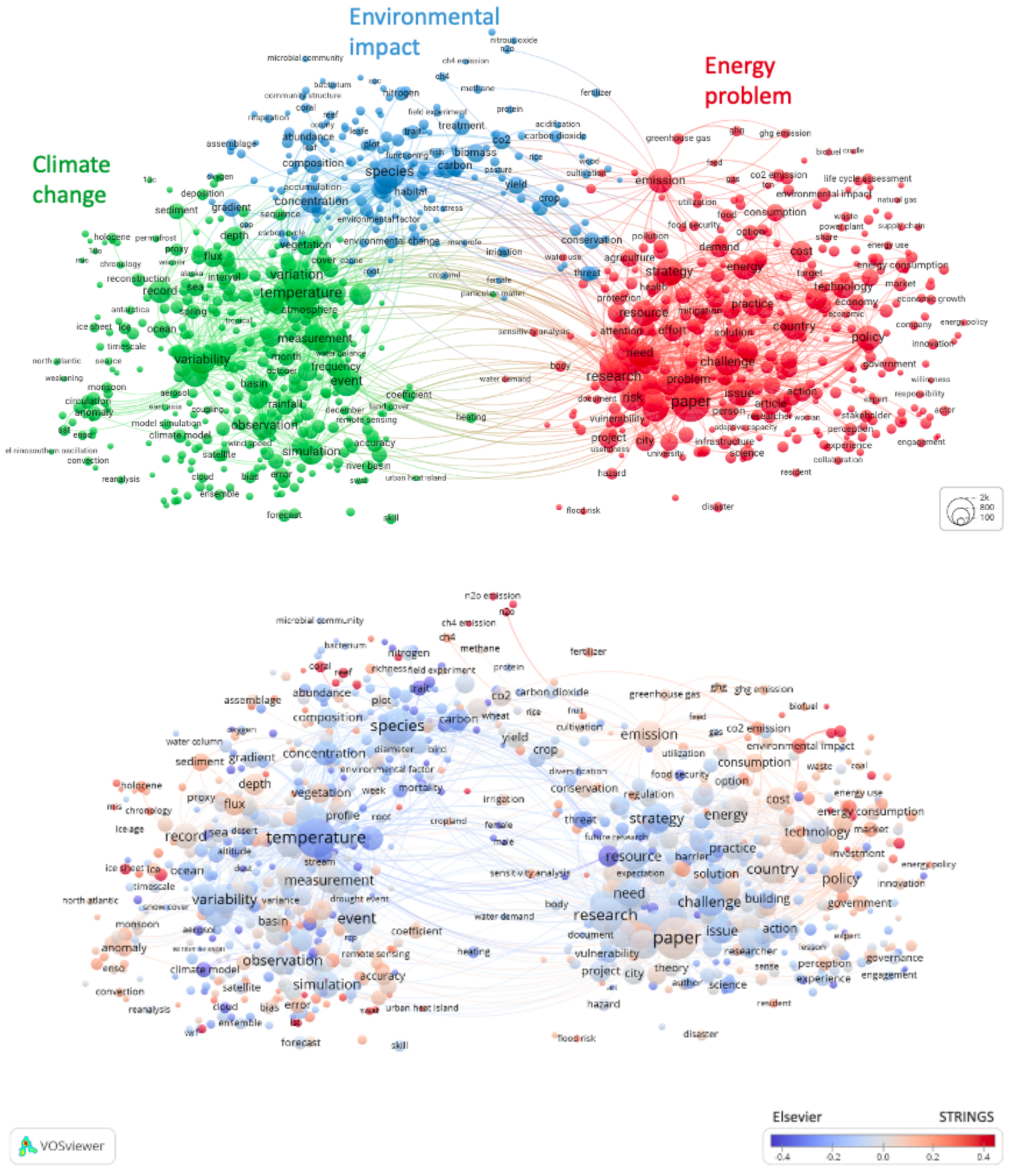


Figure 3.3 Elsevier vs STRINGS - Click to navigate interactive live map

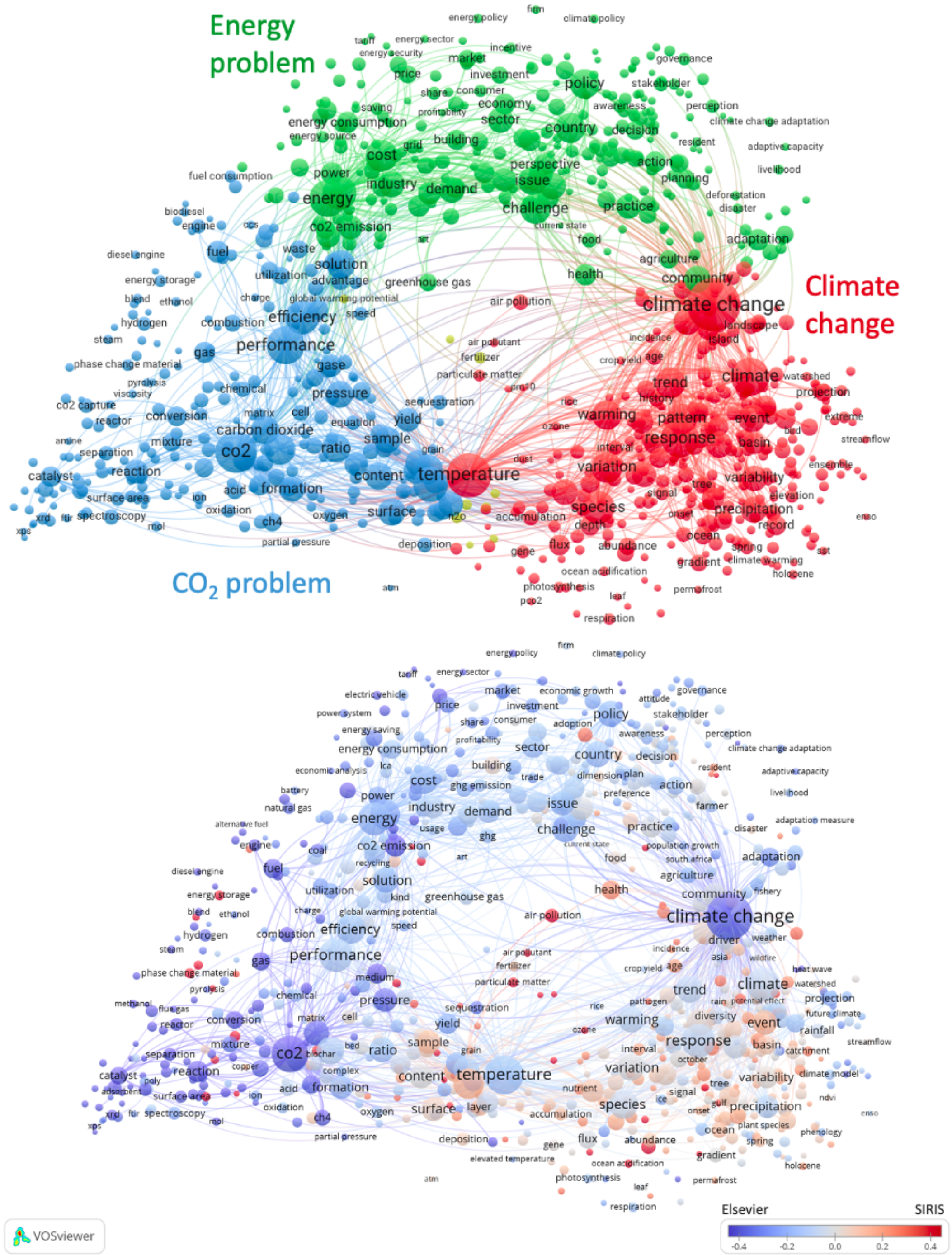


Figure 3.4 Elsevier vs SIRIS Click to navigate interactive live map

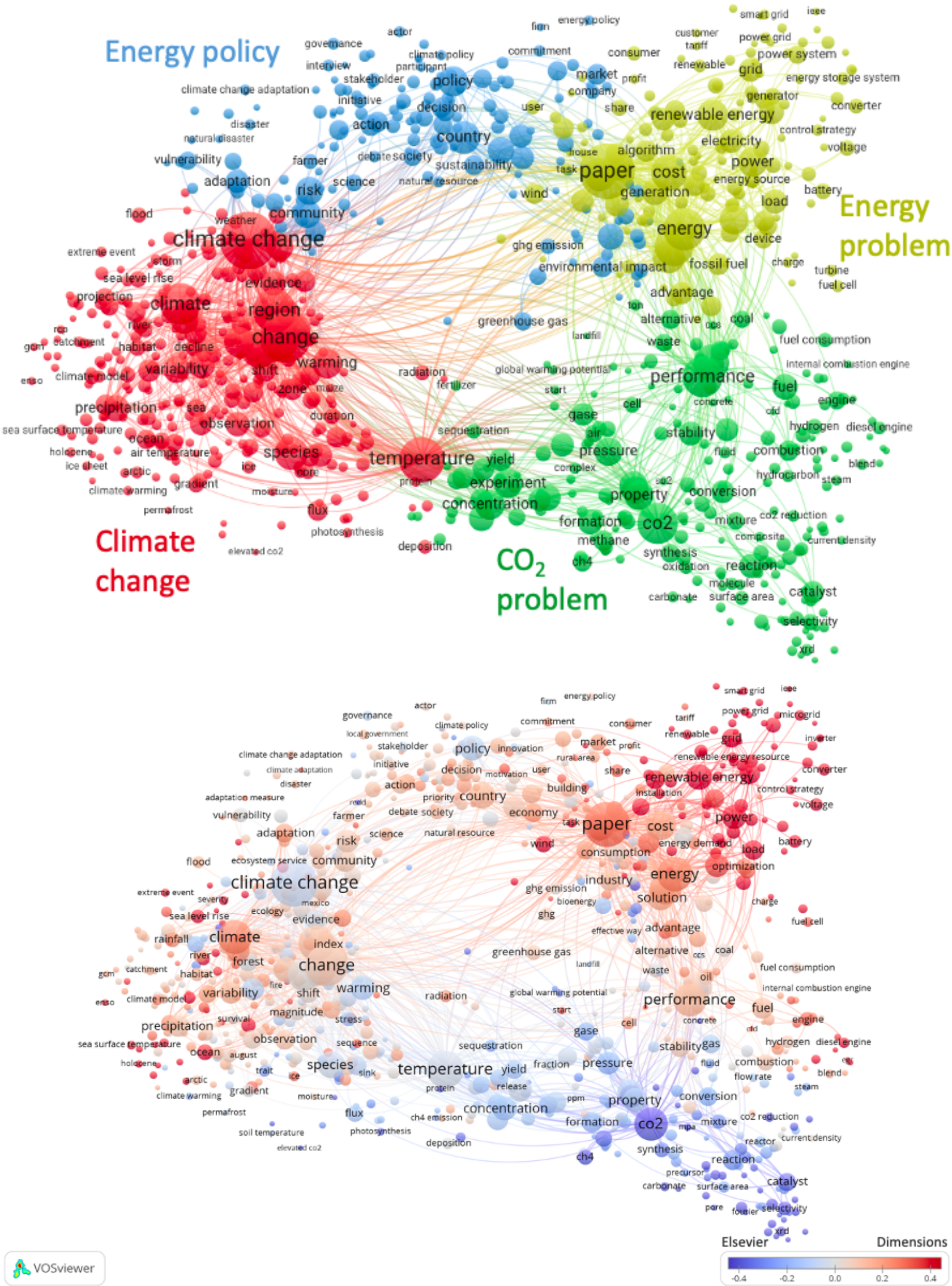


Figure 3.5 Elsevier vs Dimensions Click to navigate interactive live map

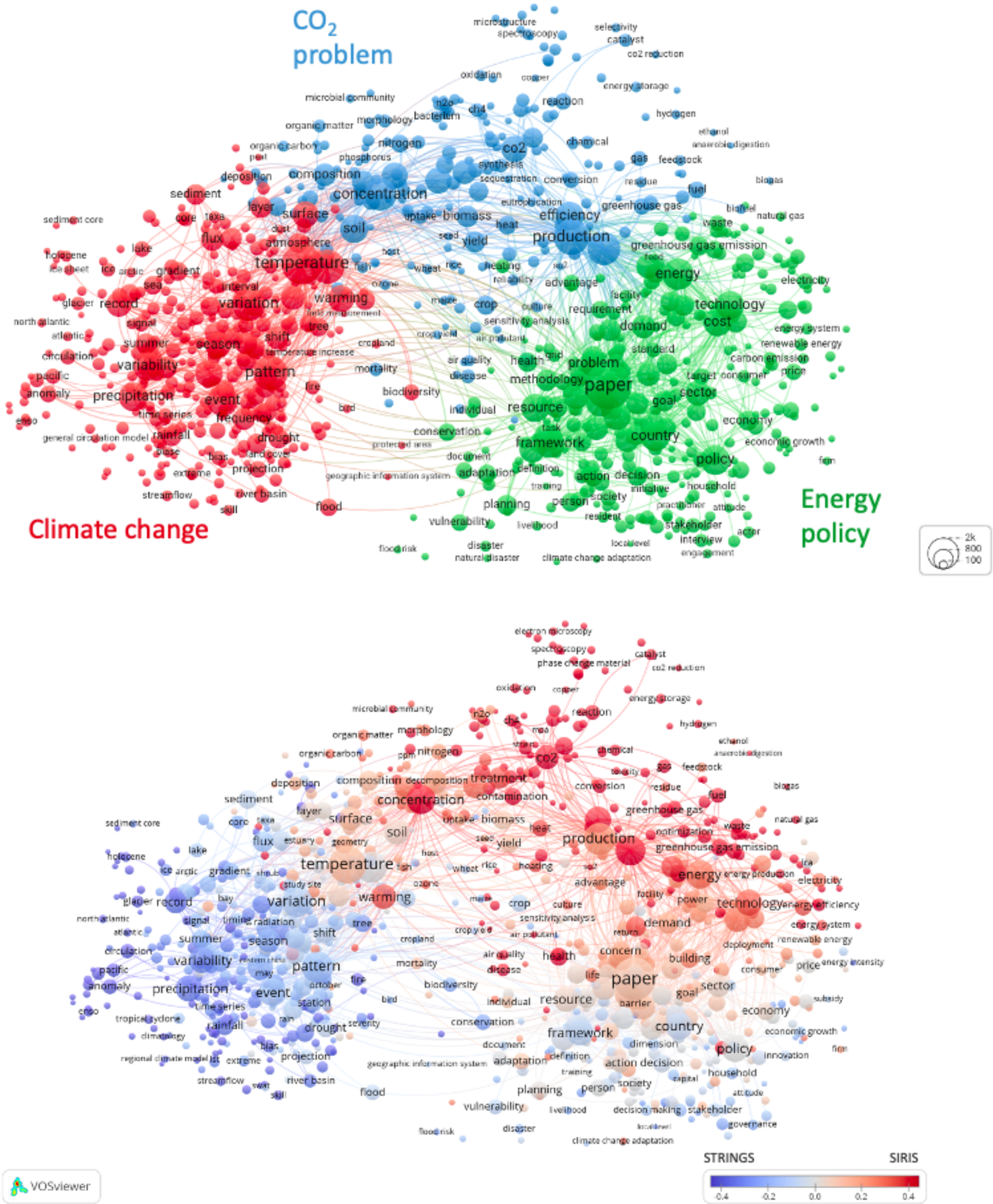


Figure 3.6 STRINGS vs SIRIS - Click to navigate interactive live map

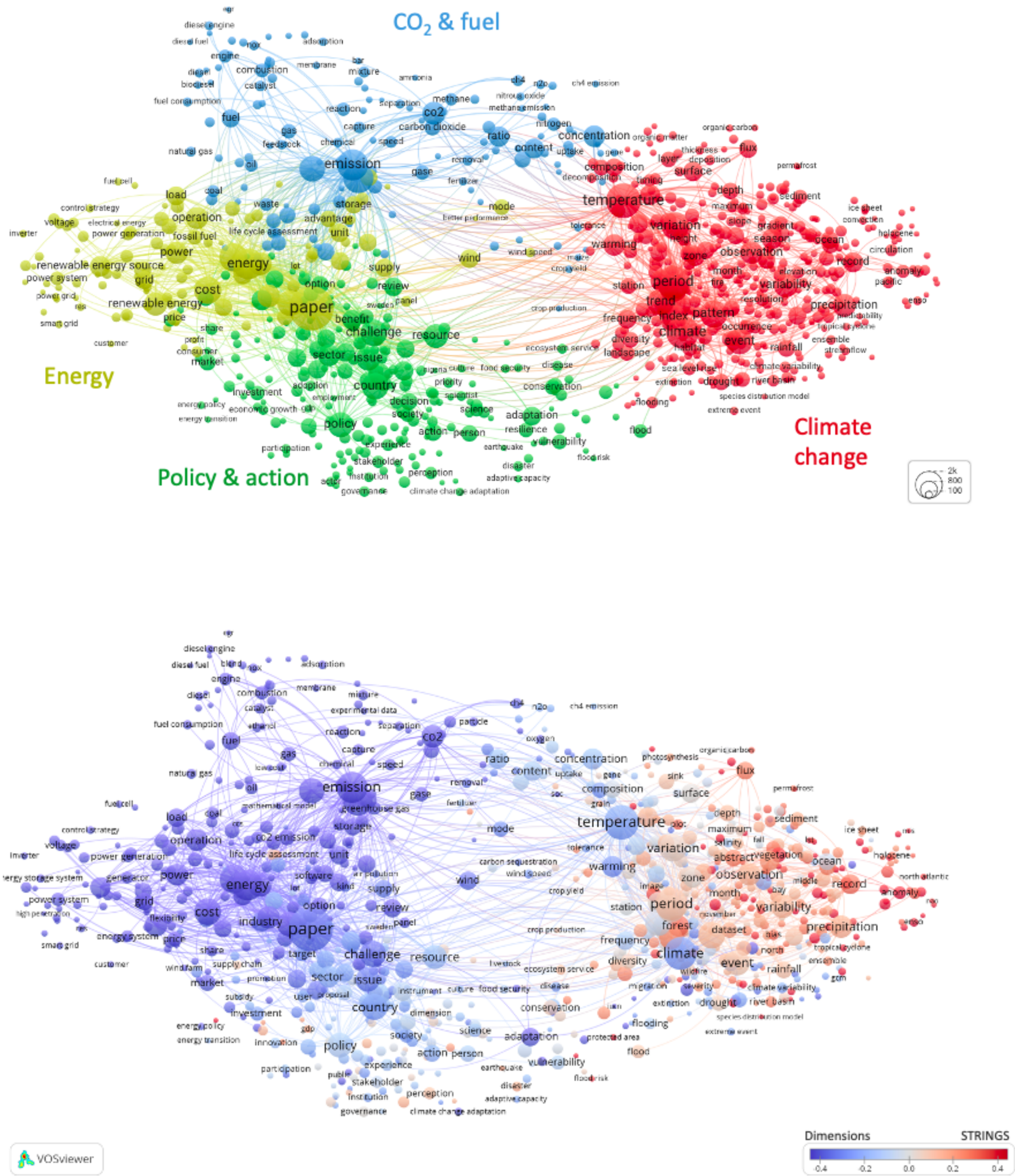


Figure 3.7 Dimensions vs STRINGS - Click to navigate interactive live map

.

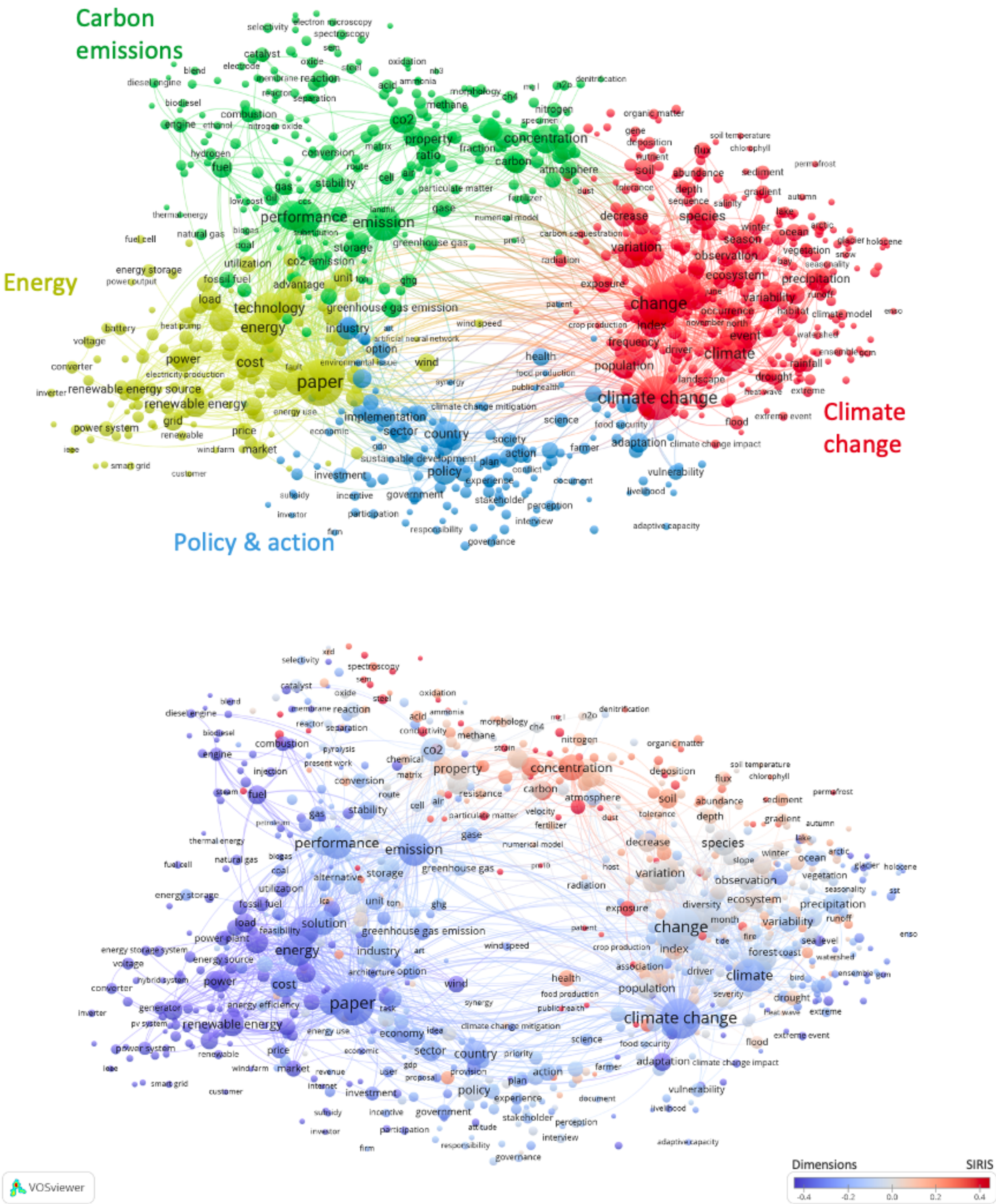


Figure 3.8 Dimensions vs SIRIS - Click to navigate interactive map

The four methods studied have all identified SDG 13-related publications but each with a discernible topical emphasis. Each map presents the terms most frequently found in a combined set of publications for the methods compared. The terms may appear in publications identified by both methods, but the colour indicates which method identified them more frequently.

The first three VOSviewer maps (figures 3.3 – 3.5) show terms expressed in research publications identified by the Elsevier method compared with the other three methods. In these maps, there is no clearly discernible pattern. In the comparison with STRINGS (figure 3.3), the Elsevier method has perhaps identified papers that more frequently use terms related to the environment and climate. This might be due to the inclusion of large numbers of related search terms in the Elsevier seed keywords. Alternatively, Elsevier's machine learning enhancement might have trained the search engine to identify these publications, or perhaps Scopus has indexed more papers in this field than the other databases.

In the three comparisons with STRINGS (figures 3.3, 3.6, and 3.7), patterns are easier to see. The colouring of terms related to climate change indicated they occurred more frequently in the publications identified by STRINGS than the other methods, although this distinction was less clear in the comparison with Elsevier (figure 3.3). STRINGS used broader, more encompassing search terms, e.g. 'climate change', than the other methods. This broad search strategy might have contributed to recall of a larger set of publications that contained related terms. STRINGS then introduced more publications to its dataset by adding all papers in clusters related by citation links. We did not quantify these additions, but entire clusters of publications were added if 15% or more records in the cluster contained the keywords. We assume this approach added many publications on climate change and contributed to their prominence in the maps. As STRINGS uses Web of Science as its data source, it is also possible that database indexes publications on climate change more frequently than Scopus or Dimensions. If that were the case, it would at least partially explain the prominence of the records in the STRINGS comparisons. However, SIRIS also used Web of Science and in the pairwise comparison, STRINGS clearly found climate change publications more frequently.

The SIRIS method appears to have retrieved publications more focused on the technical nature of carbon emissions. SIRIS used a relatively large number of keywords, and they were highly technical in their nature. SIRIS avoided broad terms like 'carbon emissions' but instead used 32 more specific terms containing the word 'carbon' such as 'orbiting carbon observatory' and 'personal carbon trading'. Construction of SIRIS search terms was supported by natural language processing and we speculate that has resulted in the more frequent inclusion of publications with technical terms as seen in the maps. Again, database coverage would provide an alternative explanation if Web of Science indexed more technical publications than the other databases. However, the comparison with STRINGS (figure 3.6), which also used Web of Science, showed SIRIS to find more technical publications, making database coverage a less likely explanation.

Dimensions demonstrated some prominence in publications with terms related to energy and policy. Dimensions searches the term 'renewable energy' which retrieves a large quantity of publications, whereas the other methods require that term to be used in combination. An explanation might partially lie in the interpretation of the subject matter experts of the term 'climate action'. Experts might differ in their emphasis with some focusing on the 'action' part while others may see the term more synonymous with climate change in general. If the experts used in the Dimensions method wanted to focus on the action, then it would make sense to choose terms containing verbs such as 'reduce emissions' and 'limit global temperature rise'. This point of view might also lead experts to select the names of agreements and forums among its search terms as places where action is discussed. Elsevier took a conscious decision not to include 'renewable energy' as a stand-alone search term for SDG 13 publications to reduce overlap with SDG 7 (Affordable and clean energy). Dimensions is the broadest of the three data sources used and its larger journal coverage might also have contributed to its large surpluses against the other methods.

3.5 Discussion

In this study we compared the publication sets retrieved by four different methods of identifying research related to SDG 13: Climate action. Each method begins by selecting relevant keywords from the SDG goal and its related targets. These keywords are then combined to create a query that is searched on a bibliographic database. Each method then enhanced its results in different ways. The resulting set of publications from the four methods overlapped very little, given that they all started with the same task.

Overlap is defined as publications that were retrieved by two methods directly compared with each other. Any publications found by one method but not the other are discrepancies. The fact that each method comprises multiple stages means that we cannot easily determine the source of any discrepancy. The method in effect becomes a black box. Our inputs are the keywords, and the resulting publication set the output. Discrepancies between the publication sets may be the result of any stage of the methods compared. Those designing methods of identifying SDG related research should be encouraged to open the black box by publishing each element of their method so end users can choose from a more informed perspective, see Table 3.6.

Table 3.6 Elements of the black box

Element	Description
Seed keyword selection	Description of source documents and how keywords were selected
Use of experts	Type of expertise, time invested, and instructions given
Operationalisation of search strategies	List of concatenated search terms
Reference sets used to assess recall	Description of reference sets and how they were constructed
Random sampling of the reference sets	Sample publications from the reference sets
Source database selection	Database, edition, and any additional parameters used
Enhancement techniques	Detailed description of methods used to enhance the dataset

Random sampling of retrieved publications	Samples of publications retrieved before and after enhancement

To begin with, the set of seed keywords selected by the four methods were very different with up to a ten-fold difference in the number and type of keywords used. This level of difference is of primary interest and raises questions around interpretation of the goal. Each method used experts or analysts with familiarity of the topic to select the keywords so why were they so different? The manual element of building the search strings is crucial because human decision controls which terms are included and how they are combined. People with deep knowledge of the field will be likely to produce terms of a more technical nature. These terms will increase recall while minimising less relevant publications found by broader terms. However, none of the methods provide the identity of the experts, say how they were selected, how much time they spent, the precise instructions they were given, or how they resolved differences in expert opinion.

This missing information is key because experts might differ in their precise field of expertise. Some will be more knowledgeable about technical details of the problems surrounding SDGs and select more technical terms. Others might be more familiar with the details of the climate agreements and choose more policy related terms. Even experts with similar levels of knowledge will have their own views as to what is relevant and what is not. For instance, is research on nuclear energy relevant or not in the context of SDG 13? What about medicine's role in mitigating the effects of climate change on health? Each expert will have their own views on these and other questions, and the choices they make is a likely source of divergence in publication retrieval.

Similarly, the combination of keywords is of great importance and construction of the search queries varied between the four methods. Each method used a combination of broad, collective search terms that increase recall, and highly technical terms designed to maximise precision. Broad terms are good for recall but raise the prospect of contaminating the final dataset with less relevant papers. Elsevier 2021 and SIRIS used many more terms than the other methods. They included highly specific, technical terms that found publications in more concentrated fields. Search strategies that use many narrow, specific terms might produce precise datasets but require many more such terms to build up a corpus of publications.

The creation of thematic datasets undoubtedly involves an element of subjectivity due to the human dimension. For topics as complex as the SDGs, this is even more challenging due to the diverse nature of subject matter experts. Their expertise will always be different making it difficult for them to reach reliable consensus on what research is relevant and what is not. Under such circumstances, operationalising a specific definition or interpretation of SDG 13 in the form of a reference dataset is critical. This reference set of publications can be used to test the recall, i.e., what share of the reference set is retrieved by the implemented search queries? The query can be systematically expanded until a certain minimum threshold of recall is reached. The query should be tweaked during this process to keep precision above a defined cut-off point. Such reference sets can be made up of specialist journals, specialised research groups, or publication clusters highly relevant to the target literature. Both the selection of the reference set and the recall rate will influence the outcome and overlap with other methods. Even if difference methods started with the same interpretation of an SDG, they would still produce different results because of discrepancies between their operationalisation processes. Unfortunately, we know too little about how each method operationalised their searches limiting our ability to compare them. This requires further investigation to reveal the causes of the low level of overlap between the methods reported in this study as well as to guide future work on the development of such datasets.

The four methods used three different databases between them. Web of Science is the most selective of the three and aims to capture scholarly literature from high impact sources. Meanwhile Scopus has become more inclusive, and Dimensions searches an even larger corpus of literature. That STRINGS and SIRIS had relatively small coverage surpluses against the other methods confirmed expectations as Web of Science indexes fewer publications than the other databases. Conversely, Dimensions found tens of thousands of additional publications because they are not indexed in Scopus or Web of Science. This is expected because we know that Dimensions covers many publications not indexed in Scopus or Web of Science (Visser et al., 2021).

The databases have different coverage policies and the publications indexed therefore vary. So even running the same search query on different databases will produce different datasets depending on the emphasis of coverage. Consequently, differences in topic emphasis identified in this study might easily be due to the choice of data source rather than nuances of the search queries. It should be noted that the SIRIS approach was designed to be database agnostic. We applied the SIRIS search strategy to the Web of Science but would expect different results if the same strategy were applied to other bibliographic databases. One method of isolating the impact of database is to run the keyword searches of one method across different databases. The Bergen Group (Armitage et al., 2020a) made an attempt at this by translating Elsevier's 2019 keyword search strings into Web of Science syntax. However, this requires great skill, is not always possible, and might raise questions over differences in understanding between the original author and the translator – a common problem in language translation (van Nes et al., 2010). In our study, we isolated the impact of data source by separating the surplus into that caused by data source and the remaining portion that we could attribute the rest of the method.

Finally, all the methods enhanced their datasets but each in a different way. Elsevier 2021 and Dimensions used experts in multiple rounds of relevance checking and then used machine learning algorithms to increase recall. STRINGS added or removed publications depending on whether they were in a relevant topic cluster of publications. SIRIS employed natural language processing at an early stage to produce a long and specific list of technical keywords. The effect of these enhancements could be assessed in a series of controlled studies that only assess the effect of the enhancement. For example, Elsevier has documented (Rivest et al., 2021) a comparison between its pre- and post-enhancement datasets and some details behind the machine learning algorithms. We could potentially use the enhancement technique of one of the methods and apply it to the keywords of each method in the comparison. At this stage, we do not have access to all the details of all the enhancement methods and therefore did not attempt analysis of the enhancements or their impacts.

We have established that great differences between datasets exist but what do they mean? The differences described above will necessarily compound one another to produce the datasets and it is perhaps not surprising that they overlap so little. Publications found by one method but not another might be intentional. Each of the methods involves human decision based on interpretation of the intended outcome, selection of relevant keywords, and construction of the search strings. There may be legitimate differences between the understanding and aims of one group of experts and another. To properly identify the source of the differences between datasets, we need to analyse each stage of the identifications methods in isolation to better understand their contribution to the overall differences. In the present study, the comparison between STRINGS and SIRIS eliminates the effect of the database because they both used Web of Science.

As the study of identifying SDG research intensifies, the methods used will come under greater scrutiny. Any groups designing such methods should therefore fully and publicly document their approach step by step. It is important for readers to know details of the search strategy such as which keywords were used, who selected them, and how they were combined into search strings. Database selection is also important because it will determine which records are available for retrieval and impact the size of the final dataset. Any enhancements should be described in full, and algorithms deposited in a public repository. The more details provided, the easier the method will be to justify. This is an area of growing interest, and peers will be pleased to help improve on methods.

During the course of this study, many questions were raised that could be the subject of follow up studies. The four methods compared were complex and to fully understand their differences would require a systematic controlled comparison at each step. Other limitations to the study should also be considered.

This study focused entirely on one SDG and any conclusions drawn can only be interpreted in that context. As we did not gain a good understanding of the reasons for the discrepancies between methods, we cannot predict whether they would be similar if we used a different SDG in the case study. Broader studies could look at multiple SDGs to detect any patterns.

We chose the DOI as the unique identifier to compare overlapping coverage between data sources because of the extent of its use in academic publishing. Most records in the SDG datasets have a DOI. However, a small fraction of records was not included in the comparisons because they did not have a DOI.

There is also a small share of publications with discrepancy in the publication year between different bibliographic databases. Both Scopus and Web of Science assign the publication year of an article as the official date of publication of the journal issue. Dimensions assigns the publication year based on the date the article was first available – usually the online version (Digital Science, 2021). Consequently, our datasets may exclude a small number of publications from Elsevier, STRINGS, and SIRIS from the latter part of the time window while including the same records in Dimensions. Likewise, some records at the beginning of the time window may be included in Elsevier, STRINGS, and SIRIS, but excluded from Dimensions. However, the overall effect of discrepancies in publication year is likely to be small.

3.6 Conclusions

Each of the four methods compared has attempted to identify research related to climate action and produced largely different results. Their search strategies were created using human judgement and ranged from broad and simple to technical and focused. Between the four methods, three different bibliographic databases were used, each with their own unique coverage. Finally, in some cases, machine learning and other artificial intelligence techniques were applied to enrich the final publication datasets.

These findings support those by earlier work by (Armitage et al., 2020a) and build further by comparing four methods and visualising their outputs in the context of their search strategies. This study also shows the relative contribution of search strategy and data source to the different publication datasets produced.

Using broader data sources to apply the search strategies increases the number of documents returned simply because of the larger coverage. Dimensions comprises more documents than either Scopus or Web of Science and might offer benefits to some methods, especially those aiming to find relevant literature beyond the constraints of highly selective journal literature. The STRINGS method makes a deliberate attempt to find search terms from grey literature and web forums. It might therefore be logical to apply these search terms against the broadest possible data source, i.e., Dimensions.

The search strategy, use of subject matter experts, and data source vary between the four methods. Each method therefore produces a different set of publications related to SDG 13. The fact that we have several different answers to the same questions produces a major implication. The overlap in publications found by these different methods is too low to be adopted by policy makers without careful method selection. The choice of method will potentially define the resulting dataset more than any other factor. Any comparison between research entities should use the same method of identifying publications. As more studies on research into climate change appear in the literature, readers should avoid the temptation to draw hasty conclusions. Published assessments of SDG-related research should state the method used along with other variables such as the time period and data source. The method used is an important influencer of the number and type of resulting publications.

Acknowledgements

Ton van Raan and Ludo Waltman for invaluable input and expert guidance throughout the study, Ed Noyons for help with creating the STRINGS and SIRIS datasets, Elsevier ICSR Lab for providing the Elsevier (2021) dataset related to SDG 13. Stephanie Faulkner (Elsevier), Ed Noyons and Ismael Rafols (STRINGS), Francesco Massucci (SIRIS), Juergen Wastl (Dimensions), and two anonymous reviewers for extensive feedback and helpful comments on earlier versions of the article.

Data availability

The Scopus and Dimensions publication data used in this paper have been made freely available to CWTS for research purposes, while the WoS data have been made available to CWTS under a paid License. We are not allowed to redistribute the publication data used in this study, however the search terms presented in table 3.4 and sample DOI data presented in table 3.5 are made available in Zenodo (Purnell, 2022a).

Conflict of interest/competing interests

The author is affiliated with the Centre for Science and Technology Studies (CWTS) at Leiden University. Colleagues at CWTS have been involved in the development of the STRINGS method.

Chapter 4

Transdisciplinary research: How much is academia heeding the call to work more closely with societal stakeholders such as industry, government, and nonprofits?[3]

[3] This chapter is based on:

Purnell, P. J. (2024). Transdisciplinary research: How much is academia heeding the call to work more closely with societal stakeholders such as industry, government, and nonprofits? https://doi.org/10.48550/arXiv.2408.14024

Abstract

Transdisciplinary research, the co-creation of scientific knowledge by multiple stakeholders, is considered essential for addressing major societal problems. Research policy makers and academic leaders frequently call for closer collaboration between academia and societal stakeholders to address the grand challenges of our time. This bibliometric study evaluates progress in collaboration between academia and three societal stakeholders: industry, government, and nonprofit organisations. It analyses the level of co-publishing between academia and these societal stakeholders over the period 2013-2022. We found that research collaboration between academia and all stakeholder types studied grew in absolute terms. However, academia – industry collaboration declined 16% relative to overall academic output while academia – government and academia – nonprofit collaboration grew at roughly the same pace as academic output. Country and field of research breakdowns revealed wide variance. In light of previous work, we consider potential explanations for the gap between policymakers' aspirations and the real global trends. This study is a useful demonstration of large-scale, quantitative bibliometric techniques for research policymakers to track the impact of decisions related to funding, intellectual property law, and nonprofit support.

Keywords

Transdisciplinary research — research stakeholders – industry – government – nonprofits

4.1 Introduction

The co-creation of knowledge focusing on complex social problems by multiple stakeholders has become known as transdisciplinary research (Hernandez-Aguilar et al., 2020) and is ideally suited to address issues that transcend scientific disciplines (Hirsch Hadorn et al., 2006). Research policymakers have recognised the potential of transdisciplinary research in tackling complicated societal challenges and frequently advocate greater research cooperation between academia and industry (Banal-Estañol et al., 2015; National Institutes of Health, 2012, 2013) citing important contributions to the economy and demonstrable benefits to all stakeholders (Lee, 2000).

Academic leaders recognise this potential and frequently call for closer collaboration with industry to bring together the 'cleverest scientists and the smartest entrepreneurs' (e.g., The Russell Group of Universities, 2023) and motivate their own academics to pursue partnerships (e.g., ETH Zürich, 2024). Large, prestigious academic institutes see such collaborations as markers of success (e.g., Tsinghua University, 2024), and appeal directly to industry stakeholders to initiate partnership proposals (e.g., Stanford University, 2024).

In parallel, collaboration between academia and government is seen as an effective method of transferring subject matter expertise to policymakers through appointing academics onto advisory boards or hiring them into dual employment positions (Glied et al., 2018; Haddon & Sasse, 2019). Meanwhile nonprofit organisations are increasingly working in partnership in sometimes chaotic, but nevertheless crucial partnerships with academia and other public bodies to address complex social, environmental, and economic problems (Cornforth et al., 2014). In each of these tandem arrangements, the potential of transdisciplinary research to address complex challenges facing communities is the driver behind the collaboration.

When the three stakeholders: academia, industry, and government work together, the collaboration is known as the ‘triple helix’ (Etzkowitz & Leydesdorff, 1995; Leydesdorff, 2003). The resulting inter-stakeholder co-creation of knowledge is said to offer benefits over the same sectors working in isolation (Etzkowitz & Leydesdorff, 2000) and has been proposed as an effective method of addressing the grand challenges of our time. The model has been extended to include representatives of civic society such as nonprofit organisations and the configuration of all four stakeholders working together is known as the quadruple helix (Arnkil et al., 2010).

Despite all the calls for a shift towards transdisciplinary research and inter-stakeholder collaboration, there is little evidence of progress on a global scale. Much of the literature on transdisciplinary research has focussed on the concept, definition, and the precise relationship between the stakeholders using case studies rather than the resulting research output (Carayannis et al., 2014; Marijan & Sen, 2022). Some studies have quantified papers that include the term ‘transdisciplinary research’ (Hernandez-Aguilar et al., 2020), however, the mere presence of the term does not mean the paper represents transdisciplinary research as defined as including contributors from multiple research stakeholders. Papers that mention transdisciplinary research are therefore not necessarily real examples of transdisciplinary research.

This study aims to use determine the extent of progress in academic collaboration with three societal stakeholders: industry, nonprofit organisations, and government over a 10-year period. Specifically, we will quantify the increase in collaborative publications in both absolute terms and as a share of the overall academic output. We expect the findings will serve as useful feedback that will help policymakers determine the success of their calls for greater inter-stakeholder collaboration. We will also conduct a breakdown of results for each of the collaboration partners by country and by field to see whether any region-specific or field-specific lessons can be learned.

4.2 Literature review

4.2.1 Modes of research and transdisciplinarity

The application of transdisciplinarity to scientific research accelerated in the 1990s in response to a perceived disconnect between the scientific community and real-world problems (Hernandez-Aguilar et al., 2020), “The world has problems, but universities have departments” (Brewer, 1999). Publication of *The new production of knowledge* by (Gibbons et al., 1994) was a key moment because the authors described the distinction between two types of research; traditional ‘Mode 1’ research in which universities behave autonomously conducting research in strict disciplines and isolation from society, and ‘Mode 2’ knowledge production through transdisciplinary collaboration by temporary research teams comprising experts with diverse backgrounds and knowledge.

Transdisciplinarity can be therefore considered a characteristic of mode 2 research (Gibbons et al., 1994) and its definition is the subject of much debate. Indeed, the lack of a universally accepted definition may offer an opportunity for evolution of the term and practice with the participation of stakeholders themselves. A longitudinal study is currently under way to assess the perspectives of diverse actors in a transdisciplinary project (Thompson et al., 2017) which may contribute to the debate. General consensus of the transdisciplinary concept describes scientists working closely with other society stakeholders to conceive, conduct, and publish research (Mauser et al., 2013), thereby co-creating knowledge with society, rather than for society (Greenhalgh et al., 2016; Seidl et al., 2013). Solutions to the grand challenges of our time such as those described by the UN sustainable development goals (SDGs) (United Nations General Assembly, 2015) will require collaborative approaches involving diverse stakeholders and decision makers (Hernandez-Aguilar et al., 2020; Thompson et al., 2017).

4.2.2 Inter-stakeholder collaboration

Many have suggested that university – industry collaboration is good for innovation and the economy (Savage, 2017) although motivation for industry scientists may be different from that of academics. Some argue that academics who collaborate with industry researchers may experience a decline in publication rates because of the company's interest in protecting research results through patents rather than publishing it. Czarnitzki et al. (2015) showed that scientists in Germany experience more than twice as much delay and secrecy, jeopardising publication disclosure of academic research when projects use industry sponsorship. Perkmann & Walsh (2009) argued however, the type of collaboration makes a difference and that more basic collaborative projects between university and company scientists results in increased publication.

The difference between academic and industry scientists may result from early career choices defined by underlying preferences and professional goals. It has been suggested that scientists driven by a 'taste for science' will more likely opt for a career in academia (Agarwal & Ohyama, 2012) while those who choose industry are more motivated by financial and other resource incentives along with greater interest in downstream research (Roach & Sauermann, 2010).

Bikard et al. (2018) observed simultaneous discoveries and found that those resulting from academia – industry partnerships were followed by effective division of follow-on work with academics dedicated to publications while company scientists concentrated on protecting intellectual property. University – industry research collaboration has been found helpful for university lecturers provided it doesn't distract them from their main academic research (Manjarrés-Henríquez et al., 2009).

Various studies (e.g., Tijssen et al., 2016) have shown geographical proximity to correlate with university-industry-government collaboration (Ponds et al., 2007). The closer the company is to the university, the more likely collaboration occurs, and research is published (Bjerregaard, 2010). In 2009, (Tijssen et al., 2009) estimated that 7% of scholarly publications featured both university and industry affiliated authors.

Since the 1990s, the Chinese government has invested consistently in research-intensive universities (Zhang et al., 2013) and at the same time, private Chinese firms have boosted their investment in R&D more than state-owned enterprises (Fang et al., 2017). In 2015, the Chinese government implemented a series of laws and policies designed to enable universities to decide their own technology transfer strategy, which may have stimulated academia – industry partnerships. Academic institutions' intellectual property is considered state-owned property, and its potential improper handling was previously seen as a deterrent to industry collaboration (Zhang & Zou, 2022).

In a US-China comparison, Zhou et al. (2016) found that Chinese universities are more likely than US universities to co-author papers with foreign industry partners. Regardless of the motivation and structure of the partnership, there seems no doubt that academia – industry collaboration is being encouraged and incentivised around the world.

Nonprofit organisations are increasing held accountable through performance assessment (Lee & Nowell, 2014), which usually encompasses financial indicators (Kim, 2016), but not output metrics, which are considered too simplistic (Rey-Garcia et al., 2017). In the book *net positive* (Polman & Winston, 2021), former Unilever CEO Paul Polman provides an entertaining personal account of steering a multinational company through turbulent times. In his view, it is vital for successful businesses to work hand in hand with government, and civic society to ensure mutual benefit, thereby prioritizing stakeholder value over shareholder value. Polman sees potential for businesses to address social challenges but emphasizes the need to do so in cooperation with other stakeholders including governments, nonprofits, and academia. As our challenges become more urgent and interdependent, we need to think of new ways of addressing them. Collaboration with nonprofit organisations will be key to finding solutions and we cannot expect governments, business, or academia to solve these challenges alone.

Government is the predominant source of research funding for the academic sector in most countries, and major changes to research funding policy influence academic publication rates, intellectual property protection, and other innovation indicators (Adams et al., 2005; Beaudry & Allaoui, 2012). The academic – government research relationship is therefore at least partially based on academics' use of government funding grants to conduct research related to its area of interest. Government research grants are sometimes the result of progammatic, mission-oriented agencies that have a need to demonstrate immediately useful research findings to achieve specific goals that do not necessarily fit traditional academic research objectives (Goldfarb, 2008). Data collected by China's Ministry of Science and Technology in 2020 showed that government – academia partnerships have been more successful than other collaboration configurations at business incubation of new R&D institutions (Zhou & Wang, 2023).

In many industrialised countries, the share of university research funding directly supported through government grants is falling while industry support is increasing (Auranen & Nieminen, 2010). In the UK, the government created UK Research and Innovation (UKRI) with the goal of unifying the voice of academia and improving cooperation with government policymakers (Haddon & Sasse, 2019). Germany spent 120 billion Euros on R&D in 2022, twice as much as France and amounting to more than one-third of the entire EU investment, while Switzerland invests 50% more in R&D than the EU average as a share of its GDP (Eurostat, 2024). Government subsidisation of R&D cooperation has stimulated innovation efficiency in Germany's regions (Broekel, 2015) and efforts to incorporate standardisation into the research and innovation process (Zi & Blind, 2015). Russia's government stimulated a competitive university landscape in the 2010s which linked funding to scientific output and impact (Ivanov et al., 2016).

4.2.3 Triple and quadruple helix models

Increasing interdependence between industry, academia, and government has the advantage of optimising conditions for creative thinkers to work alongside innovators and entrepreneurs. In 1995, Etzkowitz and Leydesdorff described the collaborative relationship in knowledge based economic development between three stakeholders of innovation, namely universities, industry, and government, as the 'Triple helix' (Etzkowitz & Leydesdorff, 1995), extending the notion of interdependence from biology. Progressively integrated models see partnership between independent but separate actors gradually transform into a system of overlapping entities generating tri-lateral networks and hybrid organisations facilitated by, but not controlled by government (Etzkowitz & Leydesdorff, 2000; Leydesdorff & Etzkowitz, 1996).

According to Kang et al. (2019), Triple Helix theory was designed for established economies of Europe, the USA, and other western countries. Kang studied collaboration dynamics between stakeholders in China's two main science and technology innovation centres located in Shanghai and Beijing. The study and others attributed a strengthening cooperation between universities and industry to the shift in focus of China's universities from traditional research to an entrepreneurial university model (Kang et al., 2019; Zhu et al., 2022).

Other stakeholders may also enter the collaboration such as civil society, which extends the model to form the quadruple helix (Arnkil et al., 2010; Carayannis & Rakhmatullin, 2014), and the environment, which makes up the quintuple helix. The helix models are not evaluated in this study but may interact with and influence conclusions drawn from studies of societal stakeholder collaboration with academia (Carayannis et al., 2020).

Early comparisons of triple helix activity between countries used combined search terms for the local language translations of the words university, industry, and government (Park et al., 2005). Macro-level quantitative analyses may offer useful information, but public-private co-publication indicators need to be valid, reliable, and robust in order to form the basis for policy decision making (Tijssen, 2012).

4.3 Data and methods

4.3.1 Data

We sourced scholarly publications from the Dimensions database because of its broad coverage (Hook et al., 2018) and because it may index some relevant content not covered by other bibliometric databases (Paez, 2017; Visser et al., 2021). Specifically, we used the Dimensions database hosted by CWTS at Leiden University and focused on records published between 2013 and 2022 (10 full years).

The affiliations in Dimensions are mapped on to the Research Organization Registry (ROR) (https://ror.org/registry/), which defines more than 100,000 research organisations by type. Every organisation is manually assigned to an organisation type as described in Table 4.1.

Table 4.1 Organisation types in the Research Organization Registry

Organisation type	Description
Education	A university or similar institution involved in providing education and educating/employing researchers.
Healthcare	A medical care facility such as hospital or medical clinic. Excludes medical schools, which should be categorized as "Education".
Company	A private for-profit corporate entity involved in conducting or sponsoring research.
Archive	An organisation involved in stewarding research and cultural heritage materials. Includes libraries, museums, and zoos.
Nonprofit	A non-profit and non-governmental organisation involved in conducting or funding research.
Government	An organisation that is part of or operated by a national or regional government and that conducts or supports research.
Facility	A specialized facility where research takes place, such as a laboratory or telescope or dedicated research area.
Other	Any organisation that does not fit the categories above.

Source: ROR Data Structure

4.3.2 Methods

In order to study the collaboration between stakeholder types, we analysed the author affiliations on the papers in our dataset. In this study, a paper was linked to a stakeholder type only if a minimum proportion of its author affiliations were linked to that stakeholder type. In the first analysis, we tested the effect of setting different thresholds for counting the paper as being authored by a given stakeholder type. We considered four thresholds: 0% (one single author affiliation), 10%, 20%, and 30% of the author affiliations in a paper. For the rest of the study, we set the minimum threshold at 20% of the author affiliations. For instance, if a paper had 10 author affiliations, then it only counted as an academic paper if at least two of the author affiliations were with academic institutions.

As the vast majority of published research papers feature academic affiliations, we began our study by retrieving all papers in which at least 20% of the author affiliations were linked to academic institutions. This was defined as the overall academic dataset. Within the academic dataset, we then identified all those papers that featured collaboration with another stakeholder type (i.e. academia – industry, academia – nonprofit, or academia – government). If a paper features a minimum 20% author affiliations with academic institutions and a minimum 20% affiliations with industry, then it counts as academia – industry collaboration. Conversely, in the case that a paper with 25 author affiliations featured only one industry affiliation, it would not be counted as an industry collaboration because the industry contribution is not sufficient to meet the 20% affiliation threshold. For academic collaboration with each of the other stakeholder types, we computed the number and proportion of collaborative papers for each year in the period 2013-2022.

We then conducted further analyses of academic collaboration with other stakeholder types by country. We focused on the 25 countries with most academic publications along with the world and European Union as benchmarks (Table 4.2). In the country analysis, we based the country only on the academic affiliations. For instance, if a paper features academic affiliations from USA and China, and industry affiliations from India and Brazil, then it would only count as a paper for the USA and China. This enabled us to view international collaboration from the perspective of the academic community in a given country.

Table 4.2 Countries and their abbreviations

Country/territory	Abbr.	Country/territory	Abbr.
United States	US	Russia	RU
European Union	EU	France	FR
China	CN	Iran	IR
United Kingdom	UK	Indonesia	ID
Japan	JP	Turkey	TR
Germany	DE	Netherlands	NL
India	IN	Poland	PL
Canada	CA	Sweden	SE
Italy	IT	Taiwan	TW
Australia	AU	Switzerland	CH
Brazil	BR	Belgium	BE
South Korea	KR	Malaysia	MY
Spain	ES	Denmark	DK

Similarly, we conducted a collaboration analysis for each of the stakeholder pairs (academia – industry, academia – nonprofit, or academia – government) based on the field of research of the paper (Table 4.3).

Table 4.3 Fields of research and their abbreviations

Field of Research	Abbr.	Field of Research	Abbr.
Agricultural, Veterinary & Food Sciences	AG	Environmental Sciences	EV
Biological Sciences	BS	Health Sciences	HS
Biomedical & Clinical Sciences	BC	History, Heritage & Archaeology	HH
Built Environment & Design	BE	Human Society	HU
Chemical Sciences	CS	Information & Computing Sciences	IC
Commerce, Management, Tourism & Services	CM	Language, Communication & Culture	LC
Creative Arts & Writing	CA	Law & Legal Studies	LL
Earth Sciences	ES	Mathematical Sciences	MS
Economics	EC	Philosophy & Religious Studies	PR
Education	ED	Physical Sciences	PS
Engineering	EN	Psychology	PY

For academia – industry collaborative papers, we conducted an additional analysis to identify the companies most frequently involved in the collaboration for selected countries. We selected the two countries with steepest growth in relative academia – industry papers, and the two countries with steepest decline (provided the share of academia – industry papers exceeded 1% of the overall academic output for that country in 2022). We then identified the eight companies most frequently represented on the papers coauthored with academia in 2013 and the eight companies most frequently represented in 2022.

4.4 Results

4.4.1 Minimum author affiliation threshold

Stakeholder participation in collaborative papers was highest when no threshold was introduced, i.e. when one single author affiliation was sufficient to count a paper as belonging to a given stakeholder type (Figure 4.1). When 10% author affiliations were required to count the paper as a collaboration the share of collaborations began to drop. As the threshold of author affiliations for counting papers as collaborations grew, the share of collaborations dropped as expected. All results presented in the remainder of the study are based on a threshold of 20% author affiliations.

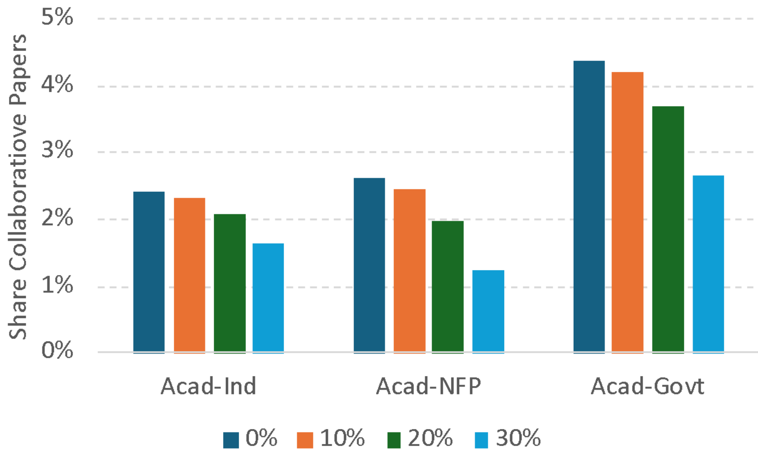


Figure 4.1 Share of authors to define stakeholder participation

4.4.2 Collaboration between stakeholders

In Figure 4.2, we show the number of papers that feature affiliations from both academic and industry organisations as a bar chart. The line graph in the same figure represents the collaborative papers as a proportion of the overall number of academic papers. While the number of collaborative papers has grown nearly 40% over the past 10 years, the share of academic papers with industry participation has dropped by 16%. This can be explained by the even greater growth in academic output that does not feature industry collaboration.

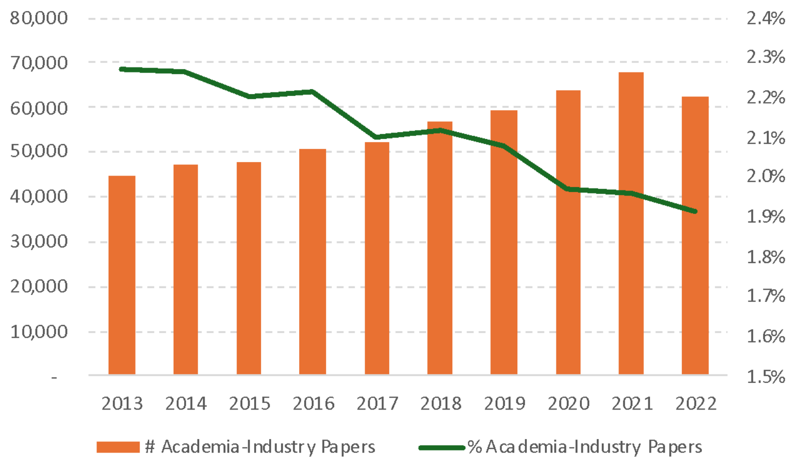


Figure 4.2 Academia - industry collaboration

In Figure 4.3, we show the same time series for academic papers that feature co-authors from nonprofit organisations. We see a similar pattern to the academia - industry collaboration described in figure 4.2.

The number of collaborative papers between academia and nonprofit (figure 4.3) has grown almost 60% over the past 10 years, while the share of academic papers with nonprofit participation has dropped slightly. This means the number of collaborative papers between academia and nonprofit sector has grown slightly less than the overall academic output.

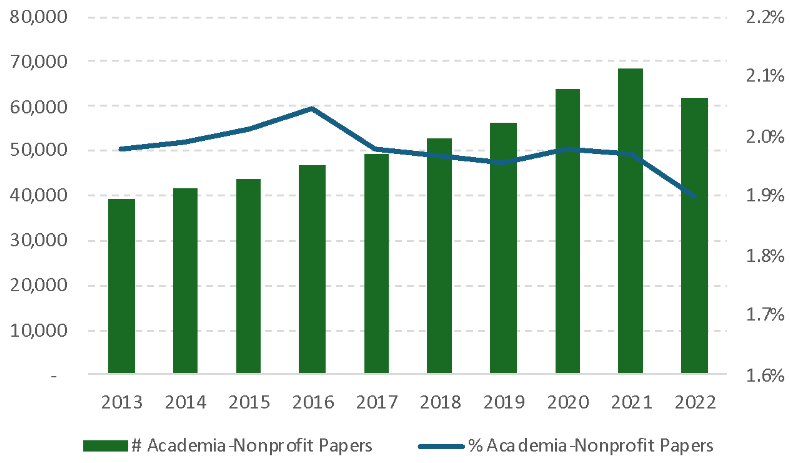


Figure 4.3 Academia - nonprofit collaboration

In Figure 4.4, we show the same time series for academic papers that feature co-authors from government organisations. Here we see that collaborative papers between academia and government have grown at almost 70% over the past 10 years. That growth has resulted in a 3% increase in the proportion of collaborative papers relative to the overall academic output.

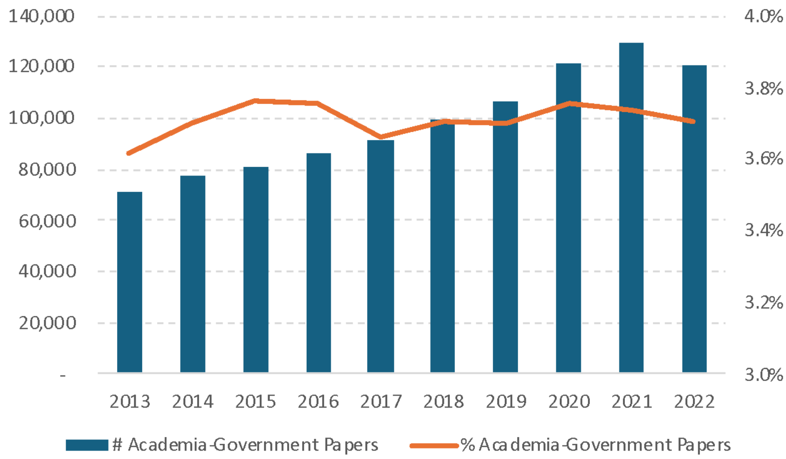


Figure 4.4 Academia - government collaboration

4.4.3 Country trends

In the next part of the analysis, we determined the 10-year collaboration trend for the 25 most productive countries along with the world and EU trends as benchmarks. For each country or territory, we determined the number of collaborative papers between academia and one of the other stakeholders as a proportion of the academia affiliated papers. In Figure 4.5, we show the share of academia – industry collaborative papers in 2013 and then again in 2022. A dot above the diagonal signifies an increase in share of academia – industry collaboration in that country or territory over the 10-year time period, while dots below it signify a decrease. China and Belgium, an outlier within the EU, showed substantial increases in academia – industry collaboration over the 10 years, while Australia, Taiwan, and Turkey appear to have maintained roughly the same share as 10 years earlier. For most countries, however, the figure shows a decline in academia - industry collaboration, in line with the overall declining trend shown in Figure 4.2. We will provide a more detailed analysis of these findings in section 4.4.5.

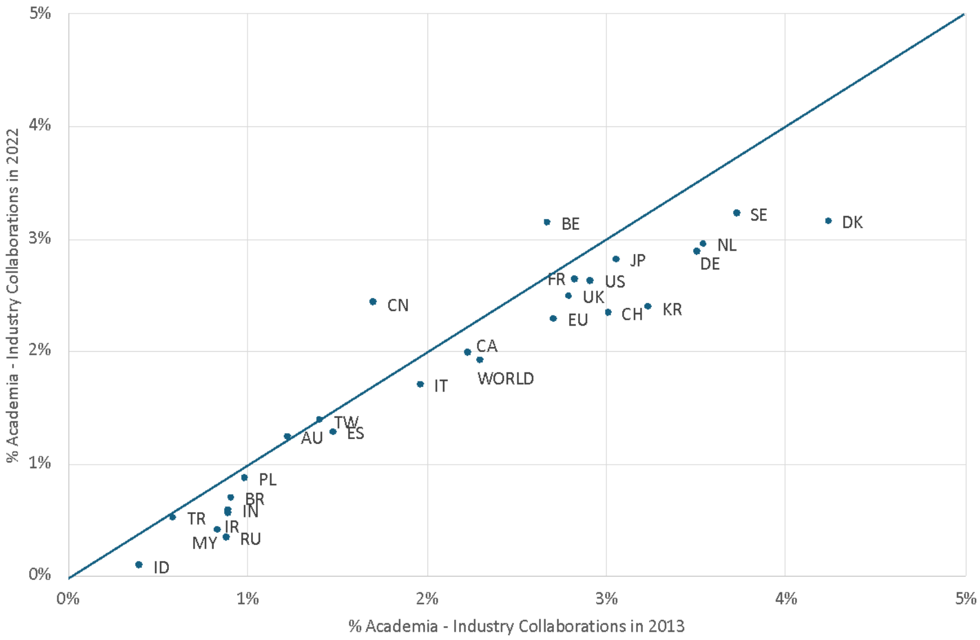


Figure 4.5 Academia - industry collaboration trend by country

Next, we present the 10-year trend for academia – nonprofit collaborations as a proportion of the overall academic output (Figure 4.6). Here we see the greatest gains in Spain, the United Kingdom, and Iran, with most of the other countries making modest growth. The most notable declines were in Taiwan, Indonesia, Belgium, and Russia.

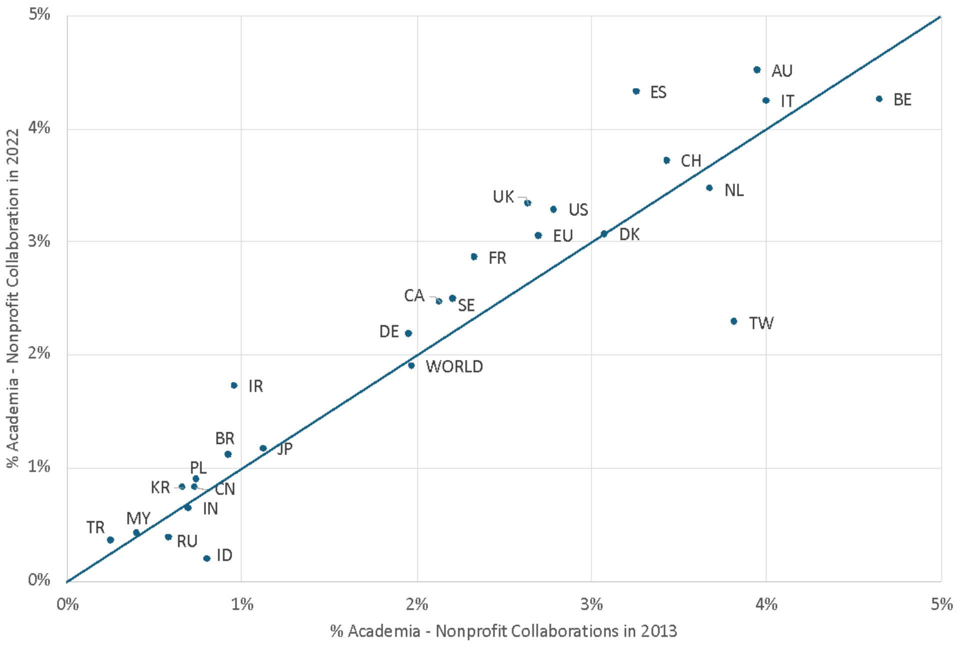


Figure 4.6 Academia - nonprofit collaboration trend by country

Next, we present the 10-year trend for academia – government collaborations as a proportion of the overall academic output (Figure 4.7). The picture for academia – government collaboration showed wide variance. The largest increases were seen in China, Malaysia, and the United Kingdom. The steepest declines were seen in France and Russia and Indonesia, Japan, Turkey, and Taiwan.

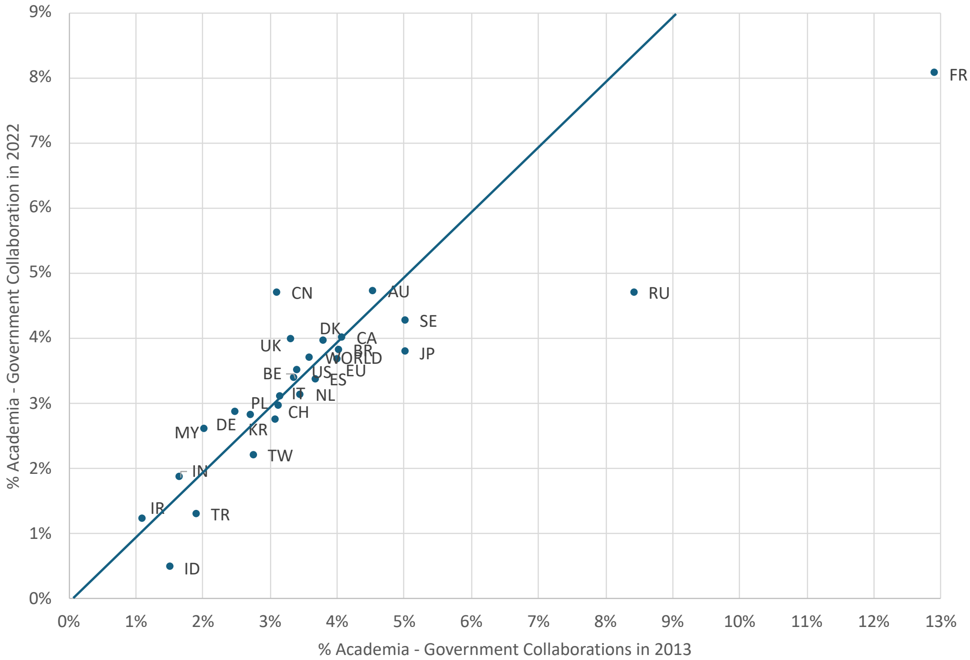


Figure 4.7 Academia - government collaboration trend by country

4.4.4 Field trends

We also analysed the trends of research for academic collaboration with each of the stakeholder types by subject field. In Figure 4.8, we see that the share of academia - industry collaborations with respect to overall academic output has fallen for all fields of research except earth sciences, which grew, and mathematical sciences, which retained almost the same share. Steeper declines were in the social sciences and humanities fields including creative arts & writing, law & legal studies, history, heritage, & archaeology. The science fields including engineering and the physical, biological, and chemical sciences declined less.

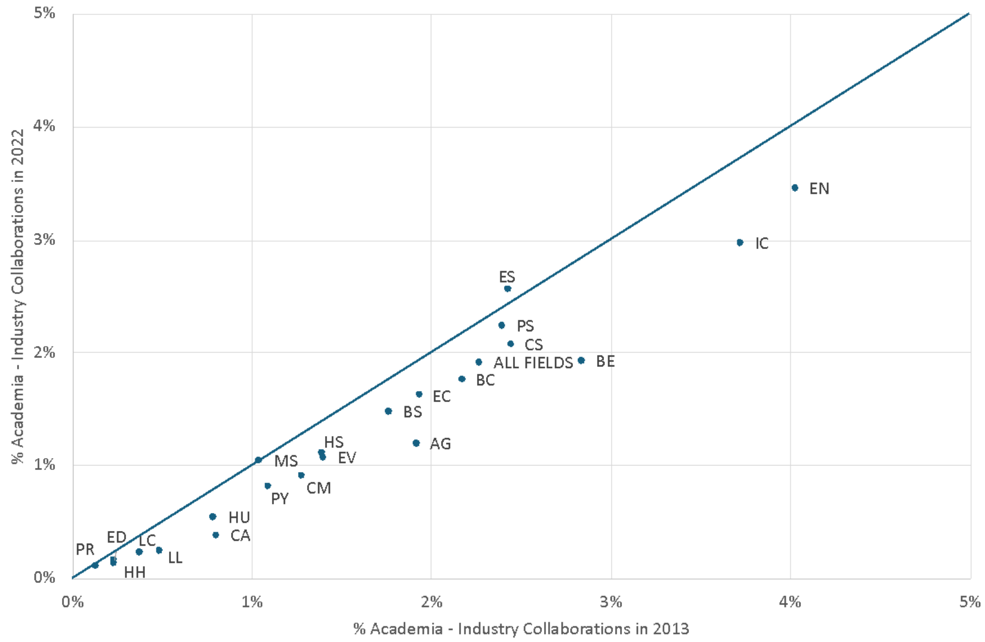


Figure 4.8 Academia - industry collaboration trend by field

The 10-year change in collaboration between academia and nonprofit organisations relative to the overall academic output is shown in Figure 4.9. Here the picture is rather diverse with marked increases in relative collaboration in philosophy & religious studies, history, heritage, & archaeology, and built environment & design. The steepest declines were noted in the creative arts & writing and the law & legal studies. The sciences tended to vary less and maintain a stable share of collaboration over the 10-year period.

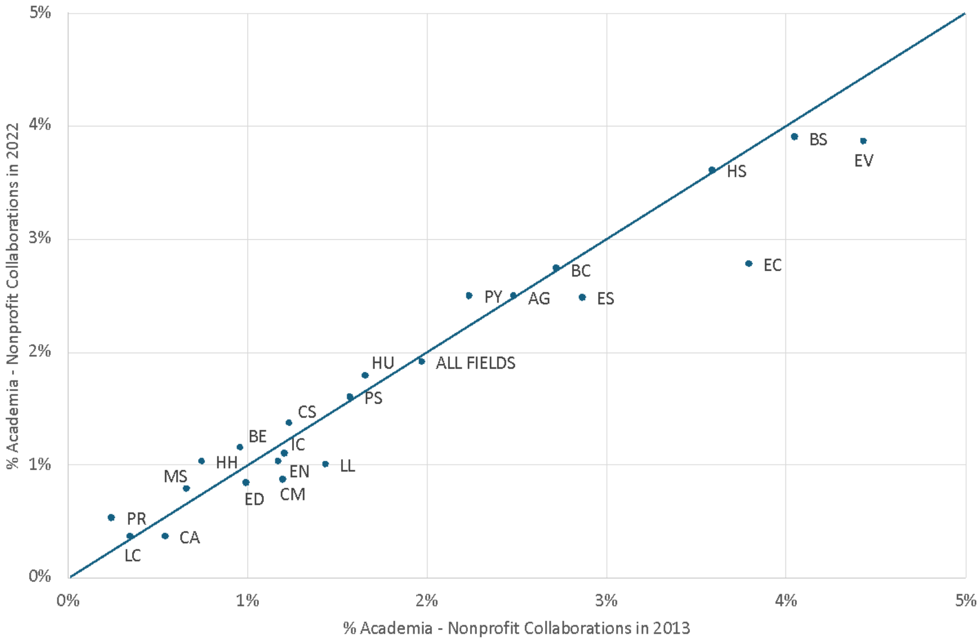


Figure 4.9 Academia - nonprofit collaboration trend by field

The collaboration between academia and government relative to the overall academic output is shown in Figure 4.10. In most fields, there was an increase in relative collaboration, most notably in the law & legal studies, philosophy & religious studies, built environment & design, history, heritage, & archaeology, and engineering. The steepest declines were seen in commerce, management, tourism, & services, economics, and environmental sciences. The sciences again maintained a relatively stable share.

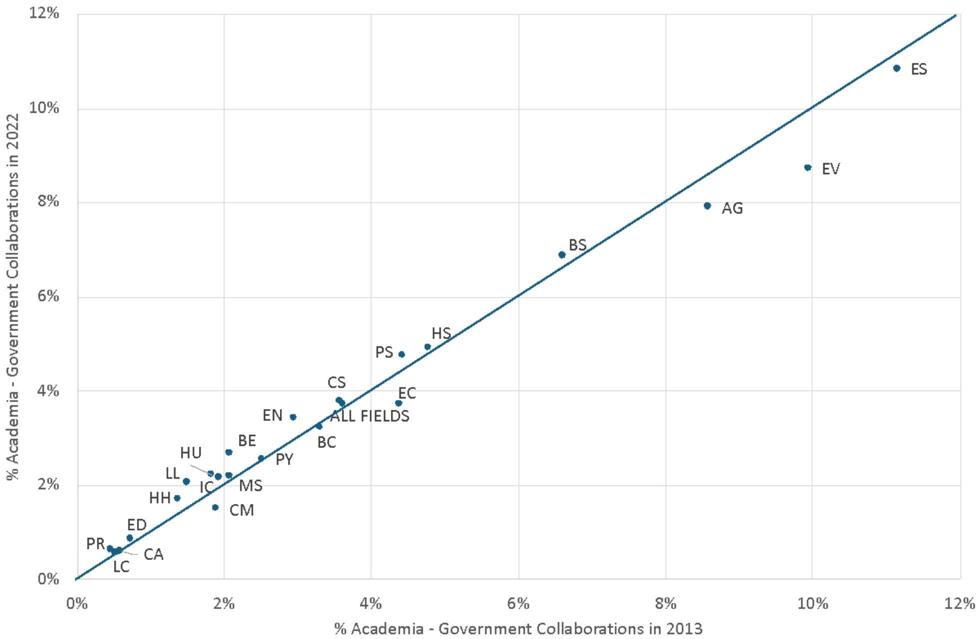


Figure 4.10 Academia - government collaboration trend by field

4.4.5 Industry partner analysis for selected countries

The findings presented in figures 4.3 and 4.4 show that academia – nonprofit and academia – government collaboration grew at roughly the same rate as overall academic output. However, figure 4.2 shows that academia – industry collaboration dropped by 16% relative to academic output. The latter was an interesting and surprising result in the context of the frequent calls on academia to close the gap with industry. The country-level findings presented in figure 4.5 could potentially shed light on whether the relative decline is a global one or if regional differences play a role. We therefore conducted an additional analysis of the academia – industry collaborations to identify the eight companies with which academia coauthored the most papers in the two countries with steepest growth (China and Belgium) and in the two countries with the steepest decline (South Korea and Switzerland) (table 4.4).

Table 4.4 Industry partners in academic collaboration: China and Belgium

China 2013		China 2022	
Company name	Papers	Company name	Papers
State Grid Corporation of China (China)	682	State Grid Corporation of China (China)	3295
China National Petroleum Corporation (China)	242	Sinopec (China)	961
Sinopec (China)	194	China National Petroleum Corporation (China)	875
Microsoft Research Asia (China)	183	China Electronics Technology Group Corporation (China)	754
China Electronics Technology Group Corporation (China)	149	China Shipbuilding Industry Corporation (China)	593
China North Industries Group Corporation (China)	132	China North Industries Group Corporation (China)	545
Aviation Industry Corporation of China (China)	127	China Southern Power Grid (China)	509
China National Offshore Oil Corporation (China)	115	Aviation Industry Corporation of China (China)	463
Belgium 2013		Belgium 2022	
Company name	Papers	Company name	Papers
Janssen (Belgium)	50	Flanders Make (Belgium)	175
Siemens (Belgium)	14	Sciensano (Belgium)	153
GlaxoSmithKline (Belgium)	12	Janssen (Belgium)	41
UCB Pharma (Belgium)	12	Sanofi (United States)	19
Novartis (Switzerland)	11	Siemens (Belgium)	15
Bekaert (Belgium)	9	Regeneron (United States)	12
Holst Centre (Netherlands)	9	Sanofi (France)	11
Novartis (United States)	8	Icometrix (Belgium)	10

In China, the top eight collaborating companies in 2022 coauthored more than four times the number of collaborations as the top eight companies in 2013. The growth appears to be generic as all the top collaborating companies in 2022 were already collaborating with academia in 2013. However, the number of co-authored papers has increased by several-fold over the 10-year time period. The biggest contributor to China's rise was by the State Grid Corporation of China, whose co-authored publication count grew nearly five-fold in the 10-year period studied.

In Belgium, collaboration with the top eight companies tripled between 2013 and 2022. The increased collaboration was largely due to over 300 collaborative papers with two companies, Flanders Make and Sciensano in 2022, with whom academia had no or negligible collaboration in 2013. Flanders Make supports the local manufacturing industry through innovative research projects to facilitate its customers' transition to industry 4.0. Flanders Make currently has open calls for research partnerships on strategic basic research projects such as 'Development and validation of a high-torque-dense actuator', and on industrial research valorisation and acceleration projects such as 'Situation-aware robust optimal vessel navigation and control' (Flanders Make, 2024). Sciensano is a healthcare company launched in 2018 that is also registered as a research institution under the Belgian Science Policy. The company aims to improve quality of life for humans and animals along with their shared environment in Belgium. Sciensano publishes details of projects on projects such as 'Research on PFAS contamination in the food chain', 'A citizen-driven crowdsourcing and feedback system to stimulate research and policy relating to Flemish and local food environments', and 'Development, testing, and implementation of the Belgian Patient Reported Experience measure for PAncreatic cancer caRE' (Sciensano, 2024).

For many countries there was a relative decline between 2013 and 2022 in collaboration between academia and industry (figure 4.5). The two countries showing the steepest decline were South Korea and Switzerland. In table 4.5, we present the top eight companies by number of publications co-authored with academic institutions in South Korea in 2013 and in 2022 followed by the corresponding companies collaborating with academic institutions in Switzerland.

Table 4.5 Industry partners in academic collaboration: South Korea and Switzerland

South Korea 2013		South Korea 2022	
Company name	Papers	Company name	Papers
Samsung (South Korea)	730	Samsung (South Korea)	587
LG Corporation (South Korea)	155	Hyundai Motors (South Korea)	120
Pohang Iron and Steel (South Korea)	78	LG Corporation (South Korea)	112
SK Group (South Korea)	68	SK Group (South Korea)	97
Korea Electric Power Corporation (South Korea)	58	Pohang Iron and Steel (South Korea)	87
Amorepacific (South Korea)	51	Hyundai Motor Group (South Korea)	53
Hyundai Motors (South Korea)	32	Korea Electric Power Corporation (South Korea)	51
CJ CheilJedang (South Korea)	26	Amorepacific (South Korea)	32
Switzerland 2013		**Switzerland 2022**	
Company name	Papers	Company name	Papers
Novartis (Switzerland)	51	Roche (Switzerland)	61
Roche (Switzerland)	31	Novartis (Switzerland)	54
ABB (Switzerland)	28	Inspire	24
Walt Disney (Switzerland)	26	Roche (United States)	21
Nestlé (Switzerland)	25	Microsoft (United States)	20
Microsoft (United States)	16	Fluxim (Switzerland)	15
Siemens (Germany)	16	Novartis (United States)	14
Novartis (United States)	11	Philochem (Switzerland)	14

In South Korea, the number of collaborative papers with top eight companies was slightly less in 2022 than it was in 2013. There was very little change in the companies participating in the collaborations, indeed the top seven companies in academic collaborations in 2013 were still among the top eight in 2022. The number of co-authored papers had dropped, which contributed to the overall decline in academic – industry collaboration in South Korea over the 10-year period studied. Notable declines were from Samsung (-20%) and LG Corporation (-28%).

In Switzerland, the number of collaborative papers with the top eight companies actually grew by 8% in absolute terms but declined relative to overall academic output. There were interesting changes in the identity of the leading companies collaborating with academia. From the list of top eight collaborating firms in 2013, ABB, Walt Disney, Nestlé, Microsoft, and Siemens all posted substantial declines in co-authored papers with academic institutions in Switzerland by 2022.

4.5 Discussion

4.5.1 Academia – Industry

We frequently hear research policymakers and higher education leaders citing the need for academia to work more closely with different societal stakeholders to maximise the benefits of research outcomes when addressing society's most pressing goals. Our results show that the number of academia – industry collaborative publications has consistently increased over time in absolute terms but has declined with respect to overall academic output. The idea that academic research in partnership with the private sector will help industry to innovate may have stimulated some collaborative research, but overall academic research output has grown faster without industry partnership than with it. Our findings show there is a wide variation in this trend depending on the country studied, with substantial relative growth in China and notable decline several countries including South Korea and Switzerland. In Europe, only Belgium has seen relative growth in academia – industry collaboration.

In 2019, China became the leading country in USPTO patent filings, with four Chinese universities among the top 10 education institutions in number of filed applications (WIPO, 2020). The impressive trend for growth in industry participation in China may have been influenced by the major government overhaul of laws and regulations surrounding research commercialisation dubbed China's Bayh-Dole Act (Huang et al., 2023; Yi & Long, 2021). In the past decade, universities have now assumed responsibility for implementing their own patents and sharing profits with researchers. This may be responsible for stimulating a spike in Chinese universities commercialising research. Industry funding of universities in China would then create conditions that could explain our reported increase in relative participation of industry in scientific research.

Two Belgian companies were responsible for three-quarters of the top eight companies' academia – industry collaborations. Flanders Make supports local manufacturers and Sciensano is a commercial research organisation that focuses on improving local human and animal health. Academic collaboration with both these companies has started recently and both promote calls for collaborative research projects on their websites. The academic community has clearly responded to calls to work together with these two companies, and policymakers could potentially use the model as an example to encourage other similar collaborations.

Despite South Korea's government boost to stimulate academia – industry collaboration, researchers in the private sector still need to develop a culture of trust in their academic partners (Lee, 2014). Our reported decline in the relative number of academia – industry collaborative papers support previous work (e.g., Jones, 2024). In Switzerland, we did not find any obvious explanation for the decline in academia – industry collaboration.

4.5.2 Academia – Nonprofit

The absolute number of academia – nonprofit collaborative research papers has shown consistent year-on-year growth in the past 10 years that was almost in line with the steep growth in overall academic output. Most of the high-income countries showed relative gains in academia – nonprofit research collaboration, but the country with the steepest increase was Iran. In Iran, it has been argued that citizen engagement on civic societal issues will be among the emerging research themes and is a potential driver behind academia – nonprofit engagement (Farazmand et al., 2019).

4.5.3 Academia – Government

The absolute number of academia – government collaborative research papers shows consistent year-on-year growth in the past 10 years and is almost in line with the overall academic output. Countries with the most marked increase in academia – government collaboration were China, Malaysia, UK, Germany, India, and Iran. Meanwhile, the steepest declines were recorded in Russia, France, and Indonesia.

Relative academia – government collaboration declined on average in EU academic sectors, which is at odds with the EU Research and Innovation policy that aims to 'ensure technological breakthroughs are developed into viable products with real commercial potential – by building partnerships with industry and governments' (European Union, 2020). However, the EU's leading gain was recorded by Germany, whose most prolific government research institution in 2022 was the Helmholtz-Zentrum Hereon, a transdisciplinary centre that spends 100 million Euros annually on support for societal institutions through its expertise in tackling climate change, biomedicine, and new energy systems.

4.5.4 Limitations

In this study we have used three pairwise comparisons, each from the perspective of the academic sector. The main reason for this is because the vast majority of published research is by authors affiliated to academic institutions. Therefore, we used academic output as the denominator in our calculations and collaboration between academia and each of the other stakeholders as the numerator. So, the question we are addressing is about the extent to which academia has collaborated with other stakeholders rather than how much the stakeholders are collaborating with each other. We might be overlooking other scenarios such as industry providing contract research services to government, which would also count as transdisciplinary research but not be counted in this study.

The Dimensions database used for this study relies on the Research Organization Registry (ROR) classification system for defining research stakeholders. It is not always straightforward to classify institutions. For instance, the State Grid Corporation of China was classified as a 'Company', whereas it could be argued that it should be 'Government'. We also noted that the classification of some organisations was changed in ROR during the course of our study. The École Polytechnique Fédérale de Lausanne (EPFL) was initially classified as 'Facility', but later changed to 'Education' and 'Funder'. Its re-classification as 'Education' meant that papers with the EPFL affiliation were included in our study and counted as academia. However, these were isolated discrepancies and we considered them artefacts in the data rather than systematic errors. Other stakeholder types categorised by ROR, such as facilities, museums, and private hospitals, were not analysed in this study, but may be of interest, especially in studies of healthcare and studies of the arts and humanities. Despite these concerns, ROR routinely updates its classification, and we consider it the most appropriate classification system of research organisations for this study. Given the changing and sometimes heterogeneous nature of organisation types, we suggest some flexibility in interpreting the results.

One of the limitations of all bibliometric studies is that we can only analyse research outputs that are published. The academic sector is by far the most prominent contributor to published research among each of the stakeholders studied. This is expected because scholarly publication is one of the most important performance indicators for academics and for their universities (Dill & Soo, 2005; Hazelkorn, 2011; Marginson & van der Wende, 2007). University academics routinely publish in journals, books, and conferences, but other stakeholders may not.

Private sector companies for instance, often conduct research aiming to bring unique products or services to market and consequently make only a fraction of their findings public. This means that we do not know what share of industry research we are counting through scholarly publications. Similarly, government participation in major challenges is important because governments can focus a nation's attention and priority, and because only governments can set national agendas and make laws. Governments habitually publish reports, guidelines, or other types of documents that do not appear in bibliographic databases as scholarly works (Bickley et al., 2020). Scholarly publications from industry, government, and nonprofits are likely to represent only a fraction of the research these actors perform.

Introduction of a 10% minimum threshold of author affiliations to count a stakeholder as having a meaningful contribution to a study reduced the number of papers in our dataset. Raising that minimum threshold to 20% and then 30% accelerated the exclusion of papers. In order to allow our threshold to eliminate the most extreme cases of lopsided collaborative papers, without unduly reducing the dataset, we set the threshold at 20% for the remainder of the study. Although the choice of 20% is somewhat arbitrary, we considered it an improvement over having no threshold at all.

4.6 Conclusion

Research policymakers and academic leaders emphasise the benefits of closing the gap between academia and industry and collaborating more closely with society to address the most difficult challenges of our time. This bibliometric study of transdisciplinary research applies large-scale quantitative techniques to assess collaboration trends between academia on the one hand and industry, government, and nonprofit organisations on the other.

There appears to be a gap between the aspirations of research policymakers who are advocating closer collaboration between academia and other research stakeholders and what is happening in reality. The overall trends along with the picture in many countries and many fields is a relative decline in collaborative research with respect to the rapid growth in research by academia alone. Although collaboration is growing in absolute terms, from a relative point of view our results do not support the notion of increased collaboration, partnerships, or transdisciplinary research.

One possible general explanation for the relative decline is the current system of incentivising academic researchers on their scholarly publications in peer-reviewed journals. Perhaps many academics find it easier and quicker to publish with fellow academics, rather than through societal engagement projects. Collaboration with industry, nonprofits, or government may introduce partners with very different ideas of successful outcomes that do not involve published papers. Faced with conflicting pressures, academics might find the pressure to publish outweighs the need to conduct societal engagement.

We performed our analysis at global and at country level. This allowed us to observe wide variation in regional trends possibly influenced by changes in national research landscapes. A clear example is the opening of Chinese intellectual property laws, which has paved the way for universities to work more closely with industry on commercialisation of research. The resulting boost in academia – industry collaboration was unique to China according to our study.

The data used for this study were from the Dimensions database, which uses the organisation classification provided by the Research Organization Registry (ROR). The ROR organisation type categories are one way of defining stakeholders, but it is open for debate as to whether it is the right or best way. For the purposes of this paper, we have chosen to work with the ROR definitions used by Dimensions for pragmatic reasons and have revealed some interesting, high-level patterns. To drill into further detail, it may be useful to run follow-on studies that consider other data sources and stakeholder definitions.

Bibliometric studies generally look backwards but can reveal patterns that may then be linked to policy changes. Therefore, research policymakers will benefit from retrospective studies when setting out their new strategies on a range of policies such as funding, academic freedom, intellectual property management.

Acknowledgements

Ton van Raan and Ludo Waltman for expert advice and guidance throughout the study. Robert Tijssen for feedback on an earlier version of the paper.

Conflict of interest/competing interests

The author declares that they have no conflict of interest.

Chapter 5

Geodiversity of research: Geographical topic focus, author location, and collaboration. A case study of SDG 2: Zero Hunger[4]

[4] This chapter is based on:

Purnell, P. J. (2024). Geodiversity of research: geographical topic focus, author location, and collaboration. A case study of SDG 2: zero hunger. Scientometrics, 129(5), 2701–2727. https://doi.org/10.1007/s11192-024-04994-5

Abstract

This bibliometric study examined three aspects of geodiversity of research, namely the geographical topic focus of the study, author location, and international collaboration dynamics. The publication dataset comprised 60,000 papers from the Dimensions database that have been associated with hunger research using Digital Science's machine learning algorithm that enhances expert led search strategies. As the research was related to hunger, papers were mapped on to the Global Hunger Index country categories as convenient classification. Only 41% of hunger-related publications that focus on countries most affected by hunger feature authors affiliated to institutions in those countries. Even fewer of those publications feature locally based authors in first or last position. These numbers gradually reverse as the level of hunger declines. We analyse sample papers in an attempt to understand the reasons for these trends. These included differences in research infrastructure, sub-authorship recognition such as acknowledgements, and limitations of the relationship between country mention and real topical focus. We did not find evidence of widespread differences between senior and overall authorship and consequently urge caution before judging international collaborations as 'helicopter' research based only on author country affiliations and authorship position.

Keywords

Geodiversity – Country focus – Hunger – Helicopter research – Author position – SDG 2

5.1 Introduction

In 2020 Nature, in collaboration with Scientific American, published a collection of articles on diversity in research (Nature, 2020). The aim of this collection was to identify sections of society that are underrepresented in research and investigate the challenges they face. Subsets of society identified in the collection included women, people with disabilities, ethnic minorities, and socially disadvantaged populations. We aim to contribute to the body of research on socially disadvantaged populations, specifically those affected by the problems associated with hunger. Alleviating hunger is a global challenge and has been defined as one of the United Nations sustainable development goals (UN SDGs) (United Nations General Assembly, 2015). We specifically investigate three geographical aspects of diversity: the geographical focus of research, author location, and collaboration patterns of researchers from different places. Rather than delving into the underlying socio-economic issues, this paper is a large-scale bibliometric study that will inform debates around the 'geodiversity of research'.

The geographical focus of research plays an important role in the sustainable development agenda because of the increasing involvement of the scientific community (Brundtland, 1987; Hassan, 2001; IUCN–UNEP–WWF, 1980; Kates et al., 2001). During many of these discussions, it has been established that studies of society's most pressing challenges offer more value when conducted at locations where most impact is felt (Balvanera et al., 2017; Mirtl et al., 2018; Sutterlüty et al., 2018). In this study, we assume that if a country name is mentioned in the title or abstract of a paper, then it signifies some topical focus on that country. Consequently, using bibliometric data we can conduct large scale studies of the geographical focus of research articles by looking at the titles and abstracts of publications. This macro level analysis enables us to build a picture of which countries researchers are focusing on.

The geographical location of researchers is equally interesting to investigate, especially when the topical focus of the study is in developing regions of the world. There is a notable difference between the global, often theoretical perspectives of well-funded academics in countries with established research infrastructures, and the action-based urgency of resource-poor local scientific communities in developing countries (Kates, 2011). There is growing concern about practices of researchers from wealthy nations using samples or data from developing regions (Amugune & Otieno-Omutoko, 2019; Bockarie et al., 2018; van Groenigen & Stoof, 2020). By using author affiliations on published research, we can contribute quantitative analyses to support this discussion.

In many cases, research in resource-poor nations can benefit from collaboration with scientists based in developed countries because they can contribute additional funding and expertise. The dynamics of cooperation between international scientists and academic communities in developing regions is key to the success and acceptance of sensitive projects. There have been concerns that such collaboration dynamics have not always been appropriate, and questions have arisen around whether due consideration has been given to local scientists. There is a growing complaint of 'helicopter' or 'parachute' research where researchers from wealthy countries are said to drop in to poorer regions, collect samples, and leave (Nature, 2022). In response, guidelines that were initially established to ensure individual scientists adhered to ethical collaborative research practices (Yakubu et al., 2018) have been expanded to call for oversight of partnerships by funders, societies, and academic publishers (e.g., Aramesh, 2019; Heinz et al., 2021; Nature, 2022; PLOS, 2021). The 2022 World Conference on Research Integrity hosted a discussion that will lead to a "Cape Town Statement" on equitable research partnerships (Horn et al., 2022). Our bibliometric analysis of international collaboration on hunger-related research is intended to provide data that will contribute to this discussion.

In our study, we use bibliometric data to investigate these three aspects of geodiversity of research, i.e., the geographical focus of the study, location of its authors, and collaboration dynamics. With nearly one in ten people today suffering from chronic hunger (von Grebmer et al., 2020), we chose SDG 2 as the focus of this case study. Our bibliometric data was sourced from the Dimensions database because it has broader coverage than databases such as Scopus or Web of Science (Hook et al., 2018) and indexes a substantial amount of short scientific documents such as meeting abstracts and scientific communications not included in Scopus (Visser et al., 2021). In our view, additional coverage beyond traditional peer-reviewed journals is important in the context of research on local issues in countries with less developed academic publishing systems. Other studies have shown advantages of the Dimensions coverage in the social sciences (van Leeuwen et al., 2022) and infectious diseases (Rahim et al., 2021). Using Dimensions is therefore a deliberate decision aimed at maximising coverage of relevant scholarly content.

Our paper on geodiversity of hunger-related research is intended to form a large-scale bibliometric analysis that can be used to support discussion around the following questions.

Research questions

1. To what extent is hunger research focused on countries affected by hunger?
2. To what extent are authors of hunger research located in countries severely affected by hunger?

3. To what extent are research publication partnerships between international experts and local scholars equal in hunger research?

5.2 Literature review

An important aspect of geodiversity of research is the geographical focus of the topic under study. We have used mentions of country names in research papers as indicators of topical focus on that country. Previous work on country naming suggests there may be an association with socio-cultural patterns related to characteristics of the topic, author location, or whether samples used in the study (Kahalon et al., 2021). Kahalon et al. (2021) found that authors are less likely to mention the country name in an article title if the study includes samples from WEIRD (Western, Educated, Industrialised, Rich, and Democratic) countries. The same authors point out that while naming countries in non-WEIRD countries can initially seem inclusive, the practice could be counterproductive by inadvertently suggesting the findings are geography-specific and not generalizable. This suggestion may reinforce the implicit belief that knowledge produced by scientists in and about individuals located in WEIRD countries represent the universal or default position (Castro Torres & Alburez-Gutierrez, 2022). The implication continues that when authors from non-WEIRD countries declare regional focus of their study, these signify exceptions to the rule thereby reducing the articles' usefulness in global research (Kahalon et al., 2021).

Another aspect of this study is the geographical location of the authors. As global sustainability challenges such as hunger, poverty, and climate change affect populations differently, author location helps us investigate the extent to which local academics publish research related to addressing problems faced by communities close to them. There is a lack of bibliometric studies on author location in hunger research, however studies on author location in related fields (e.g., Reyes-García et al., 2019) suggest that indigenous and local knowledge have the potential to contribute an additional layer of knowledge to research on social-ecological systems. The true benefit of the locally grounded research model is achieved by attracting ideas and people into a specific place where they gain access to local knowledge (Billick & Price, 2010; Gerlak et al., 2018; The British Academy, 2021).

Where local authors are underrepresented, there may be a missed opportunity to employ local knowledge to improve research outcomes. Previous studies on local author representation have lamented the lack of local authors. For instance, authors from tropical countries who have the most knowledge to contribute to conservation science and the most to suffer from its impact (Mammides et al., 2016) are underrepresented in conservation science studies. In medical fields, there is a clear under representation of authors from non-high-income countries (Campbell et al., 2023; Mooldijk et al., 2021). Meanwhile, there may be inherent sampling bias in even the most inclusive global bibliometric databases as demonstrated by the relative lack of linguistic studies on global South languages and by authors based in the global South (Bylund et al., 2023).

The third aspect of geodiversity in this study is the formation of international collaboration as represented by author affiliations and the order in which authors appear on published research papers. International collaboration including capacity building and joint research projects can play a role in the response to localised social challenges and even humanitarian tragedies (Bajoria, 2011). The 2013-2016 Ebola virus disease outbreak in West Africa generated international attention and stimulated collaboration that involved collaborative clinical studies, training of local outbreak responders, and the establishment of diagnostic and surveillance laboratories (Arias et al., 2016; Heymann et al., 2016; Yozwiak et al., 2016). Attempts to improve capacity of West African communities to respond to disease outbreaks have been termed 'rooted' collaboration (Yozwiak et al., 2016) because of the co-creation of sustainable local capacity building. Lamentably, not all the international collaborators contributed in the same spirit and some were described as sub-optimal 'parachute' research (Heymann et al., 2016).

Helicopter or parachute research (North et al., 2020) portrays the image of privileged academics who come to troubled regions from wealthy countries and avoid real collaboration with local scientists. These researchers are said to drop in, collect samples, and leave, sometimes without the knowledge or permission of authorities in the visited country (Heymann et al., 2016). Others have used the term 'colonial research' or 'neo-colonial research' (Minasny & Fiantis, 2018) to suggest that scholars from developed economies feel entitled to take samples from less well-developed areas for their own purposes and limit the input of local colleagues.

Where local scientists are excluded from collaboration, or their role is limited to locating and collecting samples, international research teams could be failing to use the chance to increase the capacity of scholars at the collection site, and instead offer little benefit to the local community (Minasny & Fiantis, 2018). Following the particularly devastating Indonesian peat fires of 2015 that destroyed vast areas of peatland, local scientists set out practical solutions as a priority, and encouraged research into responsible and effective peatland management (Minasny & Fiantis, 2018; Sabiham et al., 2018). These authors lamented international teams in Indonesia concentrating efforts on more academic research into deforestation, greenhouse gas emissions, and peat fires with less practical value.

As research becomes more collaborative, the number of authors per paper has grown more than five-fold in the last 100 years (Aboukhalil, 2014). That makes the relative contribution of authors less obvious, and readers have turned to proxies such as author position to infer leadership of research projects. In academic literature, first and last author positions are considered 'key' contributors (Mattsson et al., 2011; Wren et al., 2007). The corresponding author also holds considerable weight and has been shown to coincide most frequently with first and then last author position (Mattsson et al., 2011).

Africa-based authors are traditionally underrepresented in the scientific literature. An analysis of author affiliation position on 1,182 biomedical studies conducted in Africa, showed over 93% featured at least one Africa-based co-author (Mbaye et al., 2019). However, Africa-based co-authors featured in fewer than half the articles in first author position, and even fewer in the prestigious last author position. A similar study concurred (Hedt-Gauthier et al., 2019) and called on the research community from high income countries to challenge the established power balance. Despite increased inclusion of local scholars in Africa-based research, inequity remains regarding their relative roles in the team.

The scholarly community has begun to call for action to discourage helicopter or parachute research and instead promote ethical and sustainable collaboration with academics from low- and middle-income countries. A group of editors and researchers published a consensus statement to promote equitable authorship which includes practical advice to journal editors on evaluating manuscript submissions resulting from international collaborations (Morton et al., 2022).

A related discussion has been published in Geoderma, about the prevalence of helicopter research in soil science, along with the limited involvement of local knowledge owners, and the lack of structural improvement in the community at the place of the study (van Groenigen & Stoof, 2020). One participant from Ethiopia suggested building local capacity in young scientists though involvement of students from master's and PhD programmes who should be first authors where they conducted the work (Haile, 2020). As part of the same discussion, Giller (2020) described how the Dutch research funding agency, NWO-WOTRO Science for Global Development made its grants subject to a compulsory workshop with stakeholders from the country of the proposed study. The workshops often led to closer collaboration between the local and visiting scientists (Giller, 2020). Other journals have taken similar steps to tackle helicopter research and ethics dumping (Morton et al., 2022; Nature, 2022).

Hunger is a complex problem that is interlinked with political or military instability, e.g. South Sudan (Mayai, 2020), Yemen (De Souza, 2017). In regions under civil and military conflict, severe malnutrition contributes to mortality (Leaning & Guha-Sapir, 2013; Salama et al., 2004). Under such circumstances, the local education and research infrastructure is often limited (Lai & Thyne, 2007) and much of the work is conducted by researchers in other countries or foreign non-governmental Organizations (Ford et al., 2009; Kalleberg, 2009).

Following high profile events such as famine or war, it is natural to want to help. However, in some cases, Western researchers have arrived in affected regions with pre-conceived research protocols thereby limiting the role of local participants and scholars (Asiamah et al., 2021). A more progressive approach involves local scholars and participants in an inductive and iterative way (Firchow & Gellman, 2021; Yom, 2014). Studies of the dynamics between the researcher and the researched have demonstrated benefits in an inclusive approach where those with local knowledge are included at every stage of the research design (Riley et al., 2003) including the very question being addressed. That way, international research teams ought to engage local participants as actors with agency (Gellman, 2021) and avoid limiting their roles.

5.3 Data and methods

We extracted publications from a version of the Dimensions database hosted by the Centre for Science and Technology Studies (CWTS) at Leiden University. Dimensions uses a multistep process to tag publications considered relevant to sustainable development goals that utilises a machine learning algorithm to enhance expert driven search strategies (Wastl et al., 2020, 2021). We selected all records linked to SDG 2: Zero Hunger. The time window used was six full publication years (2016 – 2021). We included journal articles, conference papers, books, monographs, and book chapters, but excluded preprints because they have not been peer-reviewed. We also excluded papers that did not list any author affiliations because the affiliations formed an important part of our analysis.

For country population, we used the most recent UN estimates (The World Bank, 2022).

For ease of comparison, we present the results in groups of countries, rather than individual nations. As the case study is on hunger related research, we classified papers by country according to the categories described in the Global Hunger Index (GHI) 2021 report (von Grebmer et al., 2021) and shown in Figure 5.1 with full country listing in the supplementary information table S5.1.

We counted the number of publications in which a country was mentioned in either the title or abstract of the article. These were termed 'country mentions'. If one or more countries within one of the GHI categories were mentioned in a publication, then it counted as one mention for the category. If one or more countries were mentioned in the publication for two categories, then it counted as one mention for each category.

To determine the geographical location of researchers, we examined the countries in the author affiliations of each paper. These were termed 'country affiliations'. Similar to the mentions, if one or more country affiliation in any one category appeared in a publication, then it counted as one country affiliation for that category. Meanwhile, if one or more country affiliations were found from two GHI categories, it counted as one country affiliation for each category.

Figure 5.1 GHI country categories based on the severity of hunger

In order to create these categories, the GHI ranked countries using a composite score based on four indicators (Table 5.1); undernourishment, child wasting, child stunting, and child mortality (Wiesmann et al., 2015). There was sufficient data to calculate individual scores for 116 countries. An additional 12 countries were provisionally designated and for seven countries there was insufficient data for even a provisional categorisation. We assigned countries to categories if they were assigned or provisionally assigned by the GHI 2021 report but did not include the remaining seven countries in the study.

Table 5.1 Global Hunger Index composition

Dimension	Indicator	Weighting	Source
Inadequate food supply	Calorie deficiency Proportion of the population that is undernourished	1/3	UN Food and Agriculture Organisation (FAO)
Child undernutrition	Proportion of children <5 yrs. suffering from stunting	1/6	UN United Nations Children's Fund (UNICEF) World Health Organisation (WHO) World Bank
	Proportion of children <5 yrs. suffering from wasting	1/6	UN United Nations Children's Fund (UNICEF) World Health Organisation (WHO) World Bank
Child mortality	Child <5 yrs. mortality rate	1/3	UN Inter-Agency Group for Child Mortality Estimation (UN IGME)

Source: Global Hunger Index 2021 (von Grebmer et al., 2021)

The International Food Policy Research Institute (IFPRI) was founded in 1975 and organises research projects in the areas of food supply, nutrition, food trade systems, agricultural economies, and governance. It also runs country level research programmes because of the different challenges faced by each country (International Food Policy Research Institute, 2021) and the different opportunities to address them. The information gathered at country level can help us identify where the problem is most acute.

In order to address our first research question about the extent to which research on hunger focuses on countries afflicted by hunger, we calculated the aggregate number of country mentions for each category. To address our second research question about the country affiliation of hunger research scholars, we calculated the aggregate number of papers in each GHI country category. We presented the country mentions and the country affiliations as absolute numbers and normalised for combined population of the country category.

To answer the first two research questions more fully, for each of the GHI country categories and for the non-assessed countries, we combined the country mentions with the author affiliations in a Venn diagram. This enabled a visualisation of the relationship between local authorship and the severity of hunger. For the mentions-only papers and the affiliations-only papers, we performed bibliometric analyses and we conducted manual examination of random samples of the non-overlapping papers.

The mentions-only papers mention a country but do not feature any authors in the country mentioned. In order to discover where the authors are based, we quantified the number of country mentions for the country with most publications in the SDG 2 dataset from each GHI category.

We then manually examined the full text of 100 mentions-only papers (20 selected at random from each GHI category) to gain insight into the possible reasons that no locally affiliated authors were listed on the paper.

The affiliations-only papers do not mention the countries that the authors are affiliated to. We aim to find out whether these papers mention any country at all by comparing the affiliations-only papers with the SDG 2 papers that include no country mentions.

We then manually examined the full text of 100 affiliations-only papers (20 selected at random from each GHI category) to look into reasons why authors have not mentioned the country in which they are affiliated.

For the sample analyses, we grouped countries by GHI categories and assigned each paper to the most severe category it could belong. For example, a publication that mentioned a country from the low hunger category and another country from the serious category was only counted in the serious category. This way, we avoided papers being counted in multiple categories.

To address our third question about the equality of research collaborations, we determined the author position for each paper and aggregated them for each GHI country category. We then presented the share of first author country affiliations, last author affiliations, and overall country affiliations from each of the country categories.

5.4 Results

The Dimensions database contains 59,778 papers published between 2016 and 2021 that are related to SDG 2 – Zero hunger. In one third of these papers, the authors mention at least one country in the title or abstract. Within the same Dimensions SDG 2 dataset, 31,769 (53%) publications featured author affiliations in countries listed in the categories of the Global Hunger Index (GHI) 2021 report. The number of countries with their combined populations and number of country mentions and affiliations are shown in Table 5.2. There is only one country in the extremely alarming category (Somalia), and we have therefore combined the two most severe categories as 'alarming / extremely alarming'. We also show the number and combined population of countries not assessed for the GHI report. The 'not assessed' category includes many of the world's wealthy and economically developed nations. For a full list of country categorisation, see supplementary information table S5.1.

Table 5.2 GHI country categories by population, country mentions and affiliations

GHI Hunger category	# Countries	Combined population	Mentions	Mentions per pop (m)	Affiliations	Affiliations per pop (m)
Alarming / extremely alarming	10	232	434	1.87	288	1.24
Serious	37	2610	7281	2.79	12873	4.93
Moderate	31	1214	4475	3.69	7870	6.48
Low	50	2624	4776	1.82	13973	5.33
Not assessed	105	1235	4938	4.00	35834	29.02

5.4.1 Geographical focus (country mentions)

Our first research question was 'To what extent is hunger research focused on countries affected by hunger?'. We found that 26% of the SDG 2 papers in our study mentioned countries listed in the GHI 2021 report (table 5.2). Within the GHI country categories, we see an interesting pattern. First, the countries in the low category have the fewest mentions per population. The moderate countries have more than twice the number of mentions per population as the low countries. However, as the severity of hunger increases from moderate, to serious, and then to alarming / extremely alarming, the number of mentions per population declines again. The frequency of mentions per population is higher for countries not assessed by the GHI (4.00) than for any of the categories that are assessed.

5.4.2 Geographical location (author affiliations)

Our second question was 'To what extent are authors of hunger research located in countries severely affected by hunger?' We found that more than half the papers in our dataset featured at least one author whose affiliation was in one of the countries assessed in the GHI report (Table 5.2). The GHI categories with the highest number of affiliations were the low and serious categories followed by the moderate category. The alarming / extremely alarming categories featured by far the lowest (288) number of author affiliations. Due to the variation in population between the countries within the GHI categories, the population normalised country affiliations are presented on a world map in Figure 5.2.

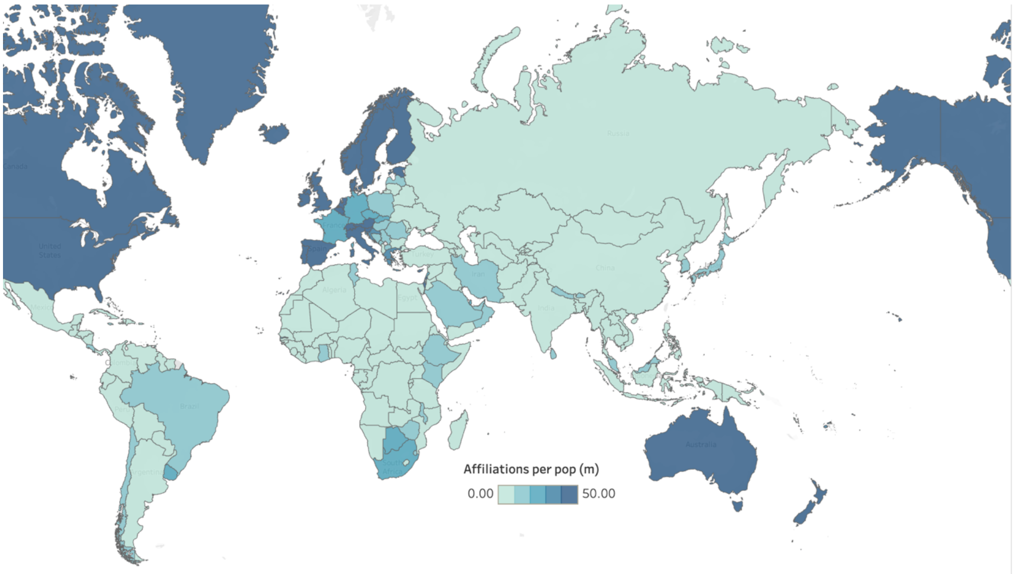


Figure 5.2 Author affiliations per million population by country

We can see at a glance that the higher number of author affiliations per population are associated with wealthy countries, while countries in developing regions such as Sub-Saharan Africa tend to have fewer author affiliations per population. Similar to the geographical focus, we found a downward trend in population-normalised geographical author affiliations as the severity of hunger increased. Within the GHI categories, the affiliations per population peaks in the moderate category and then declines markedly as the severity of hunger increases. Indeed, researchers from countries most afflicted by hunger are twenty times less likely to publish research papers on hunger than those in wealthy, non-assessed countries.

5.4.3 Relationship between regional focus and regional authorship

The relationship between country mentions and country affiliations is shown by GHI category in Figure 5.3.

Figure 5.3 Mentions and affiliations by GHI categories

For each GHI category, the circle on the left represents the number of papers which mention at least one country in that category. If a paper mentions the same country more than once or more than one country in the same category, it still only counts as one paper for that category. However, it the paper mentions one or more papers in two categories, then it counts as one paper for each category. The circle on the right follows the same rules but refers to author's country affiliation.

In the top-left diagram, the overlapping section shows that 178 of the papers that mentioned countries in the alarming / extremely alarming countries also featured at least one author affiliated to one of the countries in the same category, i.e. papers that featured local authors. Meanwhile, 256 of the mentions did not feature locally based authors and indeed did not feature any author affiliated with any country in the alarming / extremely alarming category. The 110 affiliations-only papers on the right section of the diagram tells us that 110 papers that featured authors from countries in this category did not mention any country in the same category. Figure 5.3 illustrates that the share of mentions-only papers grows, and the share of affiliations-only papers declines as the severity of hunger increases.

In Figure 5.4, we concentrate on the share of overlapping papers, i.e. those that mention countries from a GHI category and also feature at least one author from the same category.

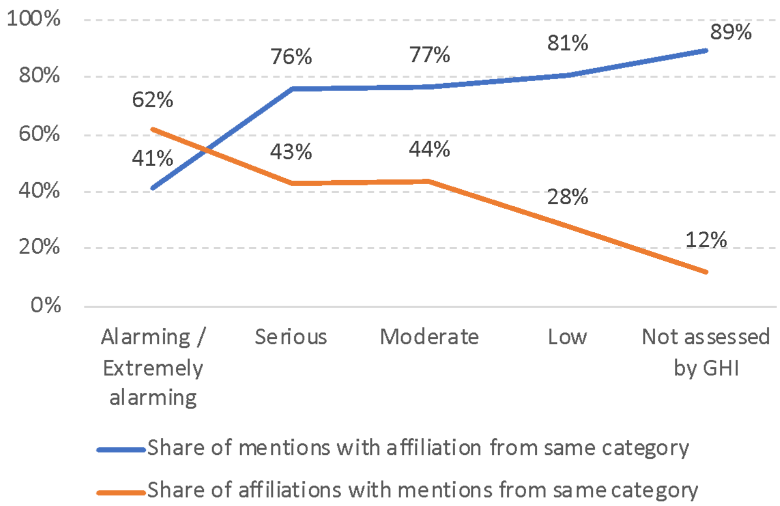


Figure 5.4 Share of mentions and affiliations by country category

The blue line shows the overlapping papers as a share of all the mentions of countries in a particular category. For instance, 41% of the papers that mention alarming or extremely alarming countries also feature at least one author affiliation from these categories. This is the lowest share and as the hunger severity decreases, the share of country mentions that also feature authors from the same category increases. The orange line shows the number of overlapping papers as a share of author affiliations in a particular GHI category. Of the papers featuring author affiliations from countries with the most severe hunger problems, 62% also mention countries from the same category. As the severity of hunger decreases, this share decreases until only 12% of the papers with authors from countries in the not assessed category mention countries in the same category.

5.5 Location of non-local authors (mentions-only papers)

The mentions-only papers do not feature authors affiliated to the countries mentioned in the article. Now we know where the authors are not located, we would like to know where they are located. In Table 5.3 we show the most frequently occurring country affiliations on the mentions-only papers for the country with the most mentions in each GHI category. In the alarming / extremely alarming categories, the Democratic Republic of the Congo (DRC) is the most frequently mentioned country. Authors that mention the DRC but are not based there are most frequently located in the USA, UK, and India.

Table 5.3 Location of authors on mentions-only papers

DRC mentions (Alarming)	Paper s	% Share	India (Serious)	Paper s	% Share	Indonesia (Moderate)	Paper s	% Share
United States	25	24%	United States	195	8%	United States	23	3%
United Kingdom	7	7%	United Kingdom	74	3%	India	15	2%
India	3	3%	Australia	41	2%	Australia	13	2%
South Africa	3	3%	China	31	1%	Japan	12	2%
Australia	3	3%	Germany	30	1%	United Kingdom	11	1%
Belgium	3	3%	Canada	29	1%	Thailand	7	1%
China	3	3%	Switzerland	17	1%	Germany	6	1%
Canada	2	2%	South Africa	14	1%	South Africa	6	1%
Denmark	2	2%	Bangladesh	14	1%	Switzerland	6	1%
France	2	2%	Japan	13	1%	France	5	1%
China (Low)	**Paper s**	**% Share**	**USA (Not assessed)**	**Paper s**	**% Share**			
United States	83	5%	India	48	3%			
India	62	4%	United Kingdom	44	2%			
United Kingdom	40	2%	China	41	2%			
Australia	30	2%	Australia	38	2%			
Germany	23	1%	Canada	30	2%			
Netherlands	20	1%	Italy	16	1%			
Japan	16	1%	Iran	14	1%			
Canada	14	1%	Germany	13	1%			

France	12	1%	Russia	12	1%			
Italy	12	1%	Brazil	11	1%			

In all groups, many of the mentions-only papers were published by authors affiliated to countries with large populations and high publication output such as the United States, United Kingdom, China, and India. There were however, some prominent appearances from regional neighbours. For instance, Bangladesh was listed among the leading author locations on papers that focused on India, and authors based in Thailand frequently published papers that focused on Indonesia.

Other relationships between countries may also have contributed to authors in one country focusing on another country that might not be a direct neighbour. For instance, authors based in South Africa and Belgium published papers that focused on the DRC that featured no locally based authors.

We also conducted a manual analysis of 100 randomly sampled mentions-only papers (20 for each of the GHI categories) to assess the reason for there being no locally based author affiliation. We examined the full text of each paper in the sample and present the results in the supplementary information table S5.2. We found two main reasons for the lack of a local author affiliation. First, while authors may not be in the same country as that mentioned in the title or abstract, it doesn't mean they are very remote. We found several papers that featured authors in neighbouring countries to those mentioned. In one case (10.1186/s12889-020-08657-x), an international group sought to improve the criteria for community-based treatment of malnutrition in South Sudan. Some of the authors were based in Kenya, just across the border with South Sudan.

In other mentions-only papers, the studies were not really focused on the countries that were mentioned in the title or abstract of the paper. This was caused by authors making passing mention of a country. For instance, in the abstract of a study of a rice production system used throughout India (DOI: 10.1007/978-981-10-3692-7), the authors mention that the method originated in Madagascar, and although Madagascar is not mentioned again, the paper has been classified as focusing on Madagascar.

5.5.1 Lack of country mentions (affiliations-only papers)

The affiliations-only papers make no mention the countries in which the authors are affiliated in their title or abstract. We wanted to know whether this is because the paper lacked geographical focus entirely or if the focus was on another geographical location. In Table 5.4 we show the number of affiliations-only papers for the country with the highest number of hunger-related research papers for each of the GHI country categories. We also show the number and share of affiliations-only papers that did not mention any country at all.

Table 5.4 Papers with no country mention

Country name	GHI category	Affiliations-only papers	Papers with no country mentioned	Share of papers that mention no country
Democratic Republic of the Congo	Alarming / Extremely alarming	37	24	65%
India	Serious	5119	4915	96%
Indonesia	Moderate	995	952	96%
China	Low	3936	3615	92%
United States	Not assessed	13058	9771	75%

In all GHI categories, a high share of affiliation-only papers did not mention any country name at all. The lowest share was in the Democratic Republic of the Congo whose authors mentioned no country in 65% of the affiliations-only papers. We also conducted an analysis of 100 randomly sampled affiliations-only papers (20 for each of the GHI categories). We examined the full text of each paper in the sample to determine the reason for the lack of country mention and the tabular results are available in the supplementary information table S5.3.

One common reason for not mentioning the country where the authors are based is that the article has no regional focus. For instance, we found several papers with author affiliations in the serious category were published by authors in India and focused on methods for improving agricultural methods and had no regional focus.

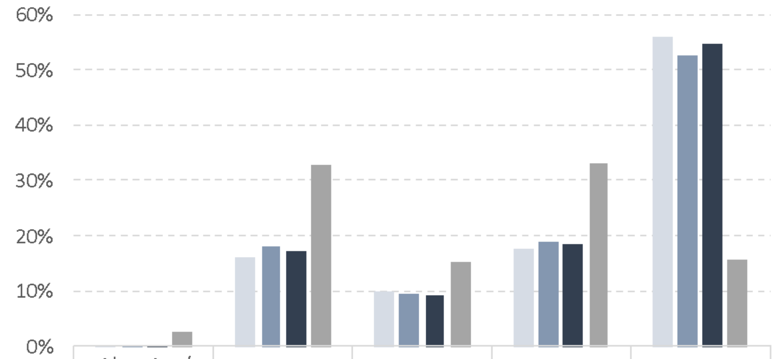
A second group of affiliations-only papers were indeed studies that focused on a specific country, but the authors had not mentioned the country at all in the title or abstract of the paper. One example (10.1155/2019/4740825) is a clinical study with all the authors based in the Democratic Republic of the Congo (DRC) and thanking a local hospital in the acknowledgements. The study is clearly focused on the DRC but the authors have not mentioned the country in the title or abstract of the paper.

In a third set of affiliations-only papers, the authors do mention the local site of the study but not necessarily the country name. We found evidence of authors using city names e.g., Royapettah (10.18203/2349-3291.ijcp20174153), islands e.g., Zanzibar (10.1093/cid/cix500), or simply referring to 'our nation' (10.1007/978-981-16-6124-2_1).

5.5.2 Partnerships

To address our third research question about the equality of international partnerships, we analysed the author position in more detail. For 55,422 (93%) of the SDG 2 papers, we could identify the country affiliation of the first and last authors, often associated with lead author positions. We found that the last author has an affiliation in a country listed in the GHI 2021 categories on 40% of the SDG 2 papers.

Considering all SDG 2 papers, only 0.35% of author affiliations were from countries in the alarming or extremely alarming GHI country categories (Figure 5.5). The share was even lower for authors in first (0.23%) and last author (0.17%) positions. Meanwhile, the combined population of the countries in the alarming and extremely alarming countries comprises 2.9% of the world's population. Authors from these countries are therefore underrepresented in scholarly research on hunger and especially underrepresented in lead author positions.



	Alarming/ Extremely alarming	Serious	Moderate	Low	Not assessed
Author affiliations	0.35%	16.3%	10.0%	17.6%	55.8%
First author	0.23%	18.1%	9.6%	19.1%	52.9%
Last author	0.17%	17.4%	9.1%	18.5%	54.8%
World population	2.93%	33.0%	15.3%	33.2%	15.6%

Figure 5.5 Author position - All SDG 2 papers

Authors from countries in the serious, moderate, and low GHI categories are also underrepresented when compared with their share of the world's population albeit to a lesser extent. The share of lead author positions is not very different from overall authorship. In those countries not assessed by the GHI, which includes relatively wealthy countries, the pattern is reversed with authors overrepresented compared with their combined share of the world's population.

5.6 Discussion

We used country mentions and author country affiliations to study the geographical diversity of researchers, using scholarly papers related to SDG 2: Zero hunger as a case study. For ease of comparison, we used the country categories as described in the 2021 Global Hunger Index (GHI) report which are based on the severity of hunger.

About one third of hunger research papers mention at least one country in their title or abstract, although this trend diminished when mentioning the country categories most severely affected by hunger. This was a counterintuitive result, and we consider possible explanations. First, there are only ten countries in the alarming and extremely alarming categories compared with 118 in the less severe categories and 106 not assessed countries. That means countries in less severe categories comprise more country names that are available to be mentioned. Similarly, the smaller relative population in the alarming categories means there are fewer local authors who can focus on their own country than in the less affected categories.

We cannot be sure of authors' motives when they decide to name a country in the title or abstract of their paper. There is evidence that authors from wealthy, developed countries with established research infrastructure, especially the United States, are less likely to mention the country names even in locally based studies (Castro Torres & Alburez-Gutierrez, 2022; Kahalon et al., 2021). Consequently, there are potentially studies with a country focus that do not mention the country of focus, and that those papers missed are predominantly in the lower GHI country categories. That means the real tendency of decreasing topical focus in countries less affected by hunger may not be as sharp as our results show.

We used authors' affiliations to determine their geographical location and defined local authors as those whose country affiliation matched the name of the country mentioned in the paper. We found the presence of local authors declined as the severity of hunger in the mentioned country increased, especially on papers that mentioned countries in the alarming and extremely alarming categories which are the most severe. We suggest potential reasons for the lack of local authors on research papers that focus on the most severely affected countries.

Countries most severely affected by hunger are also among the world's poorest countries and poverty can be a barrier to young people entering higher education. The impact of long-term poverty, especially in countries experiencing war or civil conflict, may have limited the development of a research infrastructure in some countries categorised as suffering from alarming levels of hunger. Less developed research infrastructure is therefore proposed a possible explanation for the lack of local academics and associated research publications in the most severe GHI categories. Meanwhile, academic publishing has flourished in some countries with large populations in the less severe hunger categories such as India (serious) and Indonesia (moderate). The prolific scholarly output in these countries has contributed to relatively large numbers of country mentions in hunger research.

That researchers from countries not assessed by the GHI publish the majority of hunger-related research papers is not surprising. The GHI doesn't assess developed, wealthy countries and these countries traditionally publish the most research in general. In 2015, the UN called for the whole world, not just those living in affected areas, to respond to a series of challenges. The contribution of academics in developed nations could be interpreted as an encouraging sign that the global research community is engaged in addressing the grand challenges of our time. It may also be symptomatic of a sampling bias caused by overrepresentation of country affiliations from high income countries (Bylund et al., 2023).

In all the GHI categories, our results showed in most cases that authors either mention the name of their own country or they do not mention any country at all. Papers that do not mention any country may have no regional focus, or the authors have decided not to mention the name of the country in which the study was conducted.

The appearance of local contributors mentioned in the acknowledgements section of the article but who did not appear as co-authors might also partially explain the absence of local authors. In some papers, collaborators from the country of focus were acknowledged and thanked in the publication even if they did not feature as authors of the study. Acknowledgements form part of the reward triangle bestowed by researchers on those who have helped or significantly influenced academic publications (Costas & van Leeuwen, 2012; Cronin & Weaver, 1995). The difference between a contribution that deserves 'only' acknowledgement, and one worthy of co-authorship is a key distinction. The observation of acknowledgements in our analysis of sample papers warrants further examination to see whether this phenomenon has any relation with the GHI country categories.

Author position and partnership dynamics

Author position in the context of partnerships between visiting and local researchers has come under scrutiny by the academic publishing community (e.g., Morton et al., 2022; Nature, 2022; PLOS, 2021; van Groenigen & Stoof, 2020). This study contributes to the discussion by demonstrating the relationship between author position and affiliation from both geographical focus and geographical location perspectives.

We specifically identified the country affiliation for authors who appeared in first or last author position. First, and especially last author position have been used as a proxy for lead authorship in several studies. We found that authors from countries in the most severe GHI categories were underrepresented in lead author positions. However, in all the other categories the share of lead author positions was close to the overall share of authors. In the serious and low GHI categories, local authors were slightly overrepresented compared with all author positions. However, where countries from GHI categories were the focus of research, authors from the same category featured far more frequently and were even overrepresented in lead author positions. This result is suggestive of greater participation of local academics in studies with a regional focus.

This part of our study builds on earlier work conducted in health fields that reported underrepresentation of local researchers in lead author positions (Hedt-Gauthier et al., 2019; Mbaye et al., 2019). Our results support the idea that unequal partnerships may exist in international collaborative research, but only in the countries most severely affected by hunger. In all other categories, our author position analysis showed that the share of local researchers in lead author positions was similar to the overall share of local authors. We therefore encourage further discussion on the partnership dynamics between local academics and international collaborating partners.

These findings could be interpreted as showing that international academics from wealthy nations more frequently occupy leadership roles in hunger research studies on the most severely affected regions than local authors. The underrepresentation of local authors presented in this paper may also support the idea that local contributors might play a limited role in international collaborative research in their own country such as data collection but not authorship (Asiamah et al., 2021).

Study limitations

This study was the first to examine the geodiversity of research using SDG 2 as a case study. However, we acknowledge the study has a number of limitations that could be used to identify areas for follow on studies.

First, the GHI country categories are based on a composite indicator. The cut-off scores that assign countries to one category or another are necessarily arbitrary but mean that there might be greater differences between countries at the extremities of a category than between countries separated by the cut-off score. We also hypothesise that conditions may vary within a country, such that different regions of a country would meet the conditions for different GHI country categories. We have presented evidence of community-based collaboration across international borders between countries that have landed in different GHI categories. The GHI categories are therefore not infallible in their indication of severity of hunger. We could have used nation states grouped by geographical or political regions instead, but those groupings would of course be subject to the same limitations as the GHI categories.

The definition of the body of research papers on hunger research is not unequivocal. We used the Dimensions database because of its inclusive coverage (Hook et al., 2018; Huang et al., 2020; Visser et al., 2021) and because Dimensions has developed a method that uses an AI algorithm to tag papers related to the different SDGs. However, competing bibliometric databases have used slightly different approaches to identifying SDG research which make huge differences to the papers retrieved. Comparisons have shown very limited overlap between databases in specific SDG searches (Armitage et al., 2020a; Purnell, 2022b). Using alternative publication data sources for SDG related research studies could have therefore produced different results in our study. Indeed, the varying search techniques and use of AI algorithms between publication databases is a limitation now applicable to every bibliometric study that uses SDG related research.

While Dimensions coverage is broad, it is not comprehensive and relies mainly on Crossref and PubMed as it sources. For publications to be in Crossref, they need to register a digital object identifier (DOI). This requires some expertise and a financial arrangement where publishers pay $1 US for each paper assigned a DOI. While these arrangements may be easily achievable in some regions, in other less economically developed areas of the world they might pose a barrier to publishing. In these cases, papers are more likely to appear in local university presses, not be assigned a DOI, and therefore not be indexed by Dimensions. Unfortunately, poorer countries appear in the most severe GHI categories and are at the highest risk of lower database coverage (Giménez-Toledo et al., 2017). As a consequence, there might be proportionately fewer research papers on hunger published by authors affiliated to the most affected countries in our study. Crossref has recently announced an initiative to address this obstacle (Collins, 2022).

We have used mention of a country name in the title or abstract of an article as an indicator that the study focuses on that country to a certain extent. However, our manual examination of sample papers uncovered evidence of countries that were only mentioned in passing, and that did not represent the geographical focus of the study.

This finding shows that a single appearance of a country name should not alone be accepted as a reliable indicator of geographical focus. Future studies could look for ways to improve knowledge surrounding the use of country mentions as indicators of focus such as interpreting the country name in context. The passing mentions in our study only mention a country once and consequently multiple mentions of a country name in an article might strengthen the indication of focus.

Conversely, we have assumed that the absence of country name in the article title or abstract means the study was either focused somewhere else or had no regional focus at all. This assumption might not always hold true. Scholarly papers by authors in wealthy countries with long established research infrastructure often do not mention the name of the country even in locally focused studies (Castro Torres & Alburez-Gutierrez, 2022; Kahalon et al., 2021). Some authors refer to regions of the world, e.g., Sub-Saharan Africa rather than countries, or to smaller units of countries like cities, islands, or regions. These examples serve as further evidence to support research into the use of country mentions as indicators of geographical focus.

5.7 Conclusion

This is the first large-scale bibliometric study on the geodiversity of research that used country mentions as an indicator of topic focus and author affiliations to identify geographical location. In light of the urgent and growing problem presented by hunger, we used the body of research papers related to SDG 2: Zero Hunger in the Dimensions database as a case study, and the Global Hunger Index (GHI) country categories for ease of comparison.

We found that hunger research papers focused less on the countries in the most severe GHI country categories (extremely alarming or alarming) although that may be partially explained by the comparatively low aggregate population of those countries. On the majority of papers that mentioned countries in the alarming and extremely alarming GHI country categories, the were no authors from the country mentioned. Instead, the majority of authors on the mentions-only papers were from relatively populous and wealthy countries. However, we conducted a manual examination of random samples of the mentions-only papers and found that some authors were based in neighbouring countries.

There were also methodological reasons for the low rate of mentions of countries most severely affected by hunger. There are fewer countries in the severe GHI countries to mention and the aggregate population is relatively small. Researchers based in wealthy countries are also less likely to mention the geographical focus of their study than authors in the severe GHI country categories. Use of country mention as an indicator of geographical focus might therefore not be uniform across the countries in this paper and future studies could investigate further.

Our study showed declining participation of local authors affiliated to institutions in the most severe GHI country categories. Author affiliations per population in the extremely alarming and alarming categories was one-twentieth of that in the wealthy countries not assessed by the GHI. Authors based in the most severe GHI country categories mentioned their country in most cases but not in the less severe country categories. We found in all country categories, that most affiliations-only papers did not contain mention of any country at all. Either the authors chose not to mention the country of the study, or the study had no regional focus.

Our examination of a random sample of affiliations-only papers revealed that in some studies with a localised geographical focus, the authors simply omitted to mention the site of the study. Others mentioned the city, island, or region name but not the country name. Again, this shows the shortcomings of using country names as a reliable method of capturing the whereabouts of all authors. To improve recall, the method could be extended to cover additional geographical terms. Future studies would then more accurately determine the share of studies that have no regional focus.

We observe that many of the countries in the most alarming categories have also faced civil conflict, famine, and other causes of long-term instability that may limit the development of research infrastructure and consequent publication output.

The underrepresentation of academics from countries most affected by hunger in last author position is of concern in the context of equitable and fair international research collaboration. Our manual analyses of sample papers in the extremely alarming and alarming categories showed that sometimes local contributors are mentioned in the acknowledgements section of the paper, rather than being included as co-authors. We encourage further analysis on the criteria that warrant co-authorship and whether it is justly applied across all countries. The trend was not repeated in the other categories and our findings therefore did not support the notion of widespread unequal publication practice. Indeed, the examples in our manual analyses did not seem to resonate with the 'helicopter' or 'parachute' research practices described in recent literature (e.g., Heymann et al., 2016; Minasny & Fiantis, 2018; North et al., 2020).

Manual examination of the papers revealed that other factors were at play. For instance, the assumption that non-local authors are based in safe, wealthy countries is often wrong. Many are in fact located just across an international border from a country facing sever hunger problems, or they are based in a country in an even more severe hunger category than the country mentioned. We also found that some papers that mentioned a country didn't really focus on it and the lack of authors from that country could not therefore be interpreted as evidence of questionable research practice.

In our view, reports of helicopter research that call into question researchers' motives require clear definitions and methods that include a level of qualitative assessment including at the very least, manual examination of the publication. While some high-profile examples of questionable research ethics have been published, we urge caution when extrapolating superficial metrics such as author affiliation or author position to ensure they don't lead to unfounded conclusions (Nature, 2022).

Acknowledgements

Ton van Raan and Ludo Waltman for expert advice and guidance throughout the study. Marja Spierenburg, Rodrigo Costas, Ismael Rafols, and the two anonymous reviewers for feedback on earlier versions of the paper.

Data availability

The sample records manually examined for this study and summarized in the supplementary information tables S5.2 and S5.3 are available in Zenodo (Purnell, 2023).

Conflict of interest/competing interests

The author declares that they have no conflict of interest.

Chapter 6

Conference proceedings publications in bibliographic databases: A case study of countries in Southeast Asia[5]

[5] This chapter is based on:

Purnell, P. J. (2021). Conference proceedings publications in bibliographic databases: a case study of countries in Southeast Asia. Scientometrics, 126(1), 355–387. https://doi.org/10.1007/s11192-020-03773-2

Abstract

Leading citation databases have made concerted efforts to reflect academic conference contributions in the form of proceedings papers in their databases. We studied global trends and a regional case study to determine the relative representation of conference proceedings in the global scholarly literature using the Web of Science, Scopus, and Dimensions. We designed our case study of ten Southeast Asian countries to uncover conspicuous publishing patterns obscured by global average figures. As a result, we discovered that Indonesia alone has made a recent and remarkable shift towards conference proceedings publishing. This trend was not the result of expanding database coverage but may be linked to a rapid increase in conferences locally hosted in Indonesia. The conclusion suggests that conference proceedings are increasingly indexed by major databases, and that scholars might have found advantage in publishing conference papers that were quicker and easier to publish than journal articles or book chapters. Our study is relevant to policy makers in the area of research evaluation because it highlights potential changes in academic publishing behaviour by those being assessed.

Keywords

Conference proceedings – Indonesia – ASEAN – location – promotion

6.1 Introduction

Academic scholars all over the world face mounting pressure to publish their findings in sources that are indexed in global citation databases (Abramo & D'Angelo, 2014). Those who evaluate individuals for recruitment, academic promotion, or research grant awards increasingly use some form of quantitative assessment of their publication output and citation impact when taking decisions. The methods used in such assessment are often the subject of extensive discussion (Hicks et al., 2015) and frequently use established citation indexes such as Web of Science (WoS) and Scopus (Archambault et al., 2009) and more recently, new databases such as Dimensions (Hook et al., 2018). All these databases comprise research contributions published in a variety of academic sources, most importantly journals, books and conference proceedings.

The subject of this study is the body of literature resulting from scholarly conferences and which play an important and increasingly visible role in the scholarly literature alongside journal articles, book chapters and other forms of scientific communication. In recent years, conference proceedings have been integrated with citation indexes to enable academics to search, retrieve and cite them in the same way they cite journal articles. The contribution of conference papers to the major citation indexes has not yet been well documented and this is the first topic addressed in this study.

Global trends hide lots of interesting dynamics at the level of individual countries and as the second topic addressed by our paper, we include a case study comprising the ten countries of the Southeast Asian Nations (ASEAN). The ASEAN countries provide a good case study for this topic because they hold a shared vision of identity along with similarities in geography and scientific goals.

Each country displayed a different pattern in conference proceedings output which were grouped into four clusters, and the discussion presents some possible reasons for the differences. Such reasons include database coverage, the influence of national policy on the behaviour of scholars, and the location of the conference.

In this article, we review the literature on the developing role of conference proceedings in scholarly communication, their incorporation into the major citation databases, and introduce the ASEAN region as an ideal case study. In section 6.3, we describe the databases used, the Web of Science, Scopus, and Dimensions and how we accessed their data. In the results and discussion (section 6.4), we present the trends in conference proceedings paper publishing over the last 20 years in each of the databases and how their numbers and proportions of the databases compare with one another. We then present data for the ASEAN region and its ten countries and group them into four clusters based on their conference proceedings publishing patterns. The results for Indonesia were so striking that the final part of the study examines two possible explanations for the sharp increase in conference output. First, whether the growth was due to expansion in database coverage, and the second is whether the location of the conference was related to the number of resulting proceedings papers from the hosting country. The article concludes that conference proceedings have become increasingly visible in the major citation databases, that interesting country-level patterns can be revealed by analysis, and that there has been a steep growth in preferential conference paper publishing in Indonesia coinciding with an increase in locally hosted conferences.

6.2 Literature review

6.2.1 Conference proceedings

An evolutionary model of scholarly communication was described several decades ago in which academic research is communicated first as personal correspondence and in subsequent incremental forms that include open letters, conference proceedings and finally the journal article (Garvey et al., 1972a). A detailed study on the publication behaviour of 12,000 scientists over a 5-year period demonstrated that the majority of conference material presented at national meetings in America was later published as a journal article (Garvey et al., 1972b) and this pattern became broadly accepted as common practice in the exchange of scientific information between researchers. Later evidence challenged this view with conference proceedings seen by some as an end product and being accepted as evidence of scholarly activity by university tenure and promotion committees (Drott, 1995), rather than being merely a step on the way to a journal article.

The advent of the Internet spawned fundamental changes to communication of academic research findings with some results appearing on the web before they were published in journals or as conference proceedings (Goodrum et al., 2001). Work comparing conference papers with journal articles showed that conference papers are generally shorter (González-Albo & Bordons, 2011), less cited (Drott, 1995) than journal articles and their citation peak is shorter lived (Lisée et al., 2008). In many fields assessment for promotion gives higher weighting to journal articles, which serves as an incentive for scholars to adapt conference proceedings for subsequent journal submission. However, in other fields proceedings papers and journal articles are seen as different expressions of the same work (Bar-Ilan, 2010) revealing a variance in publication behaviour between fields.

There is broad variance in the relative importance of proceedings papers depending on the field of study. One study showed that roughly half the papers in ISI Proceedings – the first conference proceedings database and later known as the Conference Proceedings Citation Index (CPCI) – were assigned to the field of Engineering and that this share increased from 43 per cent in 1994 to 61 per cent by 2002 (Glänzel et al., 2006), more than six times the corresponding share found in the Web of Science (WoS) journals. In the same study, the proportion of proceedings papers classified in the field of physics grew from 25 per cent to 32 per cent over the 8-year period while the corresponding proportion of journal articles in the WoS remained stable at around 13 per cent. Similarly, the engineering field was singled out as one in which proceedings papers receive a higher proportion of citations indicating that conference material is of greater import than in other fields (Lisée et al., 2008). This was especially the case for computer science papers.

Computer science has also been identified as a field in which conference proceedings are a major venue (Bar-Ilan, 2010) for disseminating research findings. Bar-Ilan (2010) pointed out that conference papers from Lecture Notes in Computer Science (LNCS) were already indexed in the Science Citation Index before the addition of the CPCI to the WoS in 2008. In a study comprising interviews with authors and journal editors in the software engineering field, conference papers were thought to be shorter than journal articles and contain only the exciting part of the study intended to keep a specialized audience abreast of novelties in their field (Montesi & Owen, 2008). Journal articles were found to be a more mature product designed to enable readers to replicate results and form part of an archive. Furthermore, conference papers were often reworked and later published as journal articles. An extensive study of citations to non-WoS literature in 3 social science fields showed that only 2% of cited references to academic literature outside the WoS were to conference papers in psychology, only 1% in political science and even less in economics (Nederhof et al., 2010) showing very limited influence of non-WoS literature on highly cited works.

6.2.2 Databases

The WoS originally comprised 3 journal indices; the Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) and the Arts & Humanities Citation Index (A&HCI) each consisting of metadata from academic peer reviewed journals. Where conference proceedings were published in journals in these indexes the papers were labelled as both 'article' and 'proceedings paper'. A Dutch study concluded in 2007 that it was feasible to expand the WoS to include additional conference papers to provide better coverage in the field of computer science provided some technical issues were addressed such as the treatment of different versions of the same study (Moed & Visser, 2007). In 2008, Thomson Reuters merged the content from its ISI Proceedings Science and Technology database into the WoS and renamed it the conference proceedings citation index (CPCI). Although a limited number of conference proceedings were already published in journals indexed in the database, it was the incorporation of the CPCI that contributed the majority of the conference literature to the WoS. It was this addition of conference proceedings into the Web of Science in 2008 that provided an opportunity to study these papers as a unique source of scholarly material.

At about the same time, journal articles that had been adapted from meeting presentations had their document types changed from 'paper' to 'proceedings paper', the same name given to the majority of the publications in the CPCI. All documents indexed from conference material were thus assigned the document type 'proceedings paper' although they may have originated either as journal articles that made explicit reference to initially having been presented at a scientific meeting or may instead have been published in a 'book' of proceedings. Proceedings papers published in journals have been found to vary in their proportion with respect to regular articles by field with highest proportions found in computer science/information technology and applied physics (Zhang & Glänzel, 2012). Conference proceedings in journals can be further categorized into those accepted in ordinary issues and those published as special editions or monographs. Those published in monographs have been found in library and information science to be shorter, have fewer references, pass more quickly through peer review and receive fewer citations (González-Albo & Bordons, 2011).

These nuances in publication behaviour, changes in documentation types and evolution of the WoS coverage over time mean that any study of publications is susceptible to an element of misinterpretation of the results owing to the idiosyncrasies of the database. For this reason, three databases were used; WoS, Elsevier's Scopus that was released in 2004 as a more inclusive citation index, and Dimensions launched in 2018 by Digital Science. The changes in proportions of conference proceedings by different countries over time might be partially explained by the natural expansion of any single database over time. By using three databases, we are likely to illuminate the effects of any major changes in conference proceedings coverage by any one database.

Elsevier launched its proprietary citation index, Scopus, in 2004 which now covers more journals than Web of Science (Mongeon & Paul-Hus, 2016). However, the overlapping coverage of conference proceedings between WoS and Scopus is limited. Indeed, WoS appears to cover many conference proceedings that are not indexed in Scopus, and Scopus covers many proceedings not covered by WoS (Visser et al., 2021).

Dimensions, a relative newcomer to citation indexing, was launched in early 2018 and offers a free version including access to just over 100 million scholarly records. Users can identify around six million conference proceedings among the records without requiring an organisational subscription. Dimensions relies heavily on Crossref as the backbone of its data (Hook et al., 2018). Dimensions coverage is also comparable with Scopus, indeed at least 90% of Scopus indexed papers were found in Dimensions with the exception of the most recent year in which coverage dropped to about two-thirds (Thelwall, 2018) probably due to a longer time lag in indexing in Dimensions. Many Scopus indexed conference proceedings are however absent in Dimensions (Visser et al., 2021).

The different databases do not always coincide on their coverage and definition of document types, for instance Anne-Wil Harzing found that Dimensions indexed more of her conference papers than Scopus while the WoS indexed none of them (Harzing, 2019). The document type definitions are not always uniform as Dimensions classifies all items indexed from journals as articles while the WoS and Scopus separates conference proceedings, editorial material, book reviews, letters, corrections, and others as specific document types. For instance, we observed in this study that Dimensions classified more than 100,000 *Journal of Physics Conference Series* papers as journal articles, which the WoS and Scopus consider conference proceedings. These practices all swell the number of journal articles in Dimensions which in turn makes the proceedings content seem a smaller proportion.

This indicates the coverage practices of the WoS, Scopus and Dimensions are independent of each other, and that multiple databases should be used whenever possible to build a picture of the real situation without relying too heavily on any single database.

6.2.3 ASEAN region

Publishing behaviour is subject to universal forces such as the requirement for academics to publish their work and demonstrate some form of impact. Regional differences in publication behaviour can provide an interesting basis for more specific study, for example the fast-growing emerging regions such as those of Southeast Asia where 10 countries, Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, Philippines, Singapore, Thailand and Vietnam, have formed the ASEAN group of nations. The organisation was founded in 1967 (ASEAN, 2019) in order to promote peace, collaboration, regional research, and common social, cultural and economic values.

The ASEAN countries provide an ideal case study because they share a common geographical region and have shared scientific goals. They have all rapidly increased their scientific output in recent years and it is interesting therefore to observe whether publication patterns emerge at individual country level that are obscured by global trends. Other studies have used the ASEAN countries for similar reasons.

In 2011, Nguyen and Pham studied the scientific output and impact of the ASEAN countries and described four distinct groups with Singapore alone in the most advanced bloc, Malaysia and Thailand in the second, Indonesia, The Philippines and Vietnam in the third, and finally Brunei, Cambodia, Laos and Myanmar making up the fourth cluster (Nguyen & Pham, 2011). These researchers also found correlations between scientific output and both the Knowledge Index and the Knowledge Economy Index published by the World Bank in 2008 (Chen & Dahlman, 2005) each of which confirmed the four distinct groups of ASEAN countries observed by Nguyen and Pham. The rate of economic growth of countries such as Singapore has been linked to the investment in research and development (R&D) (Nguyen & Pham, 2011). A large 40-country study of investment and publication data demonstrated a positive correlation between R&D spending and research publication output (Meo et al., 2013).

As part of a UNESCO report, Moed and Galevi described a bibliometric model to group 25 Asian countries' relative stage of development by organizing them into three clusters based on the proportion of their internationally co-authored papers and the geographical location of the collaborating countries (Moed & Halevi, 2014). The first group includes only Singapore among the ASEAN countries and whose papers were often co-authored with researchers in China, Hong Kong and Macau. The second cluster also only features one ASEAN country, Malaysia, which is more linked to research groups in India, Pakistan and Iran, while the third group contains the remaining members in a Southeast Asian cluster. These authors also draw a link between the state of a country's development and the ratio of the number of its doctoral students to the number of its publications. For example, Japan, a developed country has roughly the same number of doctoral students as Indonesia but produces 100 times the number of publications as Indonesia (Moed & Halevi, 2014). This paper built on the idea of organising the countries into clusters based on their conference proceedings publishing behaviour.

6.3 Methods and data sources

The method followed for analysing conference proceedings followed the same pattern for each database. First, publications were extracted from the databases and organised by document type and country. Proceedings papers were compared with all document types to determine the proportion of conference papers published by authors in each country. The second step was to analyse the patterns from each of the ten ASEAN countries and group them into four broad clusters based on distinctive publishing behaviour. Finally, the findings were validated through personal knowledge of database indexing, and interviews with a conference organiser in Indonesia, and an international publisher of conference proceedings.

Data for the bibliometric studies was extracted from in-house versions of the WoS, Scopus and Dimensions databases hosted by CWTS at Leiden University. The CWTS version of WoS comprises five citation indices; the Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index – Science (CPCI-S) and the Conference Proceedings Citation Index – Social Sciences & Humanities (CPCI-SSH) and is collectively referred to as WoS 5-ed meaning five editions. Neither the Book Citation Index (BkCI) nor the Emerging Sources Citation Index (ESCI) were included in the study because they are not hosted in the system. The extraction includes records up to the end of 2019, which allows fairly reliable study of all years up to and including 2018. The data for 2019 was not complete and therefore results for that year are not presented in the study. The in-house version of the WoS database hosted by CWTS is extracted multiple times every year from the database provider and subject to an extra round of cleaning, address unification and data validation. Scopus data were extracted in May 2019 and Dimensions data in June 2019 with the expectation that data for all years up to 2018 would be reliable enough to use in the study.

These stable and accurate versions of the databases are therefore more suitable for bibliometric study than the online versions that are designed for immediate reflection of the most recently indexed papers. Therefore, the interface versions of these databases will necessarily show slightly different results than those presented in this study.

The papers used in this study were those with document types ‘proceedings paper’ and ‘proceeding’ from Scopus and Dimensions respectively, and all papers in the CPCI for WoS. In each case, the proportion of conference papers was defined as the number of conference papers divided by the total output. The total output means the entire database for Scopus and Dimensions, while for WoS it means the three journal indexes plus the two conference proceedings indexes, known as WoS-5ed.

To determine Indonesia’s rank in conference output, we used author affiliations on the proceedings papers and noted Indonesia’s position. We browsed the conference series publisher websites to determine the location of the conferences and listed the proportion of conferences that were hosted in Indonesia. For this section we used only the WoS and Scopus but not Dimensions because the conference series could not be reliably identified from the in-house version of the database.

6.4 Results and discussion

6.4.1 Conference proceedings in citation indexes

As the databases have grown, the number of conference proceedings records has generally increased over time (Fig. 6.1). Looking back 20 years we can see that WoS usually indexed between 20-40% more conference papers each year than Scopus, although the WoS surge in 2003 increased the gap to 67% and the Scopus surge in 2005 closed the gap again to 23%. In the subsequent three years, 2010-12 the number of conference papers in the WoS dipped and was overtaken by Scopus which continued its rise to a peak in 2011. The conference paper indexing in WoS then increased again whilst Scopus fell back and the difference between the two databases in 2015-17 stabilised at about 40%. The 2018 figures showed a drop in WoS conference papers, although these might still change as records are belatedly indexed. The conference papers indexed in Dimensions appeared to be independent of WoS and Scopus growing at a slower rate and usually totalled between 50-62% of the Scopus papers until 2015. In the four years to 2017, Dimensions added conference papers at a rate of 9-11%, far higher than WoS and Scopus.

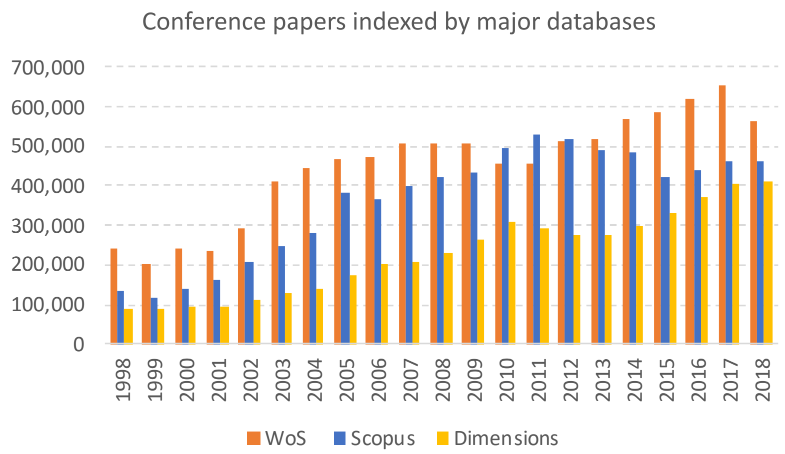


Figure 6.1. Conference papers in citation indexes

The number of conference proceedings records in the WoS was fairly even until 2001 in line with the rest of the WoS as demonstrated by the stable proportion of conference papers in the WoS (Fig. 6.2). Thereafter followed a steep increase of conference records, 77% growth between 2001 and 2003. This growth was faster than the rest of the WoS database and resulted in conference papers contributing an increasing proportional share of the database that rose from 16% to 25% in those two years. In the subsequent fourteen years the number of conference records in WoS grew only 58% and its proportion of the database reduced and stabilized at around 21%.

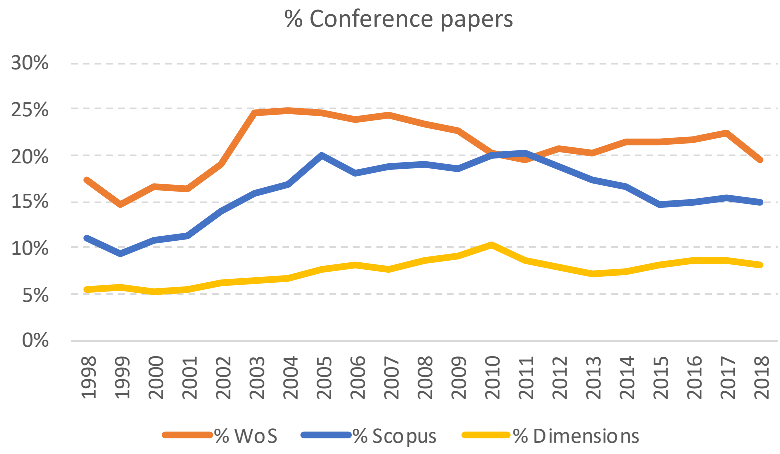


Figure 6.2. Conference papers in WoS

Tracking the share of conference papers in WoS, Scopus and Dimensions showed that over the past 20 years, WoS has consistently indexed a higher proportion of conference papers than the other two databases. The only exceptions were in the year 2005 when a peak of Scopus-indexed conference papers came close to the WoS share, and in the years 2010-11 when the proportion of conference papers in Scopus briefly surpassed that in the WoS. The share of conference papers in WoS reached its lowest point since 2002 in 2011 which coincided with the highest peak in Scopus. Dimensions consistently contained a lower share of conference papers than both the established databases.

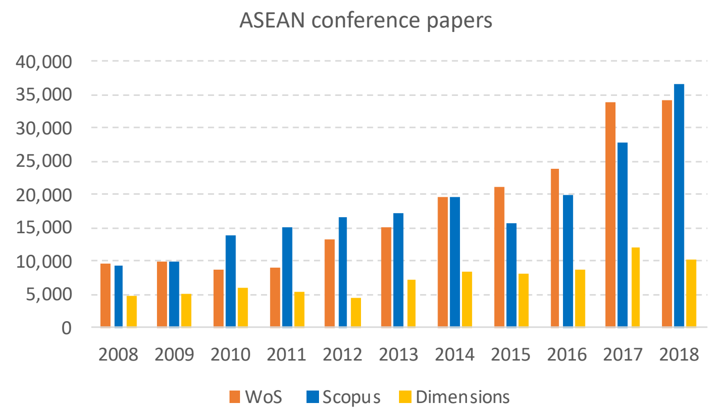
6.4.2 Regional trends in the ASEAN countries

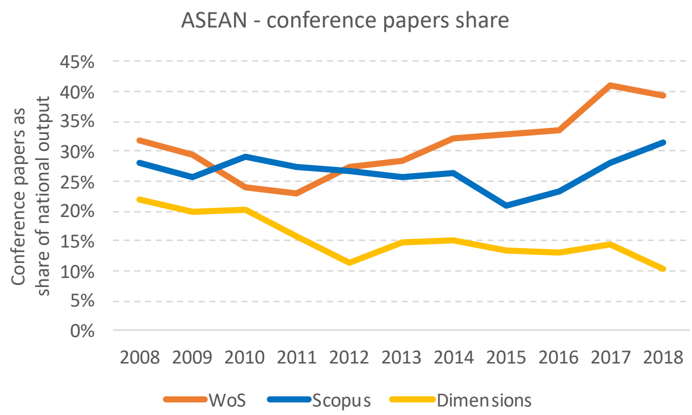
The second part of the study focussed on the specific trends observed in the ten countries of the ASEAN region. As global figures are the average of all the world's regions, it might be possible to uncover more interesting patterns by looking at individual countries. The ASEAN countries have the stated aim of working together on scientific research and provided an appropriate set of countries for study.

Results are first presented for the region as a whole and subsequently for the individual countries. We looked within the country-level findings to try to cluster countries together that exhibited similar conference proceedings publishing patterns. For instance, countries which showed increased conference output to a greater extent than the regional trend could be clustered together, while those countries whose conference publishing declined would be another cluster. A third cluster was for countries which followed the regional trend, and yet a fourth was for countries whose output was too low to assign to any other cluster with confidence.

The conference proceedings output from the ASEAN region started in the WoS and Scopus fairly late and from a very low base and therefore this paper focusses on those published in the last 10 full years plus 2018. Data for 2018 were included as almost complete but not completely reliable at the time of analysis. Therefore, the 2018 figures are included to provide insight into the latest trends but we also go back 10 full years prior to 2018 to ensure we covered a full decade of complete data.

A comparison of ASEAN conference proceedings between WoS, Scopus and Dimensions (Fig. 6.3) demonstrated some similarity with the global pattern, specifically the higher proportions of conference papers in the WoS and the lowest share in Dimensions. The share of conference papers in Scopus exceeded that of WoS only in 2010-11 and this was by a greater margin than we saw in the global graph. Individually, each of the ten countries in the ASEAN region demonstrated a different pattern and we attempted to group these patterns into four clusters as described above.





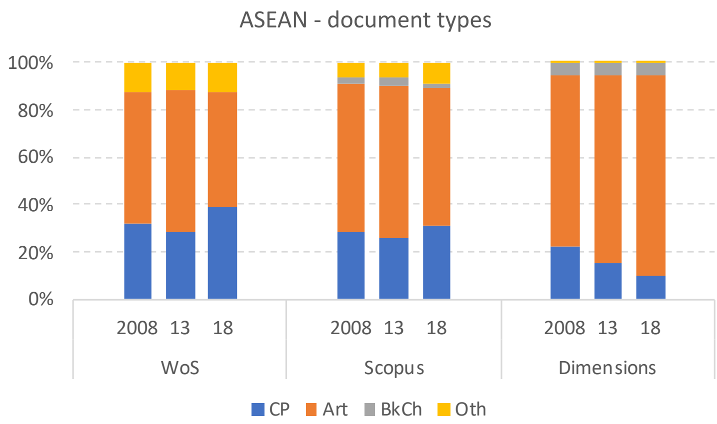


Figure 6.3. Conference proceedings publishing trends – ASEAN

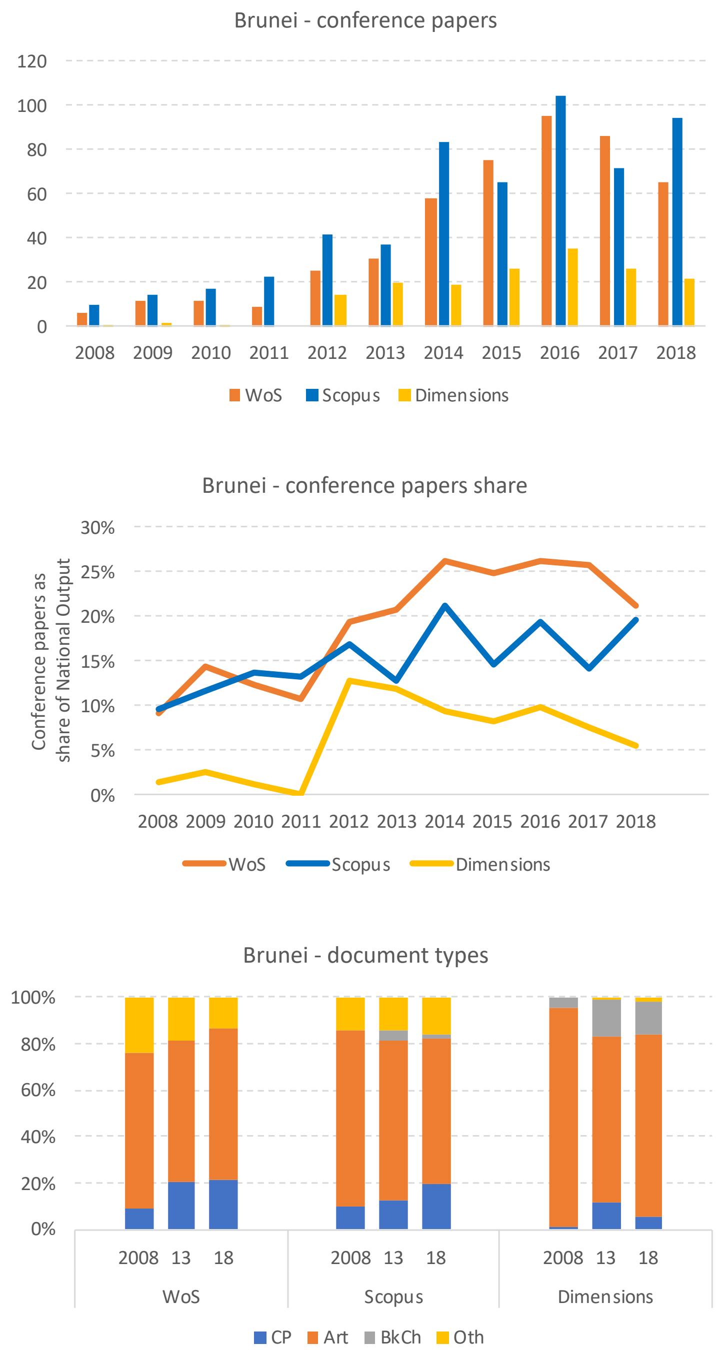
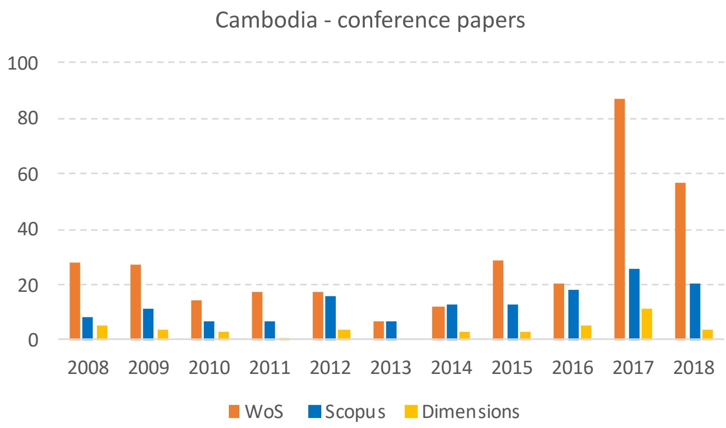
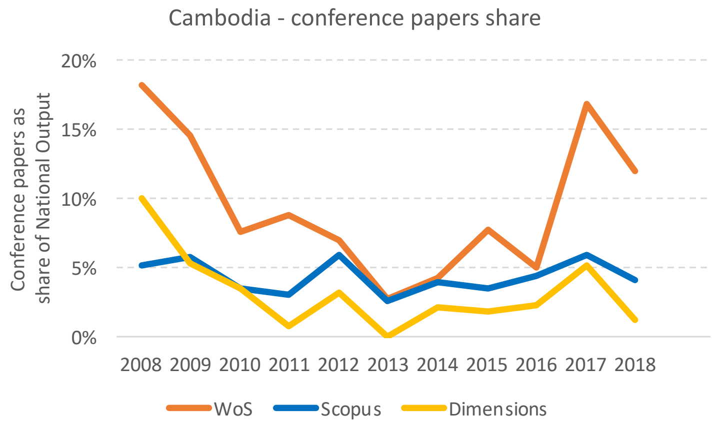


Figure 6.4. Conference proceedings publishing trends – Brunei





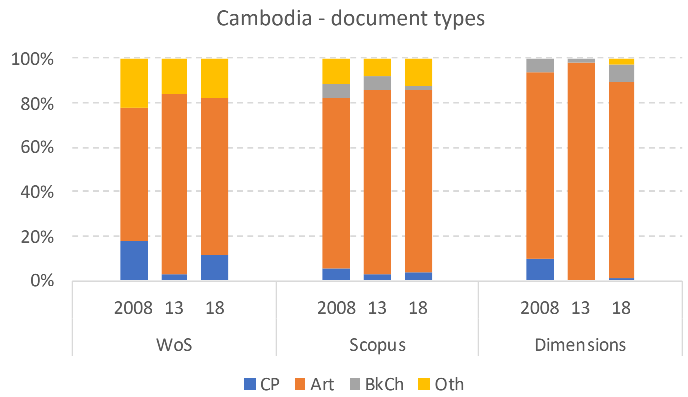
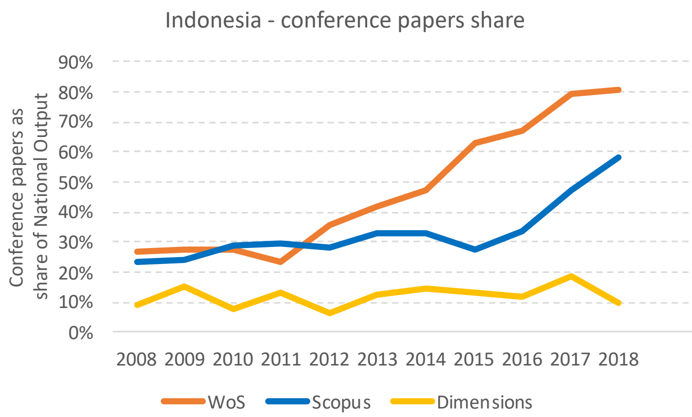


Figure 6.5. Conference proceedings publishing trends – Cambodia



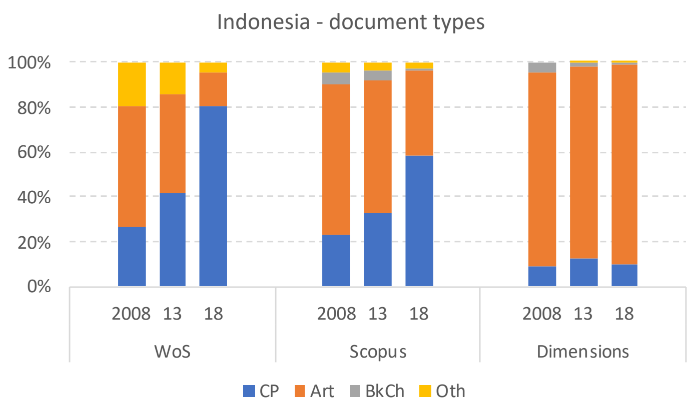


Figure 6.6. Conference proceedings publishing trends – Indonesia

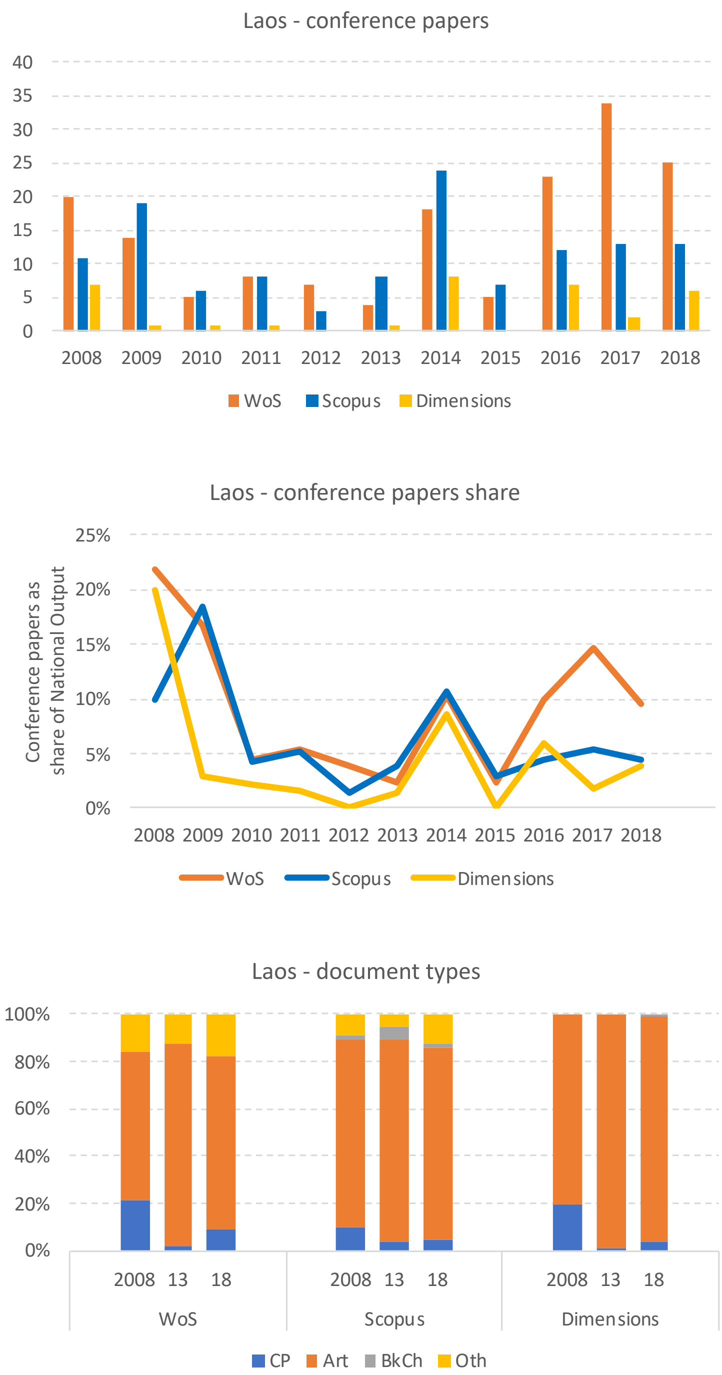


Figure 6.7. Conference proceedings publishing trends – Laos

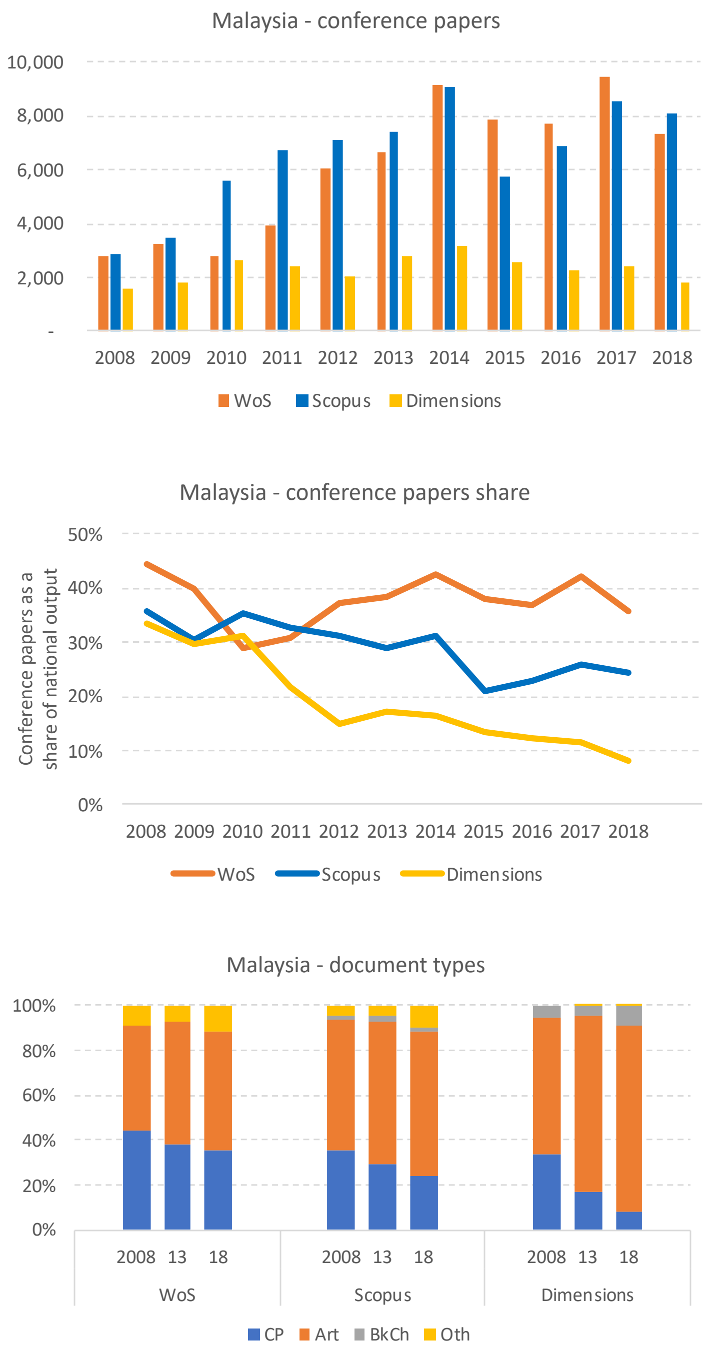


Figure 6.8. Conference proceedings publishing trends – Malaysia

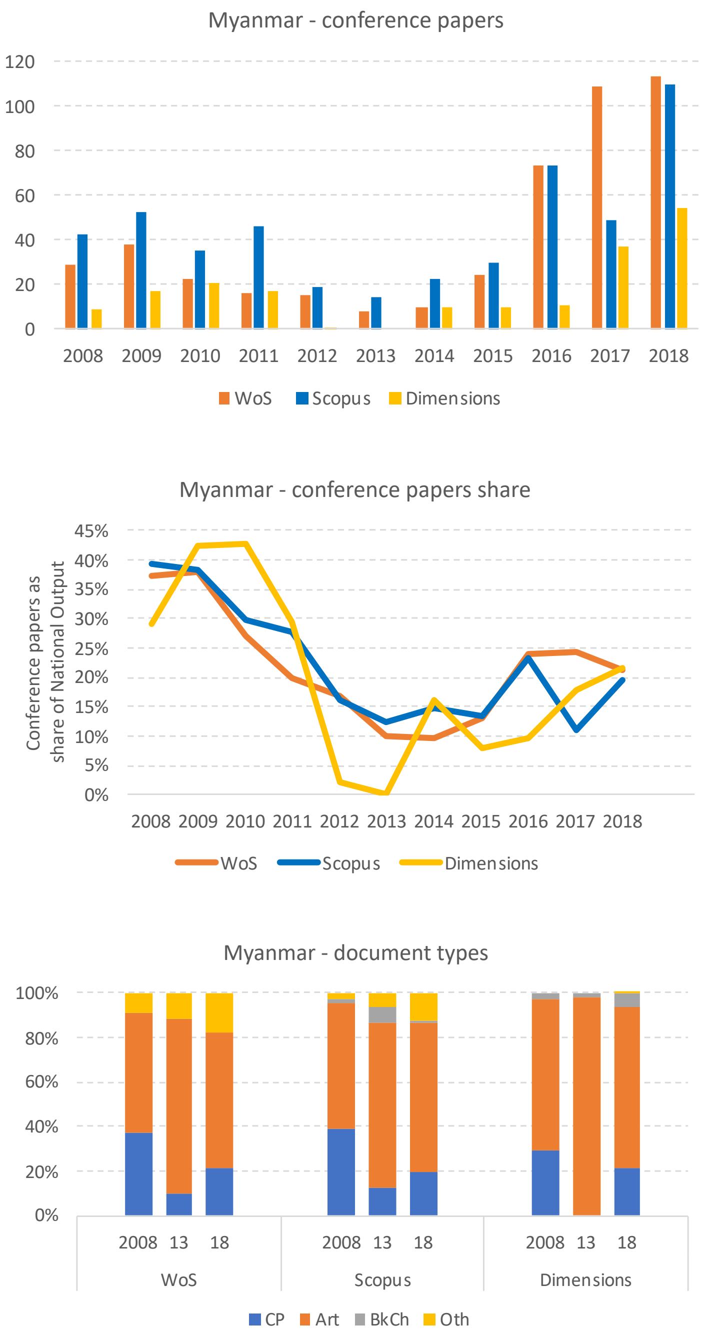


Figure 6.9. Conference proceedings publishing trends – Myanmar

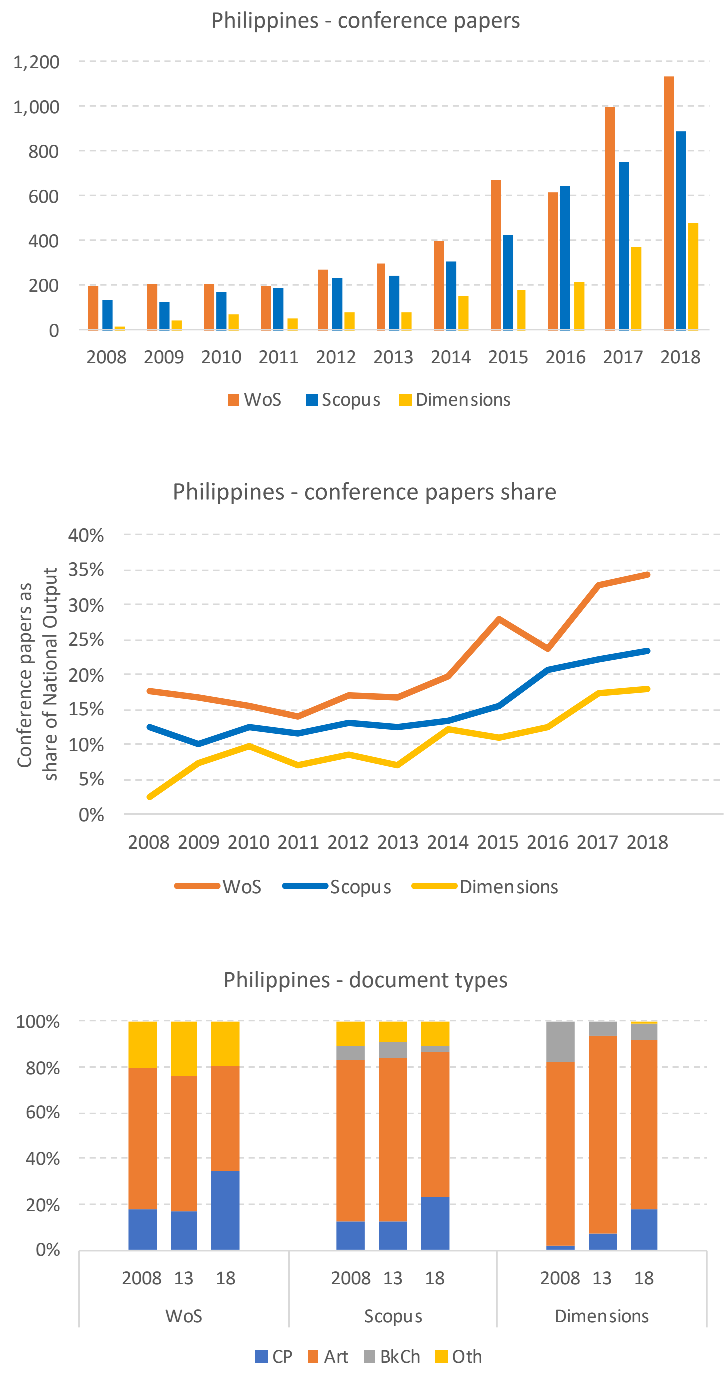


Figure 6.10. Conference proceedings publishing trends – Philippines

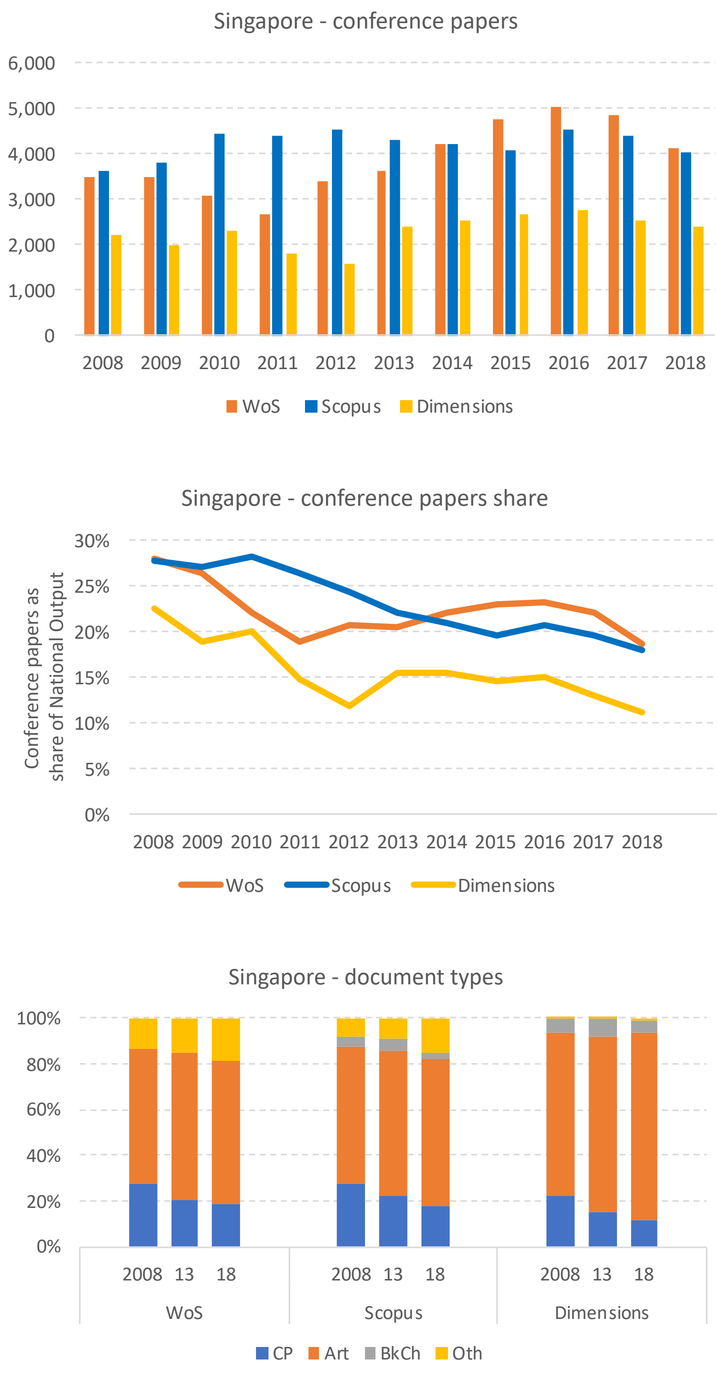
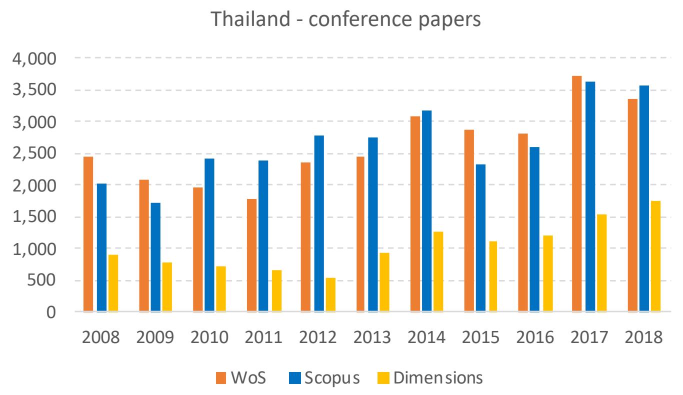
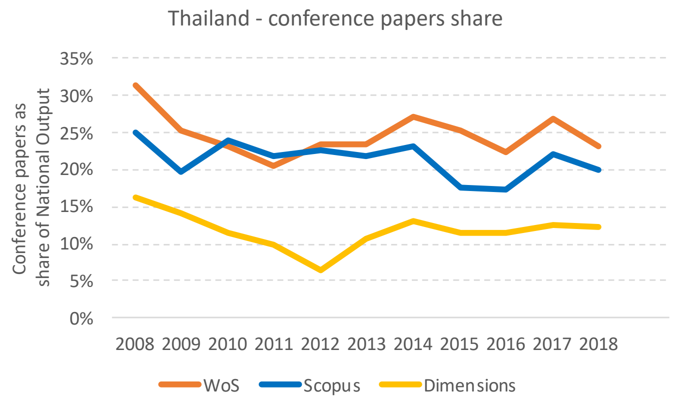


Figure 6.11. Conference proceedings publishing trends – Singapore





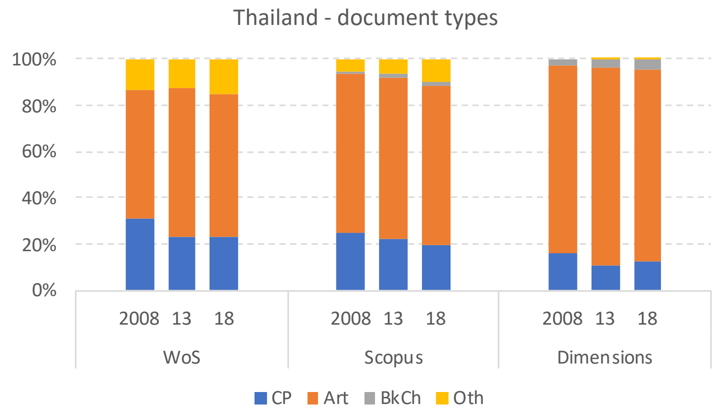
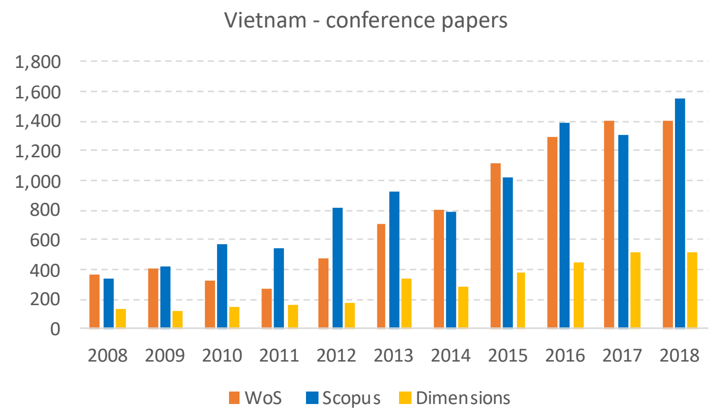
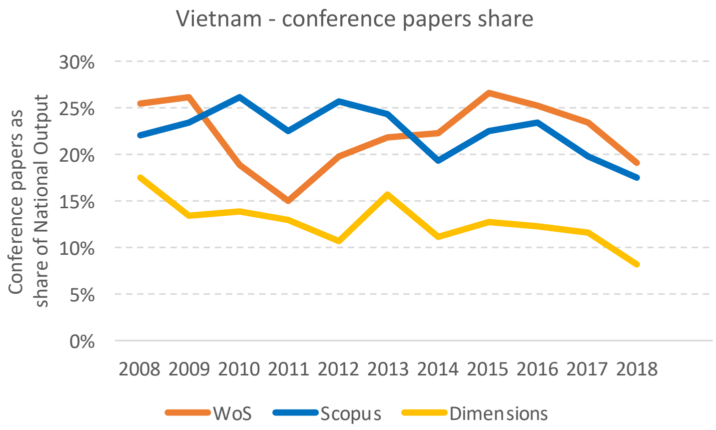


Figure 6.12. Conference proceedings publishing trends – Thailand





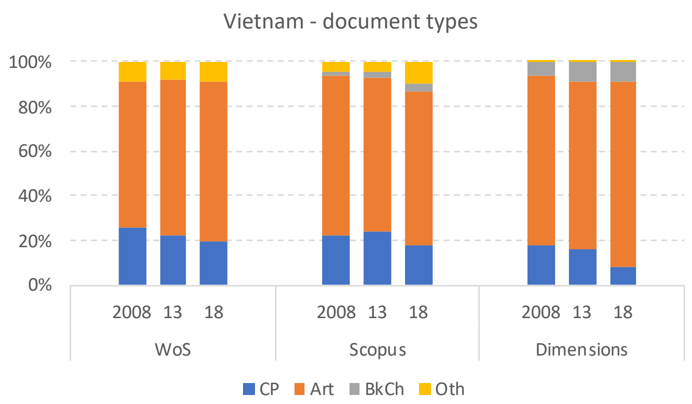


Figure 6.13. Conference proceedings publishing trends – Vietnam

6.4.3 ASEAN country clusters

The number of conference papers published by authors in the ASEAN region and their proportion of overall publications has been diverse, with each country at first glance seeming to display its own unique pattern of growth (Figs. 6.4 – 6.13). Upon closer inspection, it was possible to group the ten patterns into four clusters.

The first cluster included Indonesia and the Philippines because their patterns (Figs. 6.6 and 6.10) were both characterised by steep growth in the share of conference proceedings papers relative to other document types. The proportion of conference papers published by authors in the Philippines (Fig. 6.6) has approximately doubled since 2012 in each of the three data sources. The publishing output from the Philippines has increased in that time across all document types, however the conference output has increased more quickly. The case of Indonesia is even more remarkable (Fig. 6.10) since the share of conference papers has reached 80% in the WoS and 58% in Scopus. Contrary to expectations, the growth was not reflected at all in the Dimensions data which showed output of conference papers and journal articles growing at almost exactly the same rate and more recently the conference proceedings and book chapters losing some share to journal articles. In the WoS and Scopus databases, the expansion in the share of conference papers was clearly at the expense of journal articles, and in Scopus also of book chapters.

The second group included only Singapore, whose share of conference papers among its output has consistently fallen (Fig. 6.11) in all three data sources over the past 10 years. The conference paper output was fairly flat for Scopus and Dimensions for the entire period, whilst there was a sustained period of growth in the WoS for a five-year period from 2011 to 2016 before the output fell back again. Therefore, as the output of other document types has grown, the share of conference papers has steadily declined in each database studied. In the Scopus database, the decline in conference paper share was taken up mainly by journal articles as book chapters also saw a decline. Meanwhile in Dimensions, the book chapters dropped less, and the journal articles took up the share. These figures did not demonstrate any major shift in publishing behaviour and were most likely due to an acceleration of journal article publishing with which conference output simply did not keep pace.

The third group comprised three countries; Malaysia, Thailand, and Vietnam, each of which generally followed the patterns seen at global level. This pattern was characterised by steady growth in the conference paper output that was in line with growth in publications of all document types and which therefore resulted in a relatively flat conference papers share graph. Malaysia did not fit the pattern perfectly because the conference share declined slightly in WoS and Scopus (Fig. 6.8) and significantly in Dimensions. The proportion of Malaysian conference papers was similar in all three databases at the beginning of the study period and diverged with a higher proportion of conference papers in WoS and a lower share in Dimensions. For Thailand and Vietnam (Figs. 6.12 and 6.13), the share of conference papers in Scopus tracked that of WoS and challenged WoS to varying degrees during two periods, the year 2005 and the years 2010-12, the period when there was a general reduction in conference proceedings indexing in WoS. The loss of conference paper share was taken up in each database mainly by journal articles.

The fourth group comprised countries with very low output of academic papers and more or less stable proportions of conference papers. The group included Brunei, Cambodia, Laos, and Myanmar (Figs. 6.4, 6.5, 6.7 and 6.9). Although some differences could be observed for instance the trough in

Myanmar's output between 2012-15, the low number of papers would be expected to exaggerate the significance of such observations.

6.4.4 The case of Indonesia

The most striking result was seen in the WoS data for Indonesia that showed a sharp inflection in the proportion of conference papers in national output starting after 2011. In the following seven years, the share of conference papers rose from 23% to 80% of total annual output (Fig. 6.10). Indonesia's authors also increased their publication of other document types in the WoS, for instance journal article output has increased by 16.1% per year since 2008. However, this was far exceeded by the rate of conference papers that grew at 47.1% per year since 2008. The trend was less extreme but still marked in the Scopus database. Conference paper output from Indonesia increased more than six-fold in Scopus between 2008 and 2015 but remained within a range of 23% - 33% of Indonesia's output in the overall database. In the following three years, the number of Scopus-indexed conference papers from Indonesia multiplied by more than eight times, increasing faster than the journal article rate. By 2018 Indonesia's conference papers accounted for almost three-fifths (58%) of the country's Scopus output. The trend in the Dimensions database did not appear to follow that of the WoS and Scopus. The number of conference proceedings written by Indonesian authors fluctuated in Dimensions until 2013 and accounted for between 7% and 16% of national output. By the year 2017, the share had jumped to 18.6% accompanying a nearly four-fold increase in conference publications compared with 2015. However, the proportion fell back again in 2018, although this might have been affected by the rate at which conference papers were indexed in the database. The most recent years are less reliable, and the relative proportions might stabilise as 2018 data becomes more complete.

6.4.5 Conference coverage expansion or organic growth?

It is worth considering that Indonesia's increased conference representation in the WoS and Scopus might be due simply to the fact that these databases have added a number of conference series that contained large numbers of Indonesian papers prior to incorporation. In order to test this explanation, we looked at the six conference series with most Indonesian papers in the WoS and the corresponding six in Scopus in the last five years which was the period during which the major growth occurred. The aim was to determine whether these series were recently added to the databases and if they contained large numbers of Indonesian papers prior to their indexation.

The top six conference series in terms of Indonesia authored papers are shown for the WoS (Table 6.1) and then for Scopus (Table 6.2). Four of the top conference series were the same in WoS and Scopus while each database included two unique series in their version of Indonesia's top six. It was not possible to identify the specific conference series for Indonesian authors in the Dimensions database and therefore this section of the study used only the WoS and Scopus. Also shown is the percentage of that year's conference within the series that were physically hosted in a location in Indonesia.

The results from the WoS showed that in each of these series the number of Indonesian papers grew quickly in the five years from 2014 to 2018 (Table 6.1), more than seven-fold in the long-established *AIP Conference series* and more than 470-fold in the *IOP Conference Series: Earth & Environmental Science* (Fig. 6.14). In most of the other series, there only a handful of Indonesian papers in 2014 but thousands in 2018. In the *Advances in social science education and humanities research* there were no

Indonesian papers at all in 2014 and 2,441 in 2018. One could argue that many of these series were relatively new and have themselves grown from a very low base over the same time period of this study. That growth could be an explanation for the rapid growth in papers in these series from any country, not just Indonesia. However, the Indonesian contribution has grown a lot faster than the total as attested to by the increased proportion of papers from that country. In 2014 the proportion of published papers in these series from Indonesia ranged from 0% to 5% while by 2018 the range was 15.3% to 52.5%. Only two of the six conference series studied were added to the WoS since 2014. In one of these (*Advances in Social Sciences, Education & Humanities Research*), Indonesian papers were not present in the first year. In the other case (*AEBMR-Advances in Economics, Business & Management Research*), Indonesian scientists authored fewer than 10% of the resulting proceedings papers in its first year.

The least remarkable growth was seen in the *AIP Conference Proceedings* in which Indonesia ranked 8th by number of conference papers in 2014 and rose to 2nd by 2018. In the other five series Indonesia published more conference papers than any other country in the world in 2018. This is a notable achievement when just five years earlier, the country ranked between 17th and 41st in the world and did not publish any papers at all in one of the series.

In Scopus the figures are no less impressive. In the six conference series in Scopus with most Indonesian papers, we see the same remarkable growth (Table 6.2). Four of those conference series were the same in WoS and Scopus although the number of papers indexed varies between databases. Two of the series shown in the WoS table are not indexed in Scopus while two that are in the Scopus table are not indexed in the WoS. In both the *MATEC* and the *Es3 Web of Conference* series, there were no Indonesian authored papers in 2014 but more than a thousand in 2018 (Fig. 6.15). In the *Es3 Web of Conference* that accounted for more than a quarter (28.6%) of the world share of papers which was more than any other country in the world. In the *MATEC Web of Conferences* the Indonesian authored papers placed Indonesia as the third ranked country by output by 2018.

6.4.6 Conference hosting

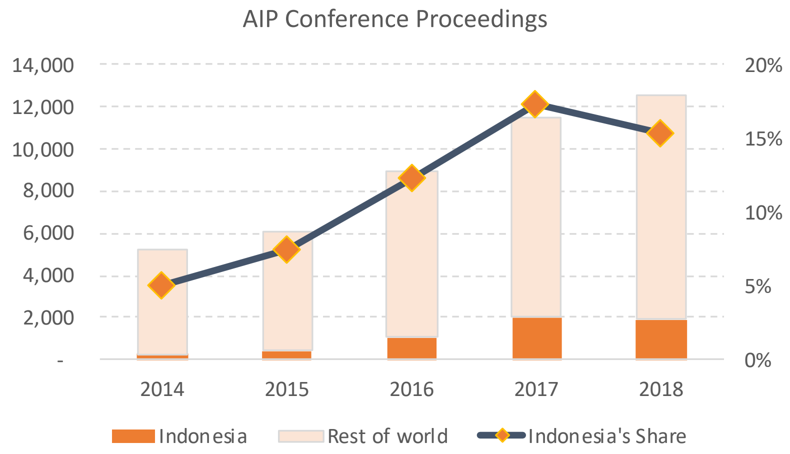
The data presented in Tables 6.1 and 6.2 show Indonesia's rank in the countries with most author affiliations on the proceedings papers in six conferences series by year. This means the publication year, and not necessarily the year the conference took place. The two columns furthest to the right show the proportion of conferences in the series that were hosted in Indonesia. The year the conference was hosted did not always coincide with the year in which the proceedings were published. Sometimes a conference is held at the end of one year, say in December and even a quick publisher will publish the conference proceedings in the following year. That means the publication numbers and conference hosting dates in Tables 6.1 and 6.2 are not infallibly related although we have made the assumption that in the majority of cases the relationship holds true. Even so, the data clearly shows an increase in the proportion of conference papers published by Indonesia based authors and an increase in the proportion of conferences hosted in Indonesia.

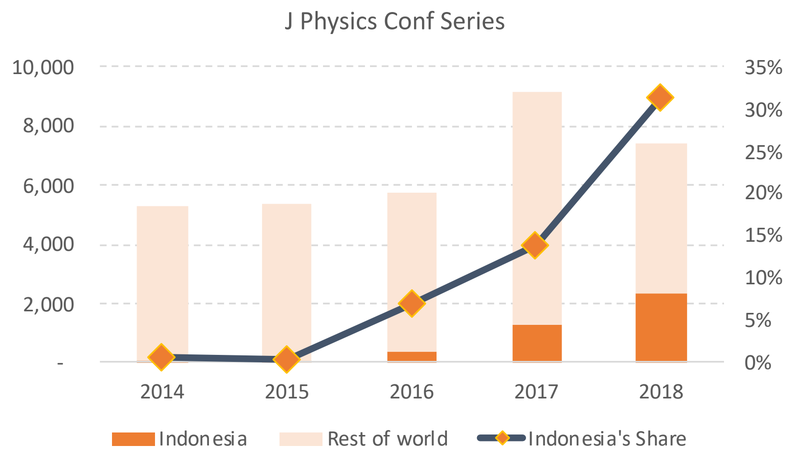
Table 6.1 Indonesian proceedings in conference series – WoS

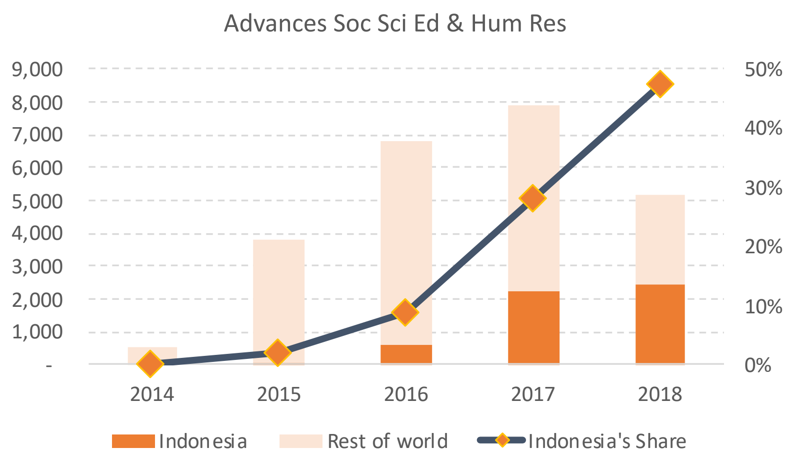
Conference proceedings 2014 and 2018	Indonesia – WoS							
	Conf papers (Indonesia)		World share (%)		Country rank by conf papers		Conf hosted (%) in Indonesia	
	2014	2018	2014	2018	2014	2018	2014	2018
AIP Conference Proceedings	260	1,918	5.0	15.3	8th	2nd	6.1	17.3
Advances in social science education and humanities research	0	2,441	0.0	47.5	Not ranked	1st	15.4	59.3
Journal of Physics Conference Series	28	2,314	0.5	31.3	41st	1st	2.2	19.6
IOP Conference Series: Earth & Environmental Science	6	2,821	0.8	42.6	17th	1st	16.7	47.9
IOP Conference Series: Materials Science & Engineering	8	1,846	1.2	19.8	25th	1st	6.3	17.1
AEBMR - Advances in Economics, Business & Management Research	Not indexed	621	Not indexed	52.5	Not indexed	1st	50.0	50.0

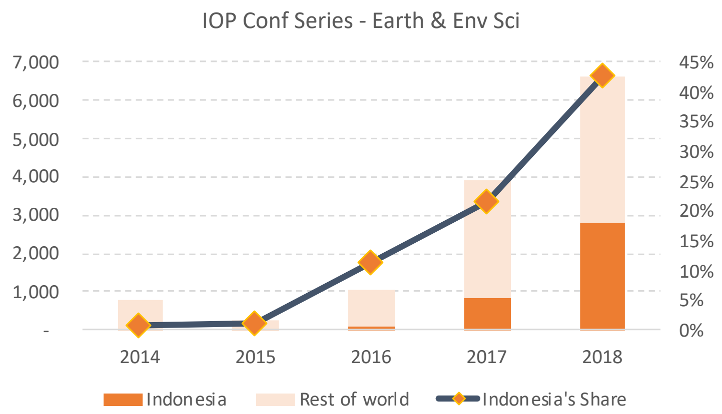
Table 6.2 Indonesian proceedings in conference series – Scopus

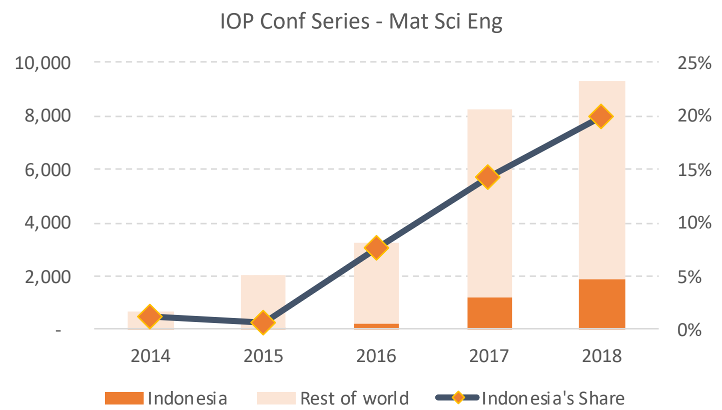
Conference proceedings 2014 and 2018	Indonesia – Scopus							
	Conf papers		World share (%)		Country rank by conf papers		Conf hosted (%) in Indonesia	
	2014	2018	2014	2018	2014	2018	2014	2018
AIP Conference Proceedings	262	1,924	6.1	15.0	7th	2nd	6.1	17.3
Journal of Physics Conference Series	32	3,778	0.6	26.2	39th	2nd	2.2	19.6
IOP Conference Series: Earth & Environmental Science	9	3,691	1.2	37.8	12th	2nd	16.7	47.9
IOP Conference Series: Materials Science & Engineering	10	2,721	1.1	17.3	24th	2nd	6.3	17.1
MATEC Web of Conferences	0	1,362	0.0	14.0	Not ranked	3rd	0.0	13.8
Es3 Web of Conferences	0	1.064	0.0	28.6	Not ranked	1st	0.0	18.8


AIP Conference Proceedings
14,000
12,000
10,000
8,000
6,000
4,000
2,000
-
20%
15%
10%
5%
0%
2014
2015
2016
2017
2018
Indonesia
Rest of world
Indonesia's Share


J Physics Conf Series
10,000
8,000
6,000
4,000
2,000
-
35%
30%
25%
20%
15%
10%
5%
0%
2014
2015
2016
2017
2018
Indonesia
Rest of world
Indonesia's Share


Advances Soc Sci Ed & Hum Res
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
-
50%
40%
30%
20%
10%
0%
2014
2015
2016
2017
2018
Indonesia
Rest of world
Indonesia's Share





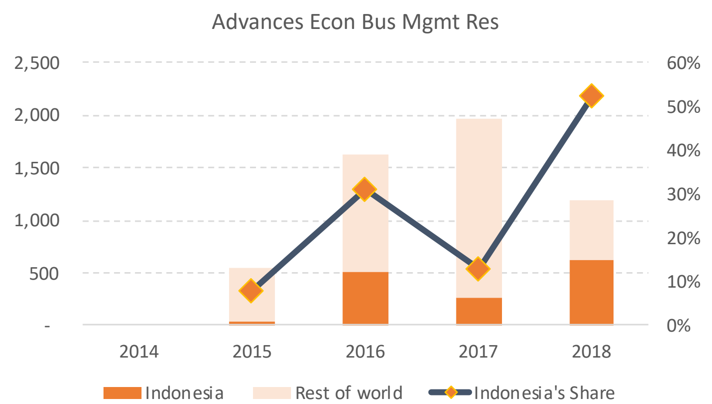
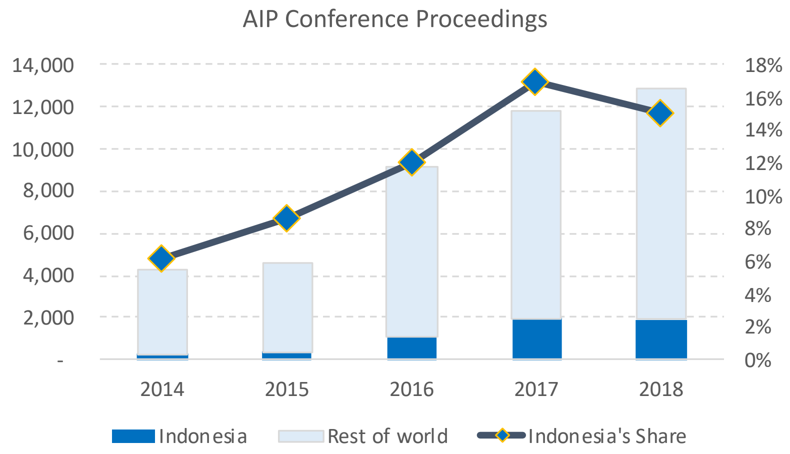
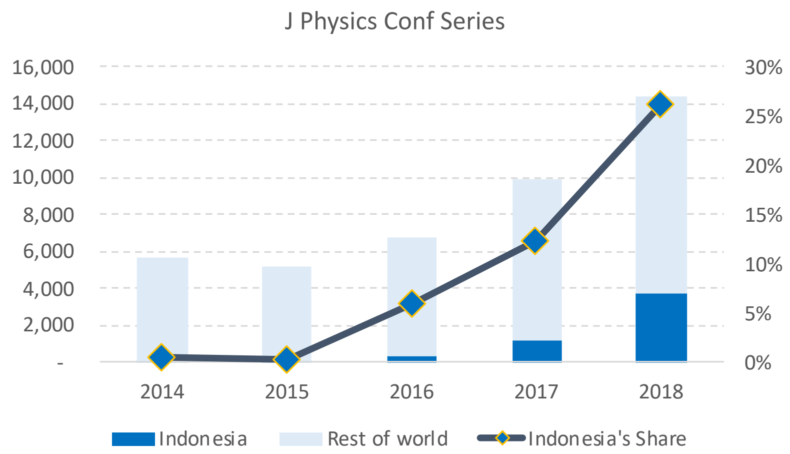


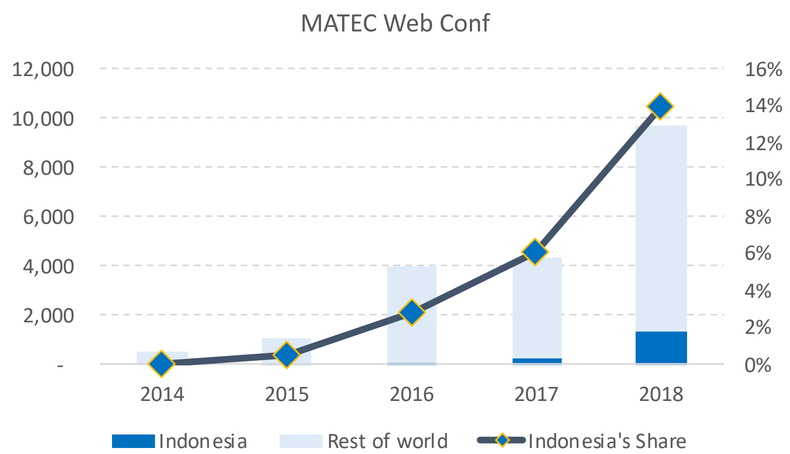
Figure 6.14. Indonesian proceedings in conference series – WoS



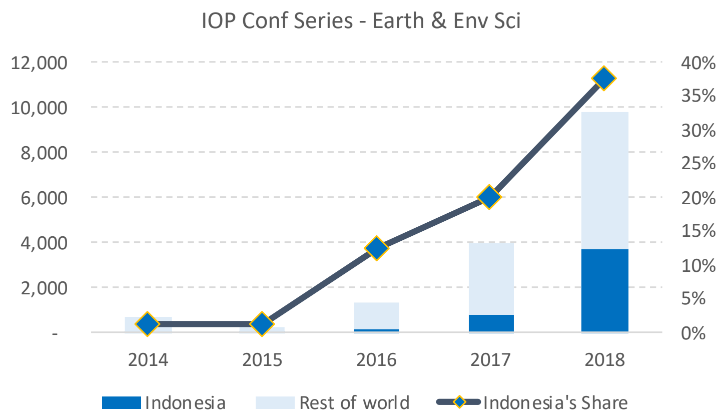
AIP Conference Proceedings
14,000
12,000
10,000
8,000
6,000
4,000
2,000
-
18%
16%
14%
12%
10%
8%
6%
4%
2%
0%
2014
2015
2016
2017
2018
Indonesia
Rest of world
Indonesia's Share



J Physics Conf Series
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
-
30%
25%
20%
15%
10%
5%
0%
2014
2015
2016
2017
2018
Indonesia
Rest of world
Indonesia's Share



MATEC Web Conf
12,000
10,000
8,000
6,000
4,000
2,000
-
16%
14%
12%
10%
8%
6%
4%
2%
0%
2014
2015
2016
2017
2018
Indonesia
Rest of world
Indonesia's Share



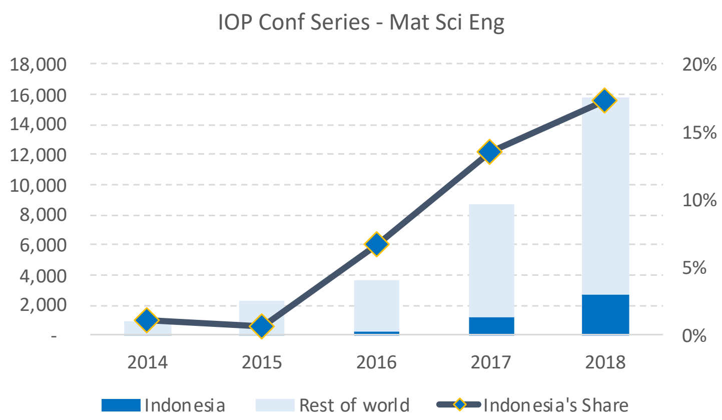


Figure 6.15. Indonesian proceedings in conference series – Scopus

6.5 Discussion

6.5.1 Conference proceedings indexing

The steep increase in conference papers coincided with the launch of a CD Rom version of 'ISI Proceedings' in the early 2000s. The database owner, then Thomson Reuters, sought to create more value in this product by adding proceedings papers from conferences held in the preceding five years[6] in addition to those that were indexed annually. The additional conferences swelled the number of proceedings papers that resulted in a peak of 25% in the proportion of conference papers in the WoS in 2003. Growth tempered after 2009 as emphasis (and resources) shifted towards the indexing of books and book chapters in preparation for the launch of the Book Citation index in 2011. Once the Book Citation Index was launched, evaluation and indexing work resumed its usual pace on proceedings. In the most recent five years, the conference papers have reached a stable ratio of around 1:4 compared with journal papers as both grew at a similar rate.

In the Scopus database, the share of conference papers showed a similar pattern to that seen in the WoS (Fig. 6.2) albeit less extreme and shifted approximately four years later. The relative delay of the phenomenon is possibly due to the later addition of conference papers to Scopus as part of the Scopus Conference Expansion Programme (Elsevier, 2015) in which over 400,000 conference papers were added to the database during the years 2011-2014. This was several years after the CPCI was added to the WoS and would explain the time difference between the two databases in the surge of conference papers.

The reason for the lower proportion of conference papers in Dimensions is likely due to greater coverage of journal articles globally combined with a smaller coverage of conference proceedings papers. It may also be due to the differences in document type classification, namely that all journal material is classified as article in Dimensions while articles in the other two databases are sub-sets of journal material which also contains other document types including conference proceedings. Journal publishers are possibly more diligent than conference organisers in generating and depositing DOIs in the Crossref database which is the backbone of the Dimensions database. That would partially explain more journal content being indexed in Dimensions than conference proceedings publications. The owners of WoS and Scopus, however, have each made concerted campaigns to boost the representation of conference material which were followed by sustained increased share thereafter.

6.5.2 ASEAN region

The clustering exercise revealed sub-regional trends with the ASEAN countries. By far the most interesting country for further study was Indonesia. Its steep and sustained increase in conference publishing (Fig. 6.10) deviated so much from the global pattern that we investigated two potential explanations. The first was whether non-indexed conference series containing large numbers of Indonesian conference papers had suddenly been indexed by the databases used in the study. The author considers this unlikely, because the indexing policies did not change significantly during this period and because the analysis showed the conferences with the largest increase of papers from Indonesia were already indexed. The second was whether hosting conferences in Indonesia had a positive impact

[6] The author is a former employee of Thomson Reuters.

on output. The only two other countries to depart from the global norm were the Philippines whose share of conference papers also increased, and Singapore whose share declined. In both of these cases, the results were notable but did not show enough difference from the global picture to warrant further investigation. Slight differences from the global average by any individual country are to be expected, and those observed from the Philippines and Singapore were most likely examples of this deviance rather than any major shift in publishing behaviour.

The differences shown in Indonesia's conference publishing pattern between the three databases might be explained by the unique coverage policy of each database and the precise timing of when large conference series were added to the databases. However, there is a clear preferential increase in Indonesian conference paper output in WoS and Scopus that implicates changing publishing behaviour in Indonesia. This suggests that the most likely explanation for the observed increase in conference paper output from Indonesia can be found within the Indonesian academic community.

Increased visibility of conference papers in global databases means research managers and administrators are now able to evaluate conference proceedings papers among other indicators of an individual's productivity. This raises the possibility that academics have chosen to preferentially publish conference proceedings rather than journal articles or book chapters. If one simply counts all output found in a citation index linked to an academic, then one will be rewarding all indexed document types equally. This might become an incentivising factor in publishing one type of paper over another in countries where academic promotion policy does not distinguish between document types. Indeed, conversations with Indonesian academics and one Indonesian conference organiser supported this hypothesis. The organiser sent ministry guidelines for academic recruitment and promotion issued in 2012, 2014, and 2017. To his knowledge, there were "no rules regarding publications in the 2000s. This policy of spurring the number of publications in Indonesia was only implemented 3-5 years ago." He went on to say Indonesians were incentivised by Scopus publications which were more easily and quickly attained through conference proceeding and that "the goal was catching up with Malaysia and Singapore in Scopus output."

An alternative possibility is that increased database coverage is responsible for greater visibility of Indonesian publication activity. Most of the conference series studied were already indexed in the citation databases before any significant contribution from Indonesia, thereby dismissing the sudden addition of specific conferences to the databases as a plausible explanation for the growth. The only exception was *AEBMR - Advances in Economics, Business & Management Research* which was added to the WoS in 2015. In that year, 44 Indonesian papers were added, and that figure rose by more than 14-fold to 621 in the following four years. The results showed that Indonesian authors have indeed greatly increased their conference proceedings publication output within the last 5 years and the presence of these papers in the databases was not explained by sudden indexing decisions. In personal communication, a leading conference publisher supported the hypothesis that some authors from Indonesia are choosing the conference route as quicker than the journal article alternative and found it easier to pass peer review on conference papers than journal articles.

In almost all the conference series studied, there was an increasing proportion of the conferences hosted in Indonesia. The growing interest of the Indonesian academic community in attending conferences and publishing conference papers may well have stimulated the increased number of conferences organised and hosted in Indonesia. The conference publisher mentioned earlier also suggested that investment by Indonesia in supporting science conferences has led to the increasing number of events physically hosted in Indonesia. The combination of government investment in science, ministry guidelines incentivising publishing, and academics' preference for conference proceedings are the likely cause of the increased number of locally hosted conferences.

6.6 Conclusions

Both the WoS and Scopus have made concerted efforts at different times to add conference proceedings papers to their databases and these remain an important part of both databases. The addition of the Conference Proceedings Citation Index to the WoS caused a major jump in representation of conference material in the database. Similarly, the Scopus Expansion Project added a lot of conference proceedings to the Scopus database although this occurred several years later and consequently the conference proceedings remain more prominent in the WoS than Scopus. Dimensions indexes more journal articles and fewer conference proceedings papers than the WoS and Scopus and therefore appears to hold a lower proportion of conference proceedings. However, conference proceedings are often classified as journal articles in Dimensions which both raises the number of articles and reduces the number of proceedings papers. The backbone of the Dimensions content is Crossref and its content is therefore determined by those publishers that have deposited DOIs in the Crossref database. Journal publishers are likely to be more diligent about this process than conference organisers because DOIs are decisive when journals are evaluated for inclusion in many prestigious databases. These practices contribute to the apparently lower proportion of conference papers in Dimensions than the other two databases.

Studying the dynamics of conference proceedings publishing depends on multiple factors that affect different parts of the world to varying degrees, and these stories may be obscured when we look at the global average figures. Consequently, the second part of the paper was a case study of the ASEAN region as a well-defined group of ten component countries to identify any interesting trends. Conference proceedings output started from a low base in the ASEAN countries and on average saw a recent increase in conference proceedings publishing in the WoS and Scopus while the proportion of journal articles and book chapters and other document types declined. The ten countries were organised into four clusters with only Indonesia and the Philippines showing sustained increase in conference publishing, while Singapore declined. The other countries generally followed the global trend. Indonesia was by far the most interesting case due to the extent of the apparent wholesale shift of its country's authors towards publishing conference proceedings at the expense of other document types. Publishing of journal articles also increased but were easily outpaced by the conference papers which accounted for the majority of change in the ASEAN region's publishing pattern.

Further investigation into potential explanations for the disproportionate increase in conference proceedings publications from Indonesian academics examined the database coverage of Indonesian conference papers and the changing dynamics within the academic community. Expansion of conference proceedings coverage in both the WoS and Scopus has been shown to be partially responsible for the increased number of research papers from Russian scientists (Moed et al., 2018). The findings showed the phenomenon reflected real growth in published conference papers by Indonesian authors rather than being the result of increased database coverage. They also determined that Indonesia has rapidly and recently increased its hosting of international conferences whose proceedings are indexed in the major citation indexes. The increase in conference hosting is a plausible explanation for the corresponding growth in Indonesian conference publications, as it must be easier to attend and present a paper at home than abroad where international travel might be a barrier. Further research would be required to determine whether the locally hosted conferences bred increased output, or if improved Indonesian research had attracted international conferences to the country.

The local publishing guidelines and sometimes controversial credit-based assessment system (Singh Chawla, 2018) might have provided the conditions and stimulus for Indonesian academics to increase their publication output and advance their careers partially through a preference for publishing conference papers and those same policies could have also incentivized scientists to host international conferences in Indonesia. A recent study found an isolated case of preferential conference proceedings publishing in Europe and attributed that to the national policy for evaluating academics (Vanecek & Pecha, 2020). Follow up studies are therefore encouraged to look at the Indonesian local higher education and research promotional policies, and surveys of Indonesian scientists would be especially warranted.

Acknowledgements

Ton van Raan and Ludo Waltman for inspiring me to embarque upon this journey and guiding me expertly through it. Roy Hendroko Setyobudi for advice on the Indonesian higher education system and conferencing. Kim Eggleton for advice on conference publishing trends.

Conflict of interest/competing interests

The author declares that they have no conflict of interest.

Chapter 7

Conclusion

7.1 Summary of the main findings

Bibliometrics has developed as a scientific field in its own right based on several decades of experience with large-scale, quantitative methods and associated published studies. It has also spurned a lively debate over the extent to which we should trust the numbers. Some point to the objectivity of metrics and urge users to rely heavily on bibliometric indicators, while other commentators resist the move and claim that only expert opinion can sufficiently account for the complexities of scientific research. This dissertation neither wholly embraces, nor rejects the use of quantitative indicators. Instead, it plots a middle ground and aims to help policymakers employ quantitative bibliometric analyses with nuance and interpret results with caution in various scenarios that require assessment of research output and performance.

Chapter 1 comprised an introduction to bibliometrics and the issues raised in the context of research policy and evaluation. This introduction presented five research questions, which were addressed in Chapters 2 – 6. Each of these chapters dealt with a specific area of interest and is presented in the form of a research article either published in a peer-reviewed journal or in the process of peer review. The first area presented in chapter 2 illuminated the questions raised by the varying quality of bibliometric data, specifically looking at the way university names are presented in four major databases. Chapter 3 addressed a methodological point raised by the advent of the UN sustainable development goals as emerging scientific disciplines and the need to delineate them in new ways. The third topic in chapter 4 concerned the extent to which bibliometric analysis demonstrates the existence and trends in the much-lauded transdisciplinary collaboration. Chapter 5 described the geodiversity of research in terms of topic, researcher location, and collaboration dynamics between academics from established and developing regions of the world. Finally, chapter 6 showed how bibliometric analysis can uncover the sometimes-unintended consequences of research policy on academic publishing behaviour. Section 7.1.1 includes a summary of the insights gained relating to each of the five research questions introduced in chapter 1 of this dissertation.

7.1.1 RQ1: How does bibliometric data quality impact research policy?

In scholarly publication databases, the link between the author and their affiliated institution is used to enable bibliometric studies of universities by aggregating papers to their institutions (Calero-Medina et al., 2020). That way, we can conduct studies of the institution itself and make easy comparisons between organisations. Policymakers routinely use these techniques to allocate funding, attract talent, and identify elite institutions (Abramo et al., 2009; Debackere & Glänzel, 2004; Rinia et al., 1998).

Chapter 2 answers the research question on the impact of data quality on research policy through pairwise comparison of publications assigned to 18 selected universities between four databases: Scopus, Web of Science, Dimensions, and Microsoft Academic. In each pairwise comparison, the papers assigned to one of the universities in the primary database but not the comparator database were identified.

There are two possible explanations for university papers being retrieved in the primary database but not the comparator. One is that the databases have different coverage policies and the second is that databases assign papers to universities in different ways. The study presented in chapter 2 was more focused on understanding affiliation disambiguation and therefore separated the non-overlapping papers into one of two categories: differences due to coverage and discrepancies between affiliations. Having discarded differences due to databases coverage, it was possible to quantify papers that were present in both databases but only assigned to the university in one of the databases. It is not possible to say whether the databases were right or wrong to assign papers to any affiliation, but it was possible to quantify the discrepancies.

Reasons for discrepancies: The findings uncovered in the first part of the study reveal differences in the methods of assigning papers to specific institutions. In order to further our understanding of the possible reasons behind the discrepancies, it was necessary to manually examine random samples of papers assigned to a university in one database but not in another (although the paper was covered in both databases). The manual examination revealed four main reasons for the discrepancies which were:

- Missing affiliation: A database has indexed the article but failed to index the authors' affiliations. The paper might be found in another database with affiliations thereby causing the discrepancy between databases.
- Missing second affiliation: In cases where an author has more than one affiliation, the database has indexed the author's first affiliations but overlooked their subsequent affiliations. In this case, the paper will be retrieved only when searching the first affiliation.
- Unification: The database has indexed the paper, along with its author(s) and their address(es), but not linked the address to a unified affiliation. For instance, the study showed that some papers attributed to authors from the American University of Beirut Medical Center have been unified to AUB in Web of Science but not in Dimensions, which treats it as a separate institution.
- Assigned to wrong institution: The database has linked the author address to an affiliation, but not the one presumed intended by the author. Examples were found in which authors from LaSTRe Laboratory in Tripoli, Northern Lebanon, were erroneously affiliated to a university in Libya in the Web of Science. This probably occurred because of the appearance of the city name, Tripoli, which is the capital of Libya but also a city in Lebanon.

The comparisons in chapter 2 suggest that Web of Science and Scopus are most adept at assigning affiliation to a publication that the other databases agree with. Meanwhile, when affiliations are assigned by Microsoft Academic and Dimensions, a much larger share of the papers are not found affiliated to the same universities in the comparator databases. These findings seem to support previous reports of low completeness of affiliation metadata in Dimensions (Guerrero-Bote et al., 2021) and Microsoft Academic (Huang et al., 2020).

A major assumption in any study of universities is that the institutions are accurately identified in the source database. In chapter 2 we see this is not always the case and that there are almost limitless ways authors can identify their institutions on an academic paper. Research organisations, publishers, and database owners invest considerable effort to identify and unify address variants to a single agreed definition of all possible variants of an institution's name. The results presented in this study confirm that this practice is not always effective and that different databases often don't agree on the same definition of a university.

Chapter 2 therefore provides an example of where bibliometric studies can provide quick, large-scale analyses of university performance but the inaccuracies in affiliation data render the reports of limited use to research policymakers. As university names vary both within and between bibliometric databases, the results of any bibliometric study of universities will produce different results depending on which database is searched. Even using a single data source is not entirely reliable because some universities have more accurate affiliation unification than others. An element of expert opinion is encouraged in any assessment of university performance to balance any inaccuracies in the bibliometric data.

7.1.2 RQ2: How can bibliometrics help improve the categorisation of sustainability research?

The United Nations sustainable development goals (SDGs) were first introduced in 2016, and prior to that the goals had never been part of the subject classification of any bibliometric database. In order to assess academic contributions to progress against the SDGs, a way of identifying research publications related to the goals had to be developed. Multiple actors including bibliometric database owners and independent research groups have made attempts at developing new methods of identifying research related to each of the SDGs. One research group compared papers retrieved using its method with those retrieved by Elsevier and found a worryingly low level of agreement (Armitage et al., 2020a).

Chapter 3 of the dissertation answered this question by comparing four methods of classifying research related to the SDGs. The four methods used in the comparison were Elsevier (Jayabalasingham et al., 2019; Provençal et al., 2021; Rivest et al., 2021), STRINGS (Confraria et al., 2021; Rafols et al., 2021), SIRIS (SIRIS Academic, 2020), and Dimensions (Wastl et al., 2020). The study analysed the extent of overlapping publications retrieved from a search for SDG 13: Climate action in each method.

The study presented in chapter 3 reported low agreement and wide variation in the scholarly papers identified as related to SDG 13: Climate action by the four methods. The Elsevier and SIRIS methods agreed on only 29.2% of the papers retrieved by the two methods, which represented the highest level of agreement in the study. The lowest level of agreement was between the Dimensions and STRINGS methods, which agreed on only 13.4% of the papers found by the two methods.

Each method is made up of multiple steps, many of which are not publicly available, which makes it difficult to work out the precise cause of the differences between their results. The steps include keyword selection, operationalisation of search strategies, source database selection, subject matter expert input, assessment of recall, enhancement using artificial intelligence algorithms, and random sampling of retrieved publications.

Each method started with a list of keywords that were selected by subject matter experts. As we do not know who the experts were, what defined them as experts, or what instructions they were given, chapter 3 includes an analysis on the keywords and groups them into words related to policy, technology, or general keywords. The analysis showed that the four methods chose different volumes of keywords, Elsevier having used 458, SIRIS 228, STRINGS 98, and Dimensions 45. There were also discernible differences in emphasis of the keywords. For instance, the Elsevier method emphasised general (46%) and technical (41%) terms. The SIRIS method also prioritised general (52%) terms but then used a balance of policy and technical (both 24%) terms. Both STRINGS and Dimensions used mainly general terms (71% and 76% respectively), and then policy-related terms (24% and 20% respectively), and both methods used only 4% technical terms.

The study in chapter 3 then presented the terms most frequently mentioned in papers retrieved by each method on a series of VOSviewer maps. The comparisons enabled a visual representation of broad groups of terms favoured by each method in the comparison. The Elsevier method may have favoured terms related to the environment and climate possibly due to the relatively large number of keywords used related to these topics. The SIRIS method retrieved publications more focussed on the technical aspect of carbon emissions, having employed very specific terms such as "orbiting carbon observatory". The STRINGS method elicited papers more focussed on climate change, while Dimensions retrieved papers with more emphasis on energy and policy.

The deviation of keywords already used in the initial step of the methods suggested the differences were likely to become compounded thereafter.

The Times Higher Education Impact Ranking is based on Elsevier's method, which involves a combination of expert input, keyword validation, search operationalisation using the Scopus database, more expert input, scaling by machine learning algorithm, and fine tuning. If the same ranking were to use an alternative method, the listing would surely look rather different because of the considerable variance in the publications included in the analysis. It is important for users of this ranking to be aware that it is currently in an exploratory phase and should not be taken as a serious report on the output or impact on sustainability research.

Each method revealed only part of their workings in publicly available documentation and these parts did not overlap. It was therefore not possible to include direct comparisons for each individual step, only the final result. Each of the methods contains an element of enhancement, and there is no way for independent researchers to reproduce it. The experts used, instructions given, operationalisation of the keywords, thresholds, machine learning, and NLP techniques used are all proprietary to the owner of each method. This introduces the concept of the so-called 'black box' into an already complex method. This is unfortunate because it renders the methods immune to reverse engineering. As there are several steps in the methods, the differences between them compound one another, making it even more difficult to unpack the precise impact of the black box. Each method owner would need to provide the key, and this is not currently on offer. Instead, we are obliged to accept the presence of a black box and work with the final dataset of publications. We can learn a lot from analysis of the end result, but to really understand the impact of each step in the process, it will be necessary to systematically deconstruct the methods in subsequent studies.

The findings presented in chapter 3 provided evidence that supported the earlier study that showed broad discrepancy between Elsevier's method and two versions of the authors' own method (Armitage et al., 2020a). Similar to the study in chapter 2, these findings should alert policymakers to the risks of relying on a single source or method of identifying sustainability-related research in any assessment exercise.

7.1.3 RQ3: What can bibliometric analysis tell us about transdisciplinary research collaboration trends?

Transdisciplinary research refers to cooperation of scientists representing different societal stakeholders (Hernandez-Aguilar et al., 2020) and is considered to offer benefits when addressing complex social problems like renewable energy, healthcare, or smart and sustainable communities (Hirsch Hadorn et al., 2006; Yarime et al., 2012). Research policymakers around the world routinely call for closer collaboration, closing the gaps between societal stakeholders, and more societal engagement (Drahota et al., 2016; Frerichs et al., 2016). What they are calling for is transdisciplinary research. Despite the regular, frequent, and international calls and policies, there is still very little evidence that such gaps are being closed.

The study presented in chapter 4 addresses this question by quantifying trends in inter-stakeholder collaboration. Stakeholder publications were sourced from the Dimensions database, which relies on the Research Organization Registry (ROR) classification of organisation types. The organisation types used in this study are shown in Table 7.1.

Table 7.1 Stakeholder types

Organisation type	Description
Academia	A university or similar institution involved in providing education and educating/employing researchers.
Industry	A private for-profit corporate entity involved in conducting or sponsoring research.
Nonprofit	A non-profit and non-governmental organisation involved in conducting or funding research.
Government	An organisation that is part of or operated by a national or regional government and that conducts or supports research.

As the vast majority of scholarly research is conducted and published by academic institutions, the study comprised three pairwise analyses:

- academia – industry
- academia – nonprofit
- academia – government

The study showed an increase in the volume of each of the three collaboration combinations over the period 2013–2022. However, the overall volume of academic papers also increased substantially and therefore the collaboration was presented as a change in collaborative research as a proportion of overall academic output over the 10 years. In the case of academia – industry, the publication of collaborative papers has grown by approx. 40% but has dropped by 16% as a share of overall academic output. That means academic output has grown faster without industry collaboration than with it. The corresponding time trend for academia – nonprofit showed the relative collaborative research falling by 2% and the academia – government collaboration growing by 3% relative to the overall academic output. These results therefore did not provide evidence to support the idea that inter-stakeholder research collaboration is increasing in relation to research output as a whole.

Regional and field of research analysis

Given that regional or national laws, policies, and customs vary widely, the chapter 4 study presented the 10-year change as a scatter graph with a dot representing each of the 25 countries with the highest output and the world and EU as benchmarks. For academia – industry, China and Belgium were the only countries to register detectable increases in collaboration relative to overall academic output. In China, a change in intellectual property laws might have stimulated a lot of commercialisation activities by universities. For academia – nonprofits, Spain, the United Kingdom, and Iran showed the largest relative increase in collaboration.

The study then used a similar breakdown of results, this time by field of research for each of the three comparator pairs. The results showed that the share of academia – industry research fell during the 10-year time period for all fields of research except earth sciences, which grew, and mathematical sciences, which remained fairly stable. Scientific fields such as engineering and biological sciences declined somewhat, although the social sciences and humanities showed the steepest declines. The academic – nonprofit analysis showed a more diverse pattern in which the social sciences and humanities deviated in both directions with marked increases in relative collaboration in philosophy & religious studies, history, heritage, & archaeology, and built environment & design and steep declines in the creative arts & writing and the law & legal studies. The sciences also deviated but to a lesser extent. For instance, the mathematical and chemical sciences showed increases in relative collaboration while engineering and the earth sciences declined. The academia – government analysis showed the steepest increases and the steepest declines in relative collaboration in the social sciences and humanities, while the sciences remained relatively stable.

Given the frequent calls for closer collaboration between academia and other research stakeholders, especially in the context of complex societal challenges, policymakers will undoubtedly benefit from progress reports on collaboration trends. This study uses a large-scale bibliometric study of inter-stakeholder collaboration as a unique contribution to the literature on trends in transdisciplinary research. Despite all the calls, the findings presented showed that the academic community has not increased collaboration in relative terms and that transdisciplinary research remains a very small share of overall academic output.

7.1.4 RQ4: How can bibliometric techniques be used to describe the geodiversity of research?

Equity, diversity, and inclusion are being given considerable attention in higher education (Fuentes et al., 2020), aspiring to capture the "full range of human diversity, including race, ethnicity, gender, sexual orientation, age, religious affiliation, disability status, social class, culture, geodiversity, and other identities associated with sociocultural diversity" (American Psychological Association (APA), 2013). The study presented in chapter 5 aims to contribute specifically to our understanding of geodiversity, especially in the context of research related to hunger.

Chapter 5 answers the research question about how bibliometrics can describe geodiversity by analysing almost 60,000 papers related to hunger that were indexed in the Dimensions database between 2016 and 2021. The geodiversity of research looked at geographical focus of papers, the geographical location of their authors, and the relative author position on the papers. To simplify the study, countries were grouped according to the Global Hunger Index (GHI) report published in 2021 that quantifies the level of hunger in each country (von Grebmer et al., 2021).

About a third of the papers analysed mentioned at least one country name in the title or abstract. Counterintuitively, the frequency of country mentions declined as the severity of hunger rose. The highest number of country mentions was 7,281 mentions of countries in the GHI serious category. The serious category comprises 37 countries with a combined population of 2.6 billion people. Therefore, normalising for the combined population there were 2.79 mentions of serious countries per million population. The most severe categories are alarming and extremely alarming, which together comprise 10 countries and a combined population of 232 million people. The 434 mentions of countries in these most severe categories meant 1.87 mentions per million population. The number of mentions per million population rose as the level of hunger dropped until the countries not assessed, which includes the wealthy, Western countries, received 4.00 mentions per million population. The only exception to this trend was the low country category, which received the lowest (1.82) number of mentions per million population.

The unexpected drop in local authorship in countries most affected by hunger was potentially explained by a lack of funding, limited access to higher education, and underdeveloped research infrastructure. Several of the countries most severely affected have or have had civil or military conflict in recent decades. The presence of war in a country is likely to limit the number of available researchers and focus peoples' priorities elsewhere (De Souza, 2017; Mayai, 2020).

The study then analysed author location to examine the extent to which authors of hunger research are located in countries severely affected by hunger. Also counter to expectation, the number of author affiliations declined as the severity of hunger rose. There were 288 author affiliations in countries in the alarming and extremely alarming country categories, which meant 1.24 affiliations per million population. The number of affiliations per million population rose as the level of hunger dropped, until countries not assessed for hunger hosted 29.02 affiliations per million population. The only exception to this trend were the low country categories, where the number of affiliations per million (5.33) fell between that of the serious (4.93) and moderate (6.48) categories.

Mentions-only papers

Comparison of country mentions and author locations enabled analysis of the reasons for authors to mention in a paper a country they are not physically located in. These papers, termed mentions-only papers, were of interest in the context of 'helicopter' or 'parachute' research, in which scientists from wealthy countries drop in to poorer regions to collect samples and then conduct their analyses in the safety and comfort of their own home. Analysis of author location confirmed that the majority of mentions-only papers in each GHI category were published by authors in Western, developed countries.

Affiliations-only papers

The next part of the study looked at papers with authors from countries suffering from hunger where the country in which the authors are based was not mentioned in the paper. These papers were termed affiliations-only papers, and in most (65% - 96%) cases they didn't mention any country at all. A common explanation for papers mentioning no country was that the paper had no regional focus. Many papers published by researchers in India focus on improving agricultural methods which are not region-specific and therefore count as affiliations-only papers. Another common reason for not mentioning the name of the country where the authors are based was that the mention seemed to have been simply overlooked. For instance, manual examination of sample papers revealed a clinical study whose authors were all based in the Democratic Republic of the Congo thanked a local hospital in the acknowledgements section of the paper.

Equality of partnerships

A section of the study examined the author location and position of author affiliations on the paper. Author affiliations from alarming and extremely alarming countries accounted for 0.35% of all affiliations, while the population of these countries accounts for 2.9% of the world's population. Therefore, authors from countries in these categories are underrepresented in hunger research. The share declined even further for authors in first (0.23%) and last author (0.17%) positions. The study confirmed the low level of lead author positions among local authors on published hunger research that focused on countries most affected by hunger. This finding has been linked to complaints of 'helicopter' research in which scientists from wealthy countries are said to 'drop in' to poorer regions to collect samples and then conduct their analyses once they are safely home (Heymann et al., 2016; Minasny & Fiantis, 2018; North et al., 2020). Rather than including scientists in the country under study among the resulting paper's co-author list, visiting scholars overlook the contribution of local academics or relegate them to a mention in the acknowledgements section of the paper.

Certainly, the case against unethical research practices warrants further investigation and intervention by the global academic community. However, it may not be easy to identify malpractice simply through large-scale bibliometric analysis. The UN call to action in 2015 was to the whole world and that means that wealthy countries should help developing nations tackle problems described in the SDGs, including poverty and hunger. In an ideal scenario, scholars from well-funded countries with established research infrastructure will work closely with academics in developing areas, who are likely to have a deeper understanding of local issues and more easily transcend cultural and linguistic barriers.

7.1.5 RQ5: How can bibliometric assessment cause inadvertent behavioural change in the scientific community?

In 2012, the Indonesian government modified its research policy with the intention of incentivising academics to increase their publications indexed in the international bibliometric databases. The motivation was to catch up with regional peer countries, especially Malaysia and Singapore. The pressure on Indonesian academics to increase publishing activity is seen in many other areas of the world (Abramo & D'Angelo, 2014). Under new ministry guidelines, Indonesian scientists' pay increases and academic promotions were linked to the number of their publications in 'accredited' journals, i.e. those indexed in Scopus or Web of Science. This provided a strong incentive for Indonesian academics to prioritise publishing in Scopus or Web of Science-indexed venues at a time when both databases had just launched conference proceedings expansion programmes.

The results of the study described in Chapter 6 show a striking and unique national publication pattern of Indonesian academics. The relative share of Indonesian conference papers indexed in the Web of Science jumped from a little over 20% up to 80% of overall Indonesian publications in just six years. That means that four out of every five published papers with an Indonesian author in 2018 were conference papers. Analysis of Scopus data confirmed this trend, and although the increase in Scopus was more modest, conference papers still accounted for three out of every five published papers by 2018.

World and ASEAN regional data showed that the pattern of changing publishing behaviour was unique to Indonesia and not representative of norms seen elsewhere. Conference proceedings output also increased in absolute terms globally but accounted for a fairly stable proportion of overall publication output over a 20-year period.

Chapter 6 includes bibliometric analyses to further analyse the surprising Indonesian result. Web of Science and Scopus were competing at the time to offer more complete coverage of high-quality science by expanding their coverage of academic conferences. The Indonesian conference papers were in conference series that had already been indexed in the databases before the coverage expansion, which suggests that the spike in Indonesian papers were published at established conferences. In 2014, Indonesia ranked as the country with the 41st most conference papers in the *Journal of Physics Conference Series*, while by 2018 Indonesia had risen to 1st position. Indonesia had become the most prominent author country among papers presented at the conference. A similar pattern could be detected for other conference series in both Web of Science and Scopus.

The study goes on to note that there was a steep increase in the number of academic conferences physically hosted in Indonesia whose proceedings were then indexed in Scopus and the Web of Science. The share of AIP conferences hosted in Indonesia almost tripled between 2014 and 2018 and other conference series saw similar or even greater increases in the share of their conferences hosted on Indonesian soil. The resources and logistics must surely make it easier for academics to attend and present at conferences in their own country instead of travelling to conferences held outside the country. The increased emphasis on hosting conferences in Indonesia might therefore have been a contributing factor in the spike in indexed conference papers from Indonesian academics.

There is a perception within academia that conference papers are quicker and easier to publish than other document types including book chapters and journal articles (Bar-Ilan, 2010; González-Albo & Bordons, 2011). Therefore, it follows that academics can increase their paper count by switching their attention to presenting papers at conferences at the expense of submitting manuscripts to journals. This practice was confirmed anecdotally by one Indonesian academic who organised local academic conferences, published the proceedings, and submitted them to Scopus and Web of Science for indexing. The timing of the release of ministry guidelines suggests a causal relationship with the selective shift towards conference publishing that was unique to Indonesia.

The study in Chapter 6 illustrates the potential consequences of general research policies and publishing guidelines. A more specific policy could have included a different weighting for the various document types. In this case, bibliometrics has been allowed to take the lead and be a deciding factor in the important matter of academic promotion. An increased element of peer judgement would surely detect and rectify this anomaly.

7.2 Implications of the findings

The studies presented here have illustrated the benefits of using large-scale, quantitative bibliometric analyses described in the dissertation objectives (section 1.5.2), such as the ability to compare data quality between bibliometric databases and compare different approaches to defining new subject areas (chapters 2 and 3). The power of bibliometrics extends beyond academia and facilitated a study of transdisciplinary research collaboration (chapter 4). In chapter 5, the analysis of data on millions of publications enabled a description of the geodiversity of research and in chapter 6, bibliometric analysis illustrated a national-level behavioural change.

In each of the cases mentioned above, there was also an opportunity to discuss the limitations of bibliometrics alluded to in the general introduction of the dissertation (section 1.1). This includes data quality issues and discrepancies between data providers (chapter 2), 'black box' methods and new technologies such as artificial intelligence (chapter 3), differences between research stakeholder types in priorities regarding research outputs (chapter 4), ambiguous recognition of contributors and the potential for unfair practice concerning collaboration between academics from privileged and unprivileged backgrounds (chapter 5), and the assumptions made on academics' publication practice (chapter 6).

These implications can be grouped into three broad categories: those concerning how bibliometric assessment is in fact shaping the research system, the need to find a balance between bibliometrics and expert judgement, and the role of the bibliometric community in improving data, methods, and transparency in the pursuit of more trust among research stakeholders.

7.2.1 Bibliometric assessment is shaping the research system

The use of bibliometrics is shaping the research system by defining what can be measured and influencing the behaviour of those being evaluated using bibliometric assessments. We have seen that scientific communities will respond to bibliometric assessments by modifying their behaviour to improve their performance. In Indonesia (chapter 6), the national research policy offered financial incentives for higher publication counts in an attempt to boost national output with respect to other countries. The shift of the Indonesian academic community from publishing journal articles to publishing 'easier' conference papers is evidence of a community responding to the way it was being assessed by changing its publication behaviour.

We have also seen discussions about whether contributors to studies are being duly assigned co-author status or merely mentioned in the acknowledgements section of a paper (chapter 5). In bibliometric studies, coauthors usually receive credit for their contribution to a paper, while those mentioned in the acknowledgements section do not. The order of co-authors has also come under scrutiny, with certain author positions associated with more prestige than others. Choices are therefore being made by academics involved in scientific study that are at least partially influenced by bibliometric assessments. Questions have been raised about how those decisions are made and who is making them, especially in the context of international collaboration between researchers from wealthy regions and academics based in low- and middle-income countries.

7.2.2 Balancing bibliometrics with expert judgement

The findings presented in this dissertation demonstrate that bibliometric analyses can be helpful in policy development and research evaluation because they reveal large-scale trends, networks, and patterns in publication dynamics relatively quickly and cheaply. However, bibliometric analyses also lack the sensitivity to fully explain the observed trends and patterns. In this dissertation, bibliometric analysis is used to set the scene by revealing intriguing trends or patterns in scientific publications that would have been difficult to spot without the use of quantitative analyses. These include the wide discrepancy between approaches to defining research related to sustainability (chapter 3), the shift in Indonesian publishing patterns (chapter 6) and transdisciplinary collaboration (chapter 4).

However, in each of the studies, some form of manual analysis was used to explain the results. Sometimes this manual step confirmed what we intuitively expected from the numbers and added an extra layer of confidence in our conclusion. At other times the manual analysis revealed surprises that were not evident from simply looking at the macro-level patterns. It is important therefore for users of bibliometric reports to accept that the results represent only the beginning of the story. Findings should be scrutinised carefully, and consumers of bibliometric reports should avoid jumping to conclusions that are not directly proven by the analysis. The fine combination of quantitative analysis with expert judgement is the key to conducting good quality evaluation and policymaking. This coincides with recommendations presented in several community-led guidelines on improving the use of bibliometric analysis in research assessment, e.g., the Leiden Manifesto (Hicks et al., 2015). The true story only really becomes clear once one gets behind the numbers.

Shouldn't we then go one step further and remove the numbers entirely? This question has been debated for a long time, and it is still the subject of intense discussion, especially in the context of research assessment. For example, in a recent letter to the editors of the journal Scientometrics, Torres-Salinas et al. (2023) complain that some sectors of the scientific community are leaning towards a state of 'bibliometric denialism'. Torres-Salinas et al. use this term to highlight the practice of deliberately denying a corpus of evidence on the value of bibliometric analysis in the same way flat-earth society supporters deny the Copernican view that the earth revolves around the sun (Boden & Epstein, 2011). The letter goes on to cite recent changes in evaluation systems in Germany and Spain that have entirely removed the quantitative component of research assessment as 'detrimental' to publication behaviour. The stark language used in this letter belies a frustration in parts of the research community at the reactive ban on bibliometric methods and consequent reluctance to incorporate any quantitative analysis into routine research assessment.

Some supporters of the views described by Torres-Salinas et al. indeed promote the wholesale removal of quantitative indicators from research assessment, but as pointed out by Rushforth (2023), this has not become the prevailing view. In a measured response to Torres-Salinas et al. (2023), Rushforth suggests the claim of denialism is too strong, even alarmist, and that the research community is generally adopting a position of 'responsible research assessment'. The idea of responsible research assessment is to treat bibliometrics as one of the available tools to 'support but not replace peer review' (Rushforth & Hammarfelt, 2023). This seems to be a sensible approach, combining the benefits of both quantitative and qualitative elements. Moreover, it is in line with the approach taken in this dissertation, where various studies have begun with a large-scale bibliometric analysis and have then applied manual examination to more fully explain the findings.

7.2.3 Importance of community-based collaborative initiatives

Although the appropriate, context-specific use of bibliometric indicators has been urged since their inception (Garfield, 1972), once the data is accessible by millions of academics, it is natural for inquisitive scientists to try out new ideas on their own. Indeed, many researchers justifiably consider themselves bibliometricians and have developed opinions on research evaluation methods. Members of the research community periodically publish calls for communal progress through initiatives such as the Declaration On Research Assessment that was launched by researchers calling for improvements to bibliometric assessment techniques (*DORA*, 2023) and the Leiden Manifesto in which professional bibliometricians describe a set of principles to guide research evaluation (Hicks et al., 2015). More recently the European Commission launched the Agreement on Reforming Research Assessment (CoARA, 2022a), which aims to transform ideas like those presented in DORA into real actions and which has attracted support from research organisations around the world. The Coalition for Advancing Research Assessment (CoARA, 2022b) has been formed to enable the reform described in the Agreement on Reforming Research Assessment. CoARA proposes a move towards a research assessment system based primarily on peer review and supported by responsible use of bibliometrics, while abandoning some of the most used and most controversial indicators and ranking systems. Authoritative, community-led declarations such as DORA, the Leiden Manifesto, and the CoARA agreement are needed to address some of the bigger challenges faced in bibliometrics. Indeed, they channel peripheral and diverse conversations into more centralised, aggregate opinions and invite the wider community to support them. These collective statements then begin shaping popular opinion into community consensus. Organised initiatives supported by large sections of the community have a greater likelihood of commanding attention and of making real improvements than scattered opinions.

As the bibliometric community develops, there are opportunities for new initiatives that address other challenges in bibliometrics. Specifically, data quality and curation should be more reliable to build greater trust in the numbers and confidence in conclusions drawn from bibliometric reports. Therefore, the entire research community has a vested interest in improving the research data infrastructure and several community-led initiatives have already demonstrated success. An example is ORCID, the open, unique identifier for academic authors that enables bibliometricians to identify research from individual authors across multiple data sources. Another is the Digital Object Identifier (DOI) used by bibliometricians to identify research publications. A third example is the more recent, community-led Research Organization Registry (ROR) which is the only openly available organisation identifier, and which would resolve disambiguation discrepancies if all data producers and bibliometricians would adopt it.

All data producers face decisions when conducting affiliation disambiguation, which has led to different databases using different systems, universities having to regularly submit updated information to database producers, and some bibliometricians conducting their own disambiguation process. Some data source producers have developed their own identifiers such as Scopus Afid and Web of Science Organisation Enhanced, but these identifiers are linked only to the database they were created for. If multiple affiliation disambiguation systems exist, bibliometric reports will be difficult to compare because each will be dependent on the system used, as we have seen in chapters 2 and 6 of this dissertation. Research evaluators and policy makers should keep this front of mind when interpreting bibliometric reports. We should progressively embrace the database-agnostic ROR for organizations to build trust in bibliometric analyses of universities and research institutions.

The community could help improve data infrastructure in other areas where bibliometric databases take different approaches. There are opportunities for the development of community-led, unified approaches to classifying published research by document types, recognising acknowledgements of contributors and funders, and categorising publications by subject area. Each of these variables are currently specific to the data producer and make it difficult to make comparisons across databases.

Decisions about the organization and classification of research data are often made by experts, and expert judgement is used to design the methods that produce the findings generated by bibliometric reports. As discussed throughout this dissertation, it is important to include an element of expert judgement in research policymaking and assessment and therefore natural that experts are involved in data organization and development of methods. However, the details of expert contribution are rarely made publicly available and therefore contribute to the 'black box' of the data or method concerned. Another component of the methodological 'black box' is the use of artificial intelligence. As machine learning and natural language processing methods begin to form part of new approaches to challenges in bibliometrics, the precise algorithms are not always made public. We have seen an example of this in chapter 3 about different approaches to the categorisation of research related to the SDGs, where expert judgement creates differences early in a multi-step method that are subsequently compounded through obscure AI algorithms that lead to discrepancies between approaches.

As the research community gains more experience using and investigating new methods and approaches, it is natural and encouraged that people form opinions and debate them in open forums. These debates will ideally serve as mechanisms that aggregate scattered opinions resulting in community-wide consensus. That would lead to community-based collaborative initiatives to open up 'black box' data and methods making them more transparent. Transparent methods that can be replicated, challenged and improved will lead to greater reliability and trust. This will ensure that policymakers and other stakeholders have sufficient knowledge to make informed choices about data sources or methodological approaches.

7.3 Vision for future

The research presented in this dissertation has examined several aspects of bibliometrics in the context of research evaluation and research policy and has opened several new avenues for further study. These avenues can be broadly considered in two groups, the first focuses on how data quality and curation can improve confidence in the use of bibliometrics, while the second centres more on how bibliometrics shapes research practice.

The confidence with which policymakers can rely on bibliometric studies depends on data quality, where there are multiple areas for improvement. The discrepancies between databases in their treatment of university names, and the different approaches to categorising research related to the sustainable development goals (SDGs) suggest there are substantial benefits to expanding research and improving data curation practices in this field.

At the same time, it needs to be recognised that bibliometrics is not just a 'neutral' tool for measuring what is happening in the research system. The use of bibliometrics also shapes research practices, sometimes in positive ways and sometimes in ways that are more questionable. A deeper awareness of the ways in which bibliometrics shapes research practices is essential to ensure that the use of bibliometrics contributes to a more effective and more fair research system. Below are some suggested avenues for future prospects.

7.3.1 Data quality should be improved through persistent identifiers

Data quality and openness is a common theme frequently mentioned in the scientific community. Improvements in data structure, consistency and quality will increase confidence in the use of bibliometric analyses. An opportunity for improvement identified in this dissertation is the increased use of persistent identifiers (PIDs) in bibliometric platforms. Persistent identifiers are an important element of open research data infrastructure and enable stakeholders to identify research elements including publications, authors, institutions, and articles (Chodacki & Carpenter, 2024; Juty et al., 2020). PID-graphs including OpenAlex (OpenAlex, 2023), DataCite (Dudek et al., 2019), and OpenAIRE (Manghi et al., 2019) already link research metadata elements and the relationships between them, which facilitates an increasingly connected global research infrastructure.

This dissertation illustrated the impact of discrepancies between databases in their management of author affiliation unification. That papers may or may not be linked to the right university is a problem. Incorporating a single open PID such as the Research Organization Registry (ROR) into all the major databases would help identify research organisations in a similar way to articles and researchers. This would allow studies to rely on a consistent, open, and transparent identifier that remains the same across different data sources. Dimensions and OpenAlex already both use ROR as the identifier for research organisations while Scopus and Web of Science still use their proprietary systems. Adoption of a single open PID by all databases would make it easier for users to switch between databases and reduce the administrative burden on universities that currently maintain their affiliation variant list with multiple database providers. The accuracy of bibliometric databases would improve and discrepancies between databases in university output would be reduced. The major university ranking systems would benefit from greater transparency in their methodologies and their rankings would be less dependent on the choice of data provider.

7.3.2 Improved data curation will improve consumer confidence in bibliometrics

Comparisons of different bibliometric data sources show that large discrepancies exist between them. Some are related to coverage, but other substantial discrepancies are caused by methods and procedures of data curation. Many of the data curation methods of data source providers are neither open, nor transparent, which means the research community is restricted in its ability to conduct detailed studies of the underlying causes of discrepancies. We should unpack the black box of data curation in areas including defining and maintaining subject categorisations, assigning or changing document types, and disambiguating organisation names and types. These are typically multi-step processes which means any methodological differences between approaches in the early steps of the process will lead to overall compounded discrepancies between the approaches. Examination of the individual steps involved in these processes by different data producers would enable a deeper understanding of how bibliometric data is handled and encourage the community to participate in improvements.

Pursuing these questions will lead data curation methods to develop over the coming years and experimentation is needed to test, fail, disagree, and build again. For example, defining the relationship between SDGs and published research papers presents a different kind of problem from resolving author name unification issues or disambiguating organisation names. It is likely we will never find universal agreement on the definition of research that is related to the SDGs because the SDGs themselves are loosely defined and mean different things to different people. Ideally, data producers and groups designing methodological approaches to complex questions should be encouraged to develop and test new approaches in an open and transparent manner. Data producers should work closely with the bibliometric community to harness its expertise. Working together towards common goals will accelerate progress towards better solutions and increased consumer confidence in bibliometric methods.

7.3.3 Inclusive contributor recognition

One of the implications raised here is the potential for inequalities in research data especially of disadvantaged scientific communities. It is in our common interest to improve inclusivity, equality, and diversity by addressing elements of the scholarly publishing and evaluation system that favour academics who already find themselves in privileged positions. For instance, it would be preferable if publishers grouped all contributors to research papers together in one section with a brief description of their contribution type. This single section should replace the author list, the acknowledgements section, and any current description of author contributions. Every contributor would be named with a brief description of the contribution which could be based on the CRediT Contributor Roles Taxonomy or an expanded version of it. The film industry uses this method in the credits that scroll up the screen at the end of the film. The same system could be applied to scholarly publishing so that no-one is excluded and bibliometricians can conduct finer level analyses of individual contribution. This way, bibliometrics would help to shape research assessment in a way that is more inclusive, transparent, and fair.

7.3.4 The influence of bibliometrics on academics' behaviour

The link between research policy and academic publishing behaviour has been established in several countries including Australia, Denmark, Indonesia, and the United Kingdom. Australia's national research policy began offering financial incentives to universities in the 1990s in return for scholarly publications in international journals (Butler, 2003b). There was then an interesting debate about whether putting a monetary value on publications was associated with behavioural changes in the Australian academic community (Butler, 2003b, 2003a, 2017; van den Besselaar et al., 2017; van den Besselaar & Sandström, 2017). The UK national research assessment exercise (RAE), later the research excellence framework (REF) has been associated with related discussion on the shift of publication behaviour among academics at UK universities towards publication types most likely to result in favourable citation impact scores (Adams, 2009; Marques et al., 2017).

These debates along with others in Denmark (Aagaard & Schneider, 2016) and Indonesia (chapter 6 of this dissertation) aimed first to establish whether the assessment had caused a change in behaviour, and then to discuss whether any change in behaviour was intended or unintended, and whether it was a positive or a negative change. Goodhart's Law tells us that a metric ceases to be an appropriate metric once it becomes a target (Goodhart, 1975). The extrapolation of that lesson to research policy is interpreted as a need to find ways to evaluate scholarly output and impact while ensuring that the community being assessed does not undergo any negative change in behaviour driven by opportunistic ways to improve their performance in the assessment. This is easier said than done because as we have seen, the bibliometric data structure, curation, indicators, and evaluation methods are already intricately linked to evaluation methods. Prominent bibliometricians representing 24 of Europe's top universities through the League of European Research Universities recently published a position paper on appropriate use of metrics in research evaluation (LERU, 2024). The LERU paper offers several recommendations to policymakers and other research stakeholders on using and interpreting bibliometric reports in conjunction with expert opinion. To reduce the risk of negative effects of bibliometrics on researcher behaviour, many of the recommendations centre on 'measuring things differently', which involves the development of next-generation indicators, new evaluation techniques, and careful balancing of bibliometric analysis with expert opinion. LERU also encourages the establishment of clear communication channels between those designing and acting on the evaluation techniques and those being evaluated.

7.3.5 The promise of artificial intelligence

We have begun to see the application of artificial intelligence in bibliometric data curation and analysis. There are opportunities for application of automation and AI prediction models throughout the data curation, publication, and impact assessment process. We have seen the application of machine learning algorithms in the categorisation of new subject fields (Jayabalasingham et al., 2019; Rivest et al., 2021; Wastl et al., 2020), which although in early stages of development, could offer policymakers nuanced options for identifying SDG-related research. The publication process itself holds huge potential for automation and AI tools. Kousha & Thelwall, (2024) have published a comprehensive review of software applications of new technologies that span article writing, journal recommendation, manuscript submission, plagiarism detection, reviewer selection, and even propose investigating potential limited uses in peer review.

Machine learning article citation prediction models hold promise to provide early post-publication insight to evaluators of research impact (Beranová et al., 2022). The ability to predict citation impact holds obvious advantages over the traditional time lag associated with waiting several years to detect citations. Alternative metrics that use social media responses to publications may shorten that gap, but AI citation prediction models should be thoroughly investigated. This would ideally lead to a scenario in which policymakers have an immediate idea of how well certain research is likely to be received and research funders gain a preview of the potential impact of research proposals. Decisions could then be taken about support at an earlier stage, thereby leading to greater concentration on high priority projects while saving resources on less promising proposals.

While artificial intelligence appears to hold huge promise, it also comes with uncertainty and should be treated with caution. One aspect of future prediction models is that they are trained on data related to things that happened in the past. Changes to the world mean the future might not behave as they did previously, which can limit the accuracy of AI predictions. Research in this area is eagerly anticipated.

References

Aagaard, K., & Schneider, J. W. (2016). Research funding and national academic performance: Examination of a Danish success story. *Science and Public Policy*, *43*(4), 518–531. https://doi.org/10.1093/scipol/scv058

Aboukhalil, R. (2014). The rising trend in authorship. *The Winnower*. https://doi.org/10.15200/winn.141832.26907

Abramo, G. (2024). The forced battle between peer-review and scientometric research assessment: Why the CoARA initiative is unsound. *Research Evaluation*, rvae021. https://doi.org/10.1093/reseval/rvae021

Abramo, G., & D'Angelo, C. A. (2014). How do you define and measure research productivity? *Scientometrics*, *101*(2), 1129–1144. https://doi.org/10.1007/s11192-014-1269-8

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Allocative efficiency in public research funding: Can bibliometrics help? *Research Policy*, *38*(1), 206–215. https://doi.org/10.1016/j.respol.2008.11.001

Adams. (1998). Benchmarking international research. *Nature*, *396*(6712), 615–618. https://doi.org/10.1038/25219

Adams, Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, *34*(3), 259–285. https://doi.org/10.1016/j.respol.2005.01.014

Adams, J. (2009). The use of bibliometrics to measure research quality in UK higher education institutions. *Archivum Immunologiae et Therapiae Experimentalis*, *57*(1), 19–32. https://doi.org/10.1007/s00005-009-0003-3

Agarwal, R., & Ohyama, A. (2012). Industry or Academia, Basic or Applied? Career Choices and Earnings Trajectories of Scientists. *Management Science*, *59*(4), 950–970. https://doi.org/10.1287/mnsc.1120.1582

American Psychological Association (APA). (2013). *APA guidelines for the undergraduate psychology major: Version 2.0*.

Amugune, B. K., & Otieno-Omutoko, L. (2019). *An African Perspective of Benefits in Social Science Research BT - Social Science Research Ethics in Africa* (N. Nortjé, R. Visagie, & J. S. Wessels, Eds.; pp. 47–56). Springer International Publishing. https://doi.org/10.1007/978-3-030-15402-8_4

Aramesh, K. (2019). *The Main Challenges and Their Ethical Nature BT - An Ethical Framework for Global Governance for Health Research* (K. Aramesh, Ed.; pp. 43–73). Springer International Publishing. https://doi.org/10.1007/978-3-030-31132-2_3

Archambault, E., Campbell, D., Gingras, Y., & Lariviere, V. (2009). Comparing of Science Bibliometric Statistics Obtained From the Web and Scopus. *Journal of the American Society for Information Science and Technology*, *60*(7), 1320–1326. https://doi.org/10.1002/asi.21062

Arias, A., Watson, S. J., Asogun, D., Tobin, E. A., Lu, J., Phan, M. V. T., Jah, U., Wadoum, R. E. G., Meredith, L., Thorne, L., Caddy, S., Tarawalie, A., Langat, P., Dudas, G., Faria, N. R., Dellicour, S., Kamara, A., Kargbo, B., Kamara, B. O., … Cotten, M. (2016). Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. *Virus Evolution*, *2*(1), vew016. https://doi.org/10.1093/ve/vew016

Armitage, C., & Kaindl, M. (2020). Getting with the program. *Nature Index*, *09 Decembe*. https://www.nature.com/articles/d41586-020-03416-9

Armitage, C., Lorenz, M., & Mikki, S. (2020a). Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results? *Quantitative Science Studies*, *1*(3), 1092–1108. https://doi.org/10.1162/qss_a_00071

Armitage, C., Lorenz, M., & Mikki, S. (2020b). *Replication data for: Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results?* (U. of Bergen, Ed.; V1 ed.). DataverseNO. https://doi.org/10.18710/98CMDR

Arnkil, R., Järvensivu, A., Koski, P., & Piirainen, T. (2010). *Exploring Quadruple Helix Outlining user-oriented innovation models* (85; 2010). https://trepo.tuni.fi/bitstream/handle/10024/65758/978-951-44-8209-0.pdf

ASEAN. (2019). *Association of Southeast Asian Nations Website*. https://asean.org

Asiamah, G. B., Awal, M. S., & MacLean, L. M. (2021). Collaboration for Designing, Conceptualizing, and (Possibly) Decolonizing Research in African Politics. *PS: Political Science & Politics*, *54*(3), 549–553. https://doi.org/10.1017/S1049096521000226

Association of Dutch Universities. (2019). *SDG-Dashboard: Impact of the Dutch Universities*. https://vsnu.nl/en_GB/sdg-dashboard-english.html

Auranen, O., & Nieminen, M. (2010). University research funding and publication performance—An international comparison. *Research Policy*, *39*(6), 822–834. https://doi.org/10.1016/j.respol.2010.03.003

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386. https://doi.org/10.1162/qss_a_00019

Bador, P., & Lafouge, T. (2005). Rédaction des adresses sur les publications: Un manque de rigueur défavorable aux universités françaises dans les classements internationaux. *La Presse Médicale*, *34*(9), 633–636. https://doi.org/10.1016/S0755-4982(05)84000-X

Bajoria, J. (2011). *Improving UN Responses to Humanitarian Crises*. UN Chronicle. https://www.un.org/en/chronicle/article/improving-un-responses-humanitarian-crises

Balvanera, P., Calderón-Contreras, R., Castro, A. J., Felipe-Lucia, M. R., Geijzendorffer, I. R., Jacobs, S., Martín-López, B., Arbieu, U., Speranza, C. I., Locatelli, B., Harguindeguy, N. P., Mercado, I. R., Spierenburg, M. J., Vallet, A., Lynes, L., & Gillson, L. (2017). Interconnected place-based social–

ecological research can inform global sustainability. *Current Opinion in Environmental Sustainability*, *29*, 1–7. https://doi.org/10.1016/j.cosust.2017.09.005

Banal-Estañol, A., Jofre-Bonet, M., & Lawson, C. (2015). The double-edged sword of industry collaboration: Evidence from engineering academics in the UK. *Research Policy*, *44*(6), 1160–1175. https://doi.org/10.1016/j.respol.2015.02.006

Bar-Ilan, J. (2010). Web of Science with the Conference Proceedings Citation Indexes: The case of computer science. *Scientometrics*, *83*(3), 809–824. https://doi.org/10.1007/s11192-009-0145-4

Baudoin, L., Akiki, V., Magnan, A., & Devos, P. (2018). Production scientifique des CHU-CHR en 2006–2015 : évolutions et positionnement national. *La Presse Médicale*, *47*(11, Part 1), e175–e186. https://doi.org/10.1016/j.lpm.2018.06.016

Beaudry, C., & Allaoui, S. (2012). Impact of public and private research funding on scientific production: The case of nanotechnology. *Research Policy*, *41*(9), 1589–1606. https://doi.org/10.1016/j.respol.2012.03.022

Beranová, L., Joachimiak, M. P., Kliegr, T., Rabby, G., & Sklenák, V. (2022). Why was this cited? Explainable machine learning applied to COVID-19 research literature. *Scientometrics*, *127*(5), 2313–2349. https://doi.org/10.1007/s11192-022-04314-9

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F., & Secret, A. (1994). The world-wide web. *Communications of the ACM*, *37*(8), 76–82. https://dl.acm.org/doi/pdf/10.1145/179606.179671

Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C. A., & Peracchi, F. (2015). Bibliometric evaluation vs. informed peer review: Evidence from Italy. *Research Policy*, *44*(2), 451–466. https://doi.org/10.1016/j.respol.2014.08.004

Bickley, M. S., Kousha, K., & Thelwall, M. (2020). Can the impact of grey literature be assessed? An investigation of UK government publications cited by articles and books. *Scientometrics*, *125*(2), 1425–1444. https://doi.org/10.1007/s11192-020-03628-w

Bikard, M., Vakili, K., & Teodoridis, F. (2018). When Collaboration Bridges Institutions: The Impact of University–Industry Collaboration on Academic Productivity. *Organization Science*, *30*(2), 426–445. https://doi.org/10.1287/orsc.2018.1235

Billaut, J.-C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics*, *84*(1), 237–263. https://doi.org/10.1007/s11192-009-0115-x

Billick, I., & Price, M. V. (2010). *The ecology of place: contributions of place-based research to ecological understanding*. University of Chicago Press. https://books.google.ae/books?hl=en&lr&id=F-UZVzTh9O4C

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, *1*(1), 363–376. https://doi.org/10.1162/qss_a_00018

Bjerregaard, T. (2010). Industry and academia in convergence: Micro-institutional dimensions of R&D collaboration. *Technovation*, *30*(2), 100 – 108. https://doi.org/10.1016/j.technovation.2009.11.002

Blasco, N., Brusca, I., & Labrador, M. (2021). Drivers for Universities' Contribution to the Sustainable Development Goals: An Analysis of Spanish Public Universities. *Sustainability*, *13*(1). https://doi.org/10.3390/su13010089

Bockarie, M., Machingaidze, S., Nyirenda, T., Olesen, O. F., & Makanga, M. (2018). Parasitic and parachute research in global health. *The Lancet Global Health*, *6*(9), e964. https://doi.org/10.1016/S2214-109X(18)30342-5

Boden, R., & Epstein, D. (2011). A Flat Earth Society? Imagining Academic Freedom. *The Sociological Review*, *59*(3), 476–495. https://doi.org/10.1111/j.1467-954X.2011.02014.x

Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. *Journal of Informetrics*, *7*(2), 286–291. https://doi.org/10.1016/j.joi.2012.12.003

Boudry, C., & Chartron, G. (2017). Availability of digital object identifiers in publications archived by PubMed. *Scientometrics*, *110*(3), 1453–1469. https://doi.org/10.1007/s11192-016-2225-6

Brankovic, J., Ringel, L., & Werron, T. (2018). How rankings produce competition: The case of global university rankings. *Zeitschrift Fur Soziologie*, *47*(4), 270–288. https://doi.org/10.1515/zfsoz-2018-0118

Brewer, G. D. (1999). The Challenges of Interdisciplinarity. *Policy Sciences*, *32*(4), 327–337. http://www.jstor.org/stable/4532473

Broekel, T. (2015). Do Cooperative Research and Development (R&D) Subsidies Stimulate Regional Innovation Efficiency? Evidence from Germany. *Regional Studies*, *49*(7), 1087–1110. https://doi.org/10.1080/00343404.2013.812781

Brundtland, G. H. (1987). *Report of the World Commission on Environment and Development: Our Common Future*.

Boden, R., & Epstein, D. (2011). A Flat Earth Society? Imagining Academic Freedom. *The Sociological Review*, *59*(3), 476–495. https://doi.org/10.1111/j.1467-954X.2011.02014.x

Butler, L. (2003a). Explaining Australia's increased share of ISI publications - the effects of a funding formula based on publication counts. *RESEARCH POLICY*, *32*(1), 143–155. https://doi.org/10.1016/S0048-7333(02)00007-0

Butler, L. (2003b). Modifying publication practices in response to funding formulas. *RESEARCH EVALUATION*, *12*(1), 39–46. https://doi.org/10.3152/147154403781776780

Butler, L. (2017). Response to van den Besselaar et al.: What happens when the Australian context is misunderstood. *JOURNAL OF INFORMETRICS*, *11*(3), 919–922. https://doi.org/10.1016/j.joi.2017.05.017

Bylund, E., Khafif, Z., & Berghoff, R. (2023). Linguistic and geographic diversity in research on second language acquisition and multilingualism: An analysis of selected journals. *Applied Linguistics*, amad022. https://doi.org/10.1093/applin/amad022

Calero-Medina, C., Noyons, E., Visser, M., & De Bruin, R. (2020). Delineating Organizations at CWTS---A Story of Many Pathways. In C. Daraio & W. Glänzel (Eds.), *Evaluative Informetrics: The Art of Metrics-Based Research Assessment : Festschrift in Honour of Henk F. Moed* (pp. 163–177). Springer International Publishing. https://doi.org/10.1007/978-3-030-47665-6_7

Cambridge University. (2024). Research Policy | Research Strategy Office. Research Policy. https://www.research-strategy.admin.cam.ac.uk/research-policy

Campbell, J. O., Gupta, A. K., Lu, A., Lim, Y. F., Mishra, N., Hewitt, J. N., Ovenden, C. D., Kovoor, J. G., Bacchi, S., Trochsler, M., & Wells, A. (2023). The geographic diversity of authorship in leading general surgery journals; A study of 24,332 authors. *The Surgeon*. https://doi.org/10.1016/j.surge.2023.08.004

Carayannis, Del Giudice, M., & Rosaria Della Peruta, M. (2014). Managing the intellectual capital within government-university-industry R&D partnerships. *Journal of Intellectual Capital*, *15*(4), 611–630. https://doi.org/10.1108/JIC-07-2014-0080

Carayannis, Acikdilli, G., & Ziemnowicz, C. (2020). Creative Destruction in International Trade: Insights from the Quadruple and Quintuple Innovation Helix Models. *Journal of the Knowledge Economy*, *11*(4), 1489–1508. https://doi.org/10.1007/s13132-019-00599-z

Carayannis, & Rakhmatullin, R. (2014). The Quadruple/Quintuple Innovation Helixes and Smart Specialisation Strategies for Sustainable and Inclusive Growth in Europe and Beyond. *Journal of the Knowledge Economy*, *5*(2), 212–239. https://doi.org/10.1007/s13132-014-0185-8

Castro Torres, A. F., & Alburez-Gutierrez, D. (2022). North and South: Naming practices and the hidden dimension of global disparities in knowledge production. *Proceedings of the National Academy of Sciences*, *119*(10), e2119373119. https://doi.org/10.1073/pnas.2119373119

Centre for Science & Technology Studies Leiden University. (2020). *Indicators*. https://www.leidenranking.com/information/indicators

Centre for Science and Technology Studies. (2022). *Leiden Ranking Indicators*. Leiden Ranking Indicators. https://www.leidenranking.com/information/indicators#collaboration-indicators

Chen, D., & Dahlman, C. (2005). *The Knowledge Economy, the KAM Methodology and World Bank Operations*. http://documents.worldbank.org/curated/en/695211468153873436/The-knowledge-economy-the-KAM-methodology-and-World-Bank-operations

Chodacki, J., & Carpenter, T. (2024). Desirable Characteristics of Persistent Identifiers. *Upstream*. https://doi.org/10.54900/c3hdq-0ev76

Clarivate. (2020a). *Data change FAQs*. https://support.clarivate.com/ScientificandAcademicResearch/s/datachanges?language=en_US

Clarivate. (2020b). *Organization name unification procedures: Unifying address variants*. http://ips.clarivate.com/m/pdfs/Clarivate-NameAddressUnificationProcess.pdf

Clarivate. (2022). *Web of Science Journal Evaluation Process and Selection Criteria*. The Web of Science Core Collection Selection Process - 24 Quality Criteria. https://clarivate.com/webofsciencegroup/journal-evaluation-process-and-selection-criteria/

CoARA. (2022a). *Agreement on Reforming Research Assessment*. The Agreement. https://coara.eu/app/uploads/2022/09/2022_07_19_rra_agreement_final.pdf

CoARA. (2022b). *Coalition for Advancing Research Assessment*. The Agreement. https://coara.eu

Cole, F. J., & Eales, N. B. (1917). The History of Comparative Anatomy: Part I.—A Statistical Analysis of the Literature. *Science Progress (1916-1919)*, *11*(44), 578–596. http://www.jstor.org/stable/43426882

Collins, S. (2022). *Introducing our new Global Equitable Membership (GEM) program*. Crossref Blog. https://www.crossref.org/blog/introducing-our-new-global-equitable-membership-gem-program/

Confraria, H., Noyons, E., & Ciarli, T. (2021). Countries research priorities in relation to the Sustainable Development Goals. *Research Policy*, *53*(3). https://doi.org/10.1016/j.respol.2023.104950

Cornforth, C., Hayes, J. P., & Vangen, S. (2014). Nonprofit–Public Collaborations: Understanding Governance Dynamics. *Nonprofit and Voluntary Sector Quarterly*, *44*(4), 775–795. https://doi.org/10.1177/0899764014532836

Costas, R., & van Leeuwen, T. N. (2012). Approaching the "reward triangle": General analysis of the presence of funding acknowledgments and "peer interactive communication" in scientific publications. *Journal of the American Society for Information Science and Technology*, *63*(8), 1647–1661. https://doi.org/10.1002/asi.22692

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558–569. https://doi.org/10.1002/asi.1097

Cronin, B., & Weaver, S. (1995). The praxis of acknowledgement: from bibliometrics to influmetrics. *Revista Española de Documentación Científica*, *18*(2 SE-Estudios), 172–177. https://doi.org/10.3989/redc.1995.v18.i2.654

CWTS Leiden University. (2007). *Scoping study on the use of bibliometric analysis to measure the quality of research in UK higher education institutions*. https://webarchive.nationalarchives.gov.uk/ukgwa/20100303154704/http://www.hefce.ac.uk/pubs/rdreports/2007/rd18_07/

Czarnitzki, D., Grimpe, C., & Toole, A. A. (2015). Delay and secrecy: does industry sponsorship jeopardize disclosure of academic research? *Industrial and Corporate Change*, *24*(1), 251–279. https://doi.org/10.1093/icc/dtu011

Davis, P. (2012). *The Emergence of a Citation Cartel*. The Scholarly Kitchen.

de Solla Price, D. (1978). Editorial statements. *Scientometrics*, *1*(1), 3–8. https://doi.org/10.1007/BF02016836

De Souza, L. R. (2017). Correlates of child undernutrition in Yemen. *Bandung: Journal of the Global South*, *4*(1), 1–27. https://doi.org/10.1186/s40728-017-0040-y

Debackere, K., & Glänzel, W. (2004). Using a bibliometric approach to support research policy making: The case of the Flemish BOF-key. *Scientometrics*, *59*(2), 253–276. https://doi.org/10.1023/B:SCIE.0000018532.70146.02

DeBruin, R. E., Braam, R. R., & Moed, H. F. (1991). Bibliometric lines in the sand. *Nature*, *349*(6310), 559–562. https://doi.org/10.1038/349559a0

Desrochers, N., Paul-Hus, A., Haustein, S., Costas, R., Mongeon, P., Quan-Haase, A., Bowman, T. D., Pecoskie, J., Tsou, A., & Larivière, V. (2018). Authorship, citations, acknowledgments and visibility in social media: Symbolic capital in the multifaceted reward system of science. *Social Science Information*, *57*(2), 223–248. https://doi.org/10.1177/0539018417752089

Desrochers, N., Paul-Hus, A., & Pecoskie, J. (2017). Five decades of gratitude: A meta-synthesis of acknowledgments research. *Journal of the Association for Information Science and Technology*, *68*(12), 2821–2833. https://doi.org/10.1002/asi.23903

DG Research and Innovation. (2020). Strategic Plan 2020-2024. European Commission. https://commission.europa.eu/document/download/5ac1ff20-d41e-4c10-9a05-048b7339292e_en

Diem, A., & Wolter, S. C. (2013). The Use of Bibliometrics to Measure Research Performance in Education Sciences. *Research in Higher Education*, *54*(1), 86–114. https://doi.org/10.1007/s11162-012-9264-5

Digital Science. (2021). *Which publication dates does Dimensions use?* https://dimensions.freshdesk.com/support/solutions/articles/23000019982-which-publication-dates-does-dimensions-use

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education*, *49*(4), 495–533. https://doi.org/10.1007/s10734-004-1746-8

Donner, P., Rimmert, C., & van Eck, N. J. (2020). Comparing institutional-level bibliometric research performance indicator values based on different affiliation disambiguation systems. *Quantitative Science Studies*, *1*(1), 150–170. https://doi.org/10.1162/qss_a_00013

DORA. (2023). Declaration on Research Assessment. https://sfdora.org

Drahota, A. M. Y., Meza, R. D., Brikho, B., Naaf, M., Estabillo, J. A., Gomez, E. D., VEJNOSKA, S. F., DUFEK, S., STAHMER, A. C., & AARONS, G. A. (2016). Community-Academic Partnerships: A Systematic Review of the State of the Literature and Recommendations for Future Research. *The Milbank Quarterly*, *94*(1), 163–214. https://doi.org/10.1111/1468-0009.12184

Drott, M. C. (1995). Reexamining the role of conference papers in scholarly communication. *Journal of the American Society for Information Science*, *46*(4), 299–305. https://doi.org/10.1002/(SICI)1097-4571(199505)46:4<299::AID-ASI6>3.0.CO;2-0

Ducut, E., Liu, F., & Fontelo, P. (2008). An update on Uniform Resource Locator (URL) decay in MEDLINE abstracts and measures for its mitigation. *BMC Medical Informatics and Decision Making*, *8*. https://doi.org/10.1186/1472-6947-8-23

Dudek, J., Mongeon, P., & Bergmans, J. (2019). DataCite as a Potential Source for Open Data Indicators. *ISSI*, 2037–2042.

Duran-Silva, N., Fuster, E., Massucci, F. A., & Quinquillà, A. (2019). *A controlled vocabulary defining the semantic perimeter of Sustainable Development Goals*. https://doi.org/10.5281/ZENODO.4118028

Else, H. (2018). How I scraped data from Google Scholar. *Nature Research*. https://doi.org/10.1038/d41586-018-04190-5

Elsevier. (2015). *Scopus Conference Expansion Programme*. Scopus Conference Expansion Programme.

Elsevier, & SciDev.Net. (2015). *Sustainability science in a global landscape*. https://www.elsevier.com/__data/assets/pdf_file/0018/119061/SustainabilityScienceReport-Web.pdf

Espeland, W. N., & Sauder, M. (2016). Engines of anxiety: Academic rankings, reputation, and accountability. In *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. Russell Sage Foundation. https://www.scopus.com/record/display.uri?eid=2-s2.0-85011865409&origin=inward&

ETH Zürich. (2024). *Collaborate with industry*. ETH Zürich. https://ethz.ch/en/industry/researchers/collaborations.html

Etzkowitz, H., & Leydesdorff, L. (1995). The Triple Helix--University-industry-government relations: A laboratory for knowledge based economic development. *EASST Review*, *14*(1), 14–19. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2480085

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research Policy*, *29*(2), 109–123. https://doi.org/10.1016/S0048-7333(99)00055-4

European Commission. (2024). #RIDaysEU: European Research and Innovation Days (R&I Days) 2024 - European Commission. European Research and Innovation Days. https://cinea.ec.europa.eu/news-events/events/ridayseu-european-research-and-innovation-days-ri-days-2024-2024-03-20_en

European Commission. (2020). *CORDIS - EU Research Results*. About CORDIS. https://cordis.europa.eu/about/en

European Union. (2020). *Leading innovation through EU research*. Research and Innovation - Priorities and Actions.

Eurostat. (2024). *GERD by sector of performance*. https://ec.europa.eu/eurostat/databrowser/view/RD_E_GERDREG/default/table?lang=en

Falahati Qadimi Fumani, M. R., Goltaji, M., & Parto, P. (2013). Inconsistent transliteration of Iranian university names: a hazard to Iran's ranking in ISI Web of Science. *Scientometrics*, *95*(1), 371–384. https://doi.org/10.1007/s11192-012-0818-2

Fang, L. H., Lerner, J., & Wu, C. (2017). Intellectual Property Rights Protection, Ownership, and Innovation: Evidence from China. *The Review of Financial Studies*, *30*(7), 2446–2477. https://doi.org/10.1093/rfs/hhx023

Farazmand, A., Danaeefard, H., Mostafazadeh, M., & Sadeghi, M. R. (2019). Trends in Public Administration Research: A Content Analysis of Iranian Journal Articles (2004-2017). *International Journal of Public Administration*, *42*(10), 867–879. https://doi.org/10.1080/01900692.2019.1598689

Firchow, P., & Gellman, M. (2021). Collaborative Methodologies: Why, How, and for Whom? *PS: Political Science & Politics*, *54*(3), 525–529. https://doi.org/10.1017/S1049096521000330

Fister, I., Fister, I., & Perc, M. (2016). Toward the Discovery of Citation Cartels in Citation Networks. *Frontiers in Physics*, *4*. https://doi.org/10.3389/fphy.2016.00049

Flanders Make. (2024). *Our Research*. Research. https://www.flandersmake.be/en/offer/research

Ford, N., Mills, E. J., Zachariah, R., & Upshur, R. (2009). Ethics of conducting research in conflict settings. *Conflict and Health*, *3*(1), 7. https://doi.org/10.1186/1752-1505-3-7

Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2015). Errors in DOI indexing by bibliometric databases. *Scientometrics*, *102*(3), 2181–2186. https://doi.org/10.1007/s11192-014-1503-4

Frerichs, L., Kim, M., Dave, G., Cheney, A., Hassmiller Lich, K., Jones, J., Young, T. L., Cene, C. W., Varma, D. S., Schaal, J., Black, A., Striley, C. W., Vassar, S., Sullivan, G., Cottler, L. B., Brown, A., Burke, J. G., & Corbie-Smith, G. (2016). Stakeholder Perspectives on Creating and Maintaining Trust in Community–Academic Research Partnerships. *Health Education & Behavior*, *44*(1), 182–191. https://doi.org/10.1177/1090198116648291

Fuentes, M. A., Zelaya, D. G., & Madsen, J. W. (2020). Rethinking the Course Syllabus: Considerations for Promoting Equity, Diversity, and Inclusion. *Teaching of Psychology*, *48*(1), 69–79. https://doi.org/10.1177/0098628320959979

Gadd, E. (2020). University rankings need a rethink. *Nature*, *587*(7835), 523–524. link.gale.com/apps/doc/A650901626/HRCA?u=anon~34896c82&sid=googleScholar&xid=246d60e1

Garfield, E. (1955). Citation indexes for science. *Science*, *122*(3159), 108–111. https://doi.org/10.1126/science.122.3159.108

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*(4060), 471–479. https://doi.org/10.1126/science.178.4060.471

Garfield, E. (1983). How to Use Citation Analysis for Faculty Evaluations, and When Is It Relevant? Part 1. *Essays of an Information Scientist*, *6*, 354–362.

Garvey, W. D., Lin, N., Nelson, C. E., & Tomita, K. (1972a). Research studies in patterns of scientific communication: I. General description of research program. *Information Storage and Retrieval*, *8*(3), 111–122. https://doi.org/10.1016/0020-0271(72)90040-X

Garvey, W. D., Lin, N., Nelson, C. E., & Tomita, K. (1972b). Research studies in patterns of scientific communication: II. The role of the national meeting in scientific and technical communication. *Information Storage and Retrieval*, *8*(4), 159–169. https://doi.org/10.1016/0020-0271(72)90001-0

Gasparyan, A. Y., Yessirkepov, M., Voronov, A. A., Maksaev, A. A., & Kitas, G. D. (2021). Article-Level Metrics. *J Korean Med Sci*, *36*(11). https://doi.org/10.3346/jkms.2021.36.e74

Gellman, M. (2021). Collaborative Methodology with Indigenous Communities: A Framework for Addressing Power Inequalities. *PS: Political Science & Politics*, *54*(3), 535–538. https://doi.org/10.1017/S1049096521000299

Gerlak, A. K., House-Peters, L., Varady, R. G., Albrecht, T., Zúñiga-Terán, A., de Grenade, R. R., Cook, C., & Scott, C. A. (2018). Water security: A review of place-based research. *Environmental Science & Policy*, *82*, 79–89. https://doi.org/10.1016/j.envsci.2018.01.009

Geuna, A., & Martin, B. R. (2003). University Research Evaluation and Funding: An International Comparison. *Minerva*, *41*(4), 277–304. https://doi.org/10.1023/B:MINE.0000005155.70870.bd

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The New Production of Knowledge : The Dynamics of Science and Research in Contemporary Societies*. SAGE Publications Ltd. http://digital.casalini.it/9781446265871

Giller, K. E. (2020). Grounding the helicopters. *Geoderma*, *373*, 114302. https://doi.org/10.1016/j.geoderma.2020.114302

Giménez-Toledo, E., Mañana-Rodríguez, J., & Sivertsen, G. (2017). Scholarly book publishing: Its information sources for evaluation in the social sciences and humanities. *Research Evaluation*, *26*(2), 91–101. https://doi.org/10.1093/reseval/rvx007

Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics*, *68*(3), 457–473. https://doi.org/10.1007/s11192-006-0124-y

Gläser, J., Glänzel, W., & Scharnhorst, A. (2017). Same data—different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics*, *111*(2), 981–998. https://doi.org/10.1007/s11192-017-2296-z

Glied, S., Wittenberg, R., & Israeli, A. (2018). Research in government and academia: the case of health policy. *Israel Journal of Health Policy Research*, *7*(1), 35. https://doi.org/10.1186/s13584-018-0230-3

Goldfarb, B. (2008). The effect of government contracting on academic research: Does the source of funding affect scientific output? *Research Policy*, *37*(1), 41–58. https://doi.org/10.1016/j.respol.2007.07.011

González-Albo, B., & Bordons, M. (2011). Articles vs. proceedings papers: Do they differ in research relevance and impact? A case study in the Library and Information Science field. *Journal of Informetrics*, *5*(3), 369–381. https://doi.org/10.1016/j.joi.2011.01.011

Goodhart, C. A. E. (1975). Monetary Relationships: A View from Threadneedle Street. *Reserve Bank of Australia*, *1*.

Goodrum, A. A., McCain, K. W., Lawrence, S., & Giles, C. L. (2001). Scholarly publishing in the Internet age: a citation analysis of computer science literature. *INFORMATION PROCESSING & MANAGEMENT*, *37*(5), 661–675. https://doi.org/10.1016/S0306-4573(00)00047-9

Gorraiz, J., Melero-Fuentes, D., Gumpenberger, C., & Valderrama-Zurián, J. C. (2016). Availability of digital object identifiers (DOIs) in Web of Science and Scopus. *Journal of Informetrics*, *10*(1), 98–109. https://doi.org/10.1016/j.joi.2015.11.008

Greenhalgh, T., Jackson, C., Shaw, S., & Janamian, T. (2016). Achieving Research Impact Through Co-creation in Community-Based Health Services: Literature Review and Case Study. *The Milbank Quarterly*, *94*(2), 392–429. https://doi.org/10.1111/1468-0009.12197

Gross, P. L. K., & Gross, E. M. (1927). College Libraries and Chemical Education. *Science*, *66*(1713), 385–389. http://www.jstor.org/stable/1651803

Guerrero-Bote, V. P., Chinchilla-Rodríguez, Z., Mendoza, A., & de Moya-Anegón, F. (2021). Comparative Analysis of the Bibliographic Data Sources Dimensions and Scopus: An Approach at the Country and Institutional Levels. *Frontiers in Research Metrics and Analytics*, *5*, 19. https://doi.org/10.3389/frma.2020.593494

Haddon, C., & Sasse, T. (2019). *How academia can work with government*. https://apo.org.au/sites/default/files/resource-files/2019-04/apo-nid245496.pdf

Haile, M. (2020). Response to "Global soil science research collaboration in the 21st century: Time to end helicopter research." *Geoderma*, *373*, 114300. https://doi.org/10.1016/j.geoderma.2020.114300

Harzing, A.-W. (2019). Two new kids on the block: How do Crossref and Dimensions compare with Google Scholar, Microsoft Academic, Scopus and the Web of Science? *Scientometrics*, *120*(1), 341–349. https://doi.org/10.1007/s11192-019-03114-y

Hassan, M. (2001). Transition to sustainability in the twenty-first century: the contribution of science and technology - Report of the World Conference of Scientific Academies held in Tokyo, Japan, 15-18 May 2000. *International Journal of Sustainability in Higher Education*, *2*(1), 70–78. https://doi.org/10.1108/1467630110380523

Hazelkorn, E. (2011). Rankings and the reshaping of higher education: The battle for world-class excellence. In *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. Palgrave Macmillan. https://doi.org/10.1057/9780230306394

Hedt-Gauthier, B. L., Jeufack, H. M., Neufeld, N. H., Alem, A., Sauer, S., Odhiambo, J., Boum, Y., Shuchman, M., & Volmink, J. (2019). Stuck in the middle: a systematic review of authorship in collaborative health research in Africa, 2014-2016. *BMJ Global Health*, *4*(5). https://doi.org/10.1136/bmjgh-2019-001853

Heinz, E., Holt, K. E., Meehan, C. J., & Sheppard, S. K. (2021). Addressing parachute research and removing barriers for LMIC researchers in Microbial Genomics. *Microbial Genomics*, *7*(10). https://doi.org/10.1099/mgen.0.000722

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies*, *1*(1), 414–427. https://doi.org/10.1162/qss_a_00022

Hernandez-Aguilar, C., Dominguez-Pacheco, A., Martínez-Ortiz, E. J., Ivanov, R., López Bonilla, J. L., Cruz-Orea, A., & Ordonez-Miranda, J. (2020). Evolution and characteristics of the transdisciplinary perspective in research: a literature review. *Transdisciplinary Journal of Engineering & Science*, *11*, 158–188. https://doi.org/10.22545/2020/00140

Herzog, C., Hook, D., & Konkiel, S. (2020). Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies*, *1*(1), 387–395. https://doi.org/10.1162/qss_a_00020

Heymann, D. L., Liu, J., & Lillywhite, L. (2016). Partnerships, Not Parachutists, for Zika Research. *New England Journal of Medicine*, *374*(16), 1504–1505. https://doi.org/10.1056/NEJMp1602278

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *NATURE*, *520*(7548), 429–431. https://doi.org/10.1038/520429a

Hirsch Hadorn, G., Bradley, D., Pohl, C., Rist, S., & Wiesmann, U. (2006). Implications of transdisciplinarity for sustainability research. *Ecological Economics*, *60*(1), 119–128. https://doi.org/10.1016/j.ecolecon.2005.12.002

Hirsch, J. E. (2005). An index to quantify an individual{\textquoteright}s scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Hood, W. W., & Wilson, C. S. (2001). The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics*, *52*(2), 291–314. https://doi.org/10.1023/A:1017919924342

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

Horn, L., Alba, S., Blom, F., Faure, M., Flack-Davison, E., Gauri, G., & Ijsselmuiden, C. (2022). Fostering Research Integrity through the promotion of fairness, equity and diversity in research collaborations and contexts: Towards a Cape Town Statement (pre-conference discussion paper). *OSF PrePrints*. https://osf.io/bf286/

Huang, C., Cao, C., & Coreynen, W. (2023). Stronger and more just? Recent reforms of China's intellectual property rights system and their implications. *Asia Pacific Journal of Innovation and Entrepreneurship*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/APJIE-04-2023-0081

Huang, M., & Liu, W. (2019). Substantial numbers of easily identifiable illegal DOIs still exist in Scopus. *Journal of Informetrics*, *13*(3), 901–903. https://doi.org/10.1016/j.joi.2019.03.019

Huang, Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies*, *1*(2), 445–478. https://doi.org/10.1162/qss_a_00031

International Food Policy Research Institute. (2021). *IFPRI Country Programs*. https://www.ifpri.org/topic/ifpri-country-programs

International Science Council. (2015). *Review of Targets for the Sustainable Development Goals: The Science Perspective (2015)*. https://council.science/publications/review-of-targets-for-the-sustainable-development-goals-the-science-perspective-2015

Ioannidis, J. P. A., Boyack, K. W., & Baas, J. (2020). Updated science-wide author databases of standardized citation indicators. *PLOS Biology*, *18*(10), 1–3. https://doi.org/10.1371/journal.pbio.3000918

Ioannidis, J. P. A., & Maniadis, Z. (2023). In defense of quantitative metrics in researcher assessments. *PLOS Biology*, *21*(12), e3002408. https://doi.org/10.1371/journal.pbio.3002408

IUCN–UNEP–WWF. (1980). *World Conservation Strategy: Living Resource Conservation for Sustainable Development*. https://portals.iucn.org/library/efiles/documents/wcs-004.pdf

Ivanov, V. V, Markusova, V. A., & Mindeli, L. E. (2016). Government investments and the publishing activity of higher educational institutions: Bibliometric analysis. *Herald of the Russian Academy of Sciences*, *86*(4), 314–321. https://doi.org/10.1134/S1019331616040031

Jayabalasingham, B., Boverhof, R., Agnew, K., & Klein, L. (2019). Identifying research supporting the United Nations Sustainable Development Goals. *Mendeley Data*, *1*. https://doi.org/10.17632/87txkw7khs.1

Jetten, T. H., Veldhuizen, L. J. L., Siebert, M., Ommen Kloeke, A. E. E. van, & Darroch, P. I. (2019). *An explorative study on a university's outreach in the field of UN Sustainable Development Goal 2*. *June*, 0–22. https://doi.org/10.18174/476199

Jones, R. S. (2024). Improving Korea's Innovation System. *The Peninsula*. https://keia.org/the-peninsula/improving-koreas-innovation-system/

Juty, N., Wimalaratne, S. M., Soiland-Reyes, S., Kunze, J., Goble, C. A., & Clark, T. (2020). Unique, Persistent, Resolvable: Identifiers as the Foundation of FAIR. *Data Intelligence*, *2*(1–2), 30–39. https://doi.org/10.1162/dint_a_00025

Kahalon, R., Klein, V., Ksenofontov, I., Ullrich, J., & Wright, S. C. (2021). Mentioning the Sample's Country in the Article's Title Leads to Bias in Research Evaluation. *Social Psychological and Personality Science*, *13*(2), 352–361. https://doi.org/10.1177/19485506211024036

Kalleberg, A. L. (2009). Precarious work, insecure workers: Employment relations in transition. *American Sociological Review*, *74*(1), 1–22. https://doi.org/10.1177/000312240907400101

Kang, W., Zhao, S., Song, W., & Zhuang, T. (2019). Triple helix in the science and technology innovation centers of China from the perspective of mutual information: a comparative study between Beijing and Shanghai. *Scientometrics*, *118*(3), 921–940. https://doi.org/10.1007/s11192-019-03017-y

Kates, R. W. (2011). *From the Unity of Nature to Sustainability Science: Ideas and Practice." CID Working Paper* (218). https://dash.harvard.edu/bitstream/handle/1/37366242/218_Unityof Nature to Sustainabiltiy.pdf

Kates, R. W., Clark, W. C., Corell, R., Hall, J. M., Jaeger, C. C., Lowe, I., McCarthy, J. J., Schellnhuber, H. J., Bolin, B., Dickson, N. M., Faucheux, S., Gallopin, G. C., Grübler, A., Huntley, B., Jäger, J., Jodha, N. S., Kasperson, R. E., Mabogunje, A., Matson, P., … Svedin, U. (2001). Sustainability Science. *Science*, *292*(5517), 641–642. https://doi.org/10.1126/science.1059386

Khedmatgozar, H., & Alipour Hafezi, M. (2015). A Basic Comparative Framework for Evaluation of Digital Identifier Systems. *Journal of Digital Information Management*, *13*, 190–197.

Kim, M. (2016). The Relationship of Nonprofits' Financial Health to Program Outcomes: Empirical Evidence From Nonprofit Arts Organizations. *Nonprofit and Voluntary Sector Quarterly*, *46*(3), 525–548. https://doi.org/10.1177/0899764016662914

Knorr-Cetina, K. D. (2013). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Elsevier. https://books.google.ae/books?hl=en&lr=&id=DPwgBQAAQBAJ&oi

Konur, O. (2013). The scientometric evaluation of the institutional research: The Inner Anatolian Universities- Part 3. *Energy Education Science and Technology Part B: Social and Educational Studies*, *5*(2), 251–266.

Körfgen, A., Förster, K., Glatz, I., Maier, S., Becsi, B., Meyer, A., Kromp-Kolb, H., & Stötter, J. (2018). It's a Hit! Mapping Austrian Research Contributions to the Sustainable Development Goals. *Sustainability*, *10*(9). https://doi.org/10.3390/su10093295

Kousha, K., & Thelwall, M. (2024). Artificial intelligence to support publishing and peer review: A summary and review. *Learned Publishing*, *37*(1), 4–12. https://doi.org/https://doi.org/10.1002/leap.1570

Lai, B., & Thyne, C. (2007). The Effect of Civil War on Education, 1980—97. *Journal of Peace Research*, *44*(3), 277–292. https://doi.org/10.1177/0022343307076631

Lammey, R. (2020). Solutions for identification problems: a look at the Research Organization Registry. *Sci Ed*, *7*(1), 65–69. https://doi.org/10.6087/kcse.192

Larivière, V., & Sugimoto, C. R. (2019). The Journal Impact Factor: A Brief History, Critique, and Discussion of Adverse Effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 3–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_1

Lawani, S. M. (1980). *Quality, collaboration and citations in cancer research: A bibliometric study*. The Florida State University.

Leaning, J., & Guha-Sapir, D. (2013). Natural Disasters, Armed Conflict, and Public Health. *New England Journal of Medicine*, *369*(19), 1836–1842. https://doi.org/10.1056/NEJMra1109877

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology*, *64*(1), 2–17. https://doi.org/10.1002/asi.22784

Lee, & Nowell, B. (2014). A Framework for Assessing the Performance of Nonprofit Organizations. *American Journal of Evaluation*, *36*(3), 299–319. https://doi.org/10.1177/1098214014545828

Lee. (2014). University–Industry R&D Collaboration in Korea's National Innovation System. *Science, Technology and Society*, *19*(1), 1–25. https://doi.org/10.1177/0971721813514262

Lee, Y. S. (2000). The Sustainability of University-Industry Research Collaboration: An Empirical Assessment. *The Journal of Technology Transfer*, *25*(2), 111–133. https://doi.org/10.1023/A:1007895322042

LERU. (2024). *Next Generation Metrics for Scientific and Scholarly Research in Europe*. https://www.leru.org/files/Publications/Next-Generation-Metrics-for-Scientific-and-Scholarly-Research-in-Europe_Paper.pdf

Leydesdorff, L. (2003). The mutual information of university-industry-government relations: An indicator of the Triple Helix dynamics. *Scientometrics*, *58*(2), 445–467. https://doi.org/10.1023/A:1026253130577

Leydesdorff, L., & Etzkowitz, H. (1996). Emergence of a Triple Helix of university-industry-government relations. *Science and Public Policy*, *23*(5), 279 – 286. https://doi.org/10.1093/spp/23.5.279

Lisée, C., Larivière, V., & Archambault, É. (2008). Conference proceedings as a source of scientific information: A bibliometric analysis. *Journal of the American Society for Information Science and Technology*, *59*(11), 1779–1784. https://doi.org/10.1002/asi.20888

Liu, N. C., & Cheng, Y. (2005). The Academic Ranking of World Universities. *Higher Education in Europe*, *30*(2), 127–136. https://doi.org/10.1080/03797720500260116

Liu, W., Hu, G., & Tang, L. (2018). Missing author address information in Web of Science - An explorative study. *Journal of Informetrics*, *12*(3), 985–997. https://doi.org/10.1016/j.joi.2018.07.008

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Mammides, C., Goodale, U. M., Corlett, R. T., Chen, J., Bawa, K. S., Hariya, H., Jarrad, F., Primack, R. B., Ewing, H., Xia, X., & Goodale, E. (2016). Increasing geographic diversity in the international conservation literature: A stalled process? *Biological Conservation*, *198*, 78–83. https://doi.org/10.1016/j.biocon.2016.03.030

Manghi, P., Bardi, A., Atzori, C., Baglioni, M., Manola, N., Schirrwagen, J., Principe Contributors, P., Artini, M., Becker, A., De Bonis, M., Houssos, N., Iatropoulou, K., Lempesis, A., Loeden, A., Martziou, S., &

Rodrigues, E. (2019). *The OpenAIRE Research Graph Data Model*. https://openportal.isti.cnr.it/data/2019/416382/2019_%20416382.pdf

Manjarrés-Henríquez, L., Gutiérrez-Gracia, A., Carrión-García, A., & Vega-Jurado, J. (2009). The Effects of University–Industry Relationships and Academic Research On Scientific Performance: Synergy or Substitution? *Research in Higher Education*, *50*(8), 795–811. https://doi.org/10.1007/s11162-009-9142-y

Marginson, S., & van der Wende, M. (2007). To Rank or To Be Ranked: The Impact of Global Rankings in Higher Education. *Journal of Studies in International Education*, *11*(3–4), 306–329. https://doi.org/10.1177/1028315307303544

Marijan, D., & Sen, S. (2022). Industry–Academia Research Collaboration and Knowledge Co-Creation: Patterns and Anti-Patterns. *ACM Trans. Softw. Eng. Methodol.*, *31*(3). https://doi.org/10.1145/3494519

Marques, M., Powell, J. J. W., Zapp, M., & Biesta, G. (2017). How does research evaluation impact educational research? Exploring intended and unintended consequences of research assessment in the United Kingdom, 1986–2014. *European Educational Research Journal*, *16*(6), 820–842. https://doi.org/10.1177/1474904117730159

Martin, B., & Whitley, R. (2010). A case of regulatory capture?'. *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*, *51*. https://books.google.es/books?hl=en&lr=&id=PgBREAAAQBAJ&oi

Mattsson, P., Sundberg, C. J., & Laget, P. (2011). Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics*, *87*(1), 99–105. https://doi.org/10.1007/s11192-010-0310-9

Mauser, W., Klepper, G., Rice, M., Schmalzbauer, B. S., Hackmann, H., Leemans, R., & Moore, H. (2013). Transdisciplinary global change research: the co-creation of knowledge for sustainability. *Current Opinion in Environmental Sustainability*, *5*(3), 420–431. https://doi.org/10.1016/j.cosust.2013.07.001

Mayai, A. T. (2020). Security sector spending and public safety in South Sudan, 2006–2018. *African Security Review*, *29*(3), 1–17. https://doi.org/10.1080/10246029.2020.1813784

Mbaye, R., Gebeyehu, R., Hossmann, S., Mbarga, N., Bih-Neh, E., Eteki, L., Thelma, O. A., Oyerinde, A., Kiti, G., Mburu, Y., Haberer, J., Siedner, M., Okeke, I., & Boum, Y. (2019). Who is telling the story? A systematic review of authorship for infectious disease research conducted in Africa, 1980-2016. *BMJ Global Health*, *4*(5). https://doi.org/10.1136/bmjgh-2019-001855

Meo, S. A., Al Masri, A. A., Usmani, A. M., Memon, A. N., & Zaidi, S. Z. (2013). Impact of GDP, Spending on R&D, Number of Universities and Scientific Journals on Research Publications among Asian Countries. *PLOS ONE*, *8*(6), 1–8. https://doi.org/10.1371/journal.pone.0066449

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press. https://books.google.ae/books?hl=en&lr=&id=zPvcHuUMEMwC

Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis*, *79*(4), 606–623. https://www.journals.uchicago.edu/doi/abs/10.1086/354848

Minasny, B., & Fiantis, D. (2018). 'Helicopter research': who benefits from international studies in Indonesia? *The Conversation*.

Mirtl, M., T. Borer, E., Djukic, I., Forsius, M., Haubold, H., Hugo, W., Jourdan, J., Lindenmayer, D., McDowell, W. H., Muraoka, H., Orenstein, D. E., Pauw, J. C., Peterseil, J., Shibata, H., Wohner, C., Yu, X., & Haase, P. (2018). Genesis, goals and achievements of Long-Term Ecological Research at the global scale: A critical review of ILTER and future directions. *Science of The Total Environment*, *626*, 1439–1462. https://doi.org/10.1016/j.scitotenv.2017.12.001

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, *4*(3), 265–277. https://doi.org/10.1016/j.joi.2010.01.002

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, *110*(2), 967–990. https://doi.org/10.1007/s11192-016-2212-y

Moed, & Halevi, G. (2014). Tracking scientific development and collaborations – The case of 25 Asian countries. In David W. Chapman and Chiao-Ling Chien (Ed.), *Higher education in Asia: Expanding Out, Expanding Up* (pp. 79–92). UNESCO Institute for Statistics. https://unesdoc.unesco.org/ark:/48223/pf0000227516

Moed, Markusova, V., & Akoev, M. (2018). Trends in Russian research output indexed in Scopus and Web of Science. *Scientometrics*, *116*(2), 1153–1180. https://doi.org/10.1007/s11192-018-2769-8

Moed, H. F., & Van Leeuwen, Th. N. (1995). Improving the accuracy of institute for scientific information's journal impact factors. *Journal of the American Society for Information Science*, *46*(6), 461–467. https://doi.org/10.1002/(SICI)1097-4571(199507)46:6

Moed, & Visser, M. (2007). *Developing Bibliometric Indicators of Research Performance in Computer Science: An Exploratory Study*. https://www.cwts.nl/pdf/nwo_inf_final_report_v_210207.pdf

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Montesi, M., & Owen, J. M. (2008). From conference to journal publication: How conference papers in software engineering are extended for publication in journals. *Journal of the American Society for Information Science and Technology*, *59*(5), 816–829. https://doi.org/10.1002/asi.20805

Mooldijk, S. S., Licher, S., & Wolters, F. J. (2021). Characterizing Demographic, Racial, and Geographic Diversity in Dementia Research: A Systematic Review. *JAMA Neurology*, *78*(10), 1255–1261. https://doi.org/10.1001/jamaneurol.2021.2943

Morton, B., Vercueil, A., Masekela, R., Heinz, E., Reimer, L., Saleh, S., Kalinga, C., Seekles, M., Biccard, B., Chakaya, J., Abimbola, S., Obasi, A., & Oriyo, N. (2022). Consensus statement on measures to promote equitable authorship in the publication of research from international partnerships. *Anaesthesia*, *77*(3), 264–276. https://doi.org/10.1111/anae.15597

Mugnaini, R., Fraumann, G., Tuesta, E. F., & Packer, A. L. (2021). Openness trends in Brazilian citation data: factors related to the use of DOIs. *Scientometrics*, *126*(3), 2523–2556. https://doi.org/10.1007/s11192-020-03663-7

Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2013a). Absolute and specific measures of research group excellence. Scientometrics, 95(1), 115–127. https://doi.org/10.1007/s11192-012-0874-7

Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2013b). Comparison of a citation-based indicator and peer review for absolute and specific measures of research-group excellence. Scientometrics, 97(3), 767–777. https://doi.org/10.1007/s11192-013-1058-9

Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2015a). Predicting results of the Research Excellence Framework using departmental h-index. Scientometrics, 102(3), 2165–2180. https://doi.org/10.1007/s11192-014-1512-3

Mryglod, O., Kenna, R., Holovatch, Yu., & Berche, B. (2015b). Predicting results of the research excellence framework using departmental h-index: revisited. Scientometrics, 104(3), 1013–1017. https://doi.org/10.1007/s11192-015-1567-9

Nabout, J. C., Parreira, M. R., Teresa, F. B., Carneiro, F. M., da Cunha, H. F., de Souza Ondei, L., Caramori, S. S., & Soares, T. N. (2015). Publish (in a group) or perish (alone): the trend from single- to multi-authorship in biological papers. *Scientometrics*, *102*(1), 357–364. https://doi.org/10.1007/s11192-014-1385-5

Nakamura, M., Pendlebury, D., Schnell, J., & Szomszor, M. (2019). *Navigating the Structure of Research on Sustainable Development Goals*. https://clarivate.com/webofsciencegroup/campaigns/sustainable-development-goals/

Nalimov, V. V., & Mulchenko, Z. M. (1969). *Naukometriya. Izuchenie Razvitiya Nauki kak Informatsionnogo Protsessa. [Scientometrics. Study of the Development of Science as an Information Process]*. Nauka. https://books.google.es/books?hl=en&lr=&id=GPP9AgAAQBAJ

Narin, F., & others. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Citeseer.

National Academies. (2024). U.S. Science and Innovation Policy | National Academies. U.S. Science and Innovation Policy. https://www.nationalacademies.org/policy/us-science-and-innovation-policy

National Institutes of Health. (2012). NIH launches collaborative program with industry and researchers to spur therapeutic development. *News Releases*. https://www.nih.gov/news-events/news-releases/nih-launches-collaborative-program-industry-researchers-spur-therapeutic-development

National Institutes of Health. (2013). NIH to fund collaborations with industry to identify new uses for existing compounds. *News Releases*. https://www.nih.gov/news-events/news-releases/nih-fund-collaborations-industry-identify-new-uses-existing-compounds

National Science Board. (1973). *Science Indicators 1972*. https://apps.dtic.mil/sti/tr/pdf/ADA387123.pdf

Nature. (2021). Tracking 20 leading cities' Sustainable Development Goals research. *Nature*. https://doi.org/10.1038/d41586-021-02406-9

Nature. (2022). Nature addresses helicopter research and ethics dumping. *Nature*, *606*(7912), 7–7. https://doi.org/10.1038/d41586-022-01423-6

Nederhof, A. J., van Leeuwen, T. N., & van Raan, A. F. J. (2010). Highly cited non-journal publications in political science, economics and psychology: a first exploration. *Scientometrics*, *83*(2), 363–374. https://doi.org/10.1007/s11192-009-0086-y

New York University. (2024). University Research Policies and Guidelines. University Research Policies and Guidelines. https://www.nyu.edu/research/research-policies.html

Nguyen, T. V, & Pham, L. T. (2011). Scientific output and its relationship to knowledge economy: an analysis of ASEAN countries. *SCIENTOMETRICS*, *89*(1), 107–117. https://doi.org/10.1007/s11192-011-0446-2

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, *41*(1), 609–641. https://tefkos.comminfo.rutgers.edu/Courses/e530/Readings/Nicolaisen citation analysis ARIST 2008.pdf

North, M. A., Hastie, W. W., & Hoyer, L. (2020). Out of Africa: The underrepresentation of African authors in high-impact geoscience literature. *Earth-Science Reviews*, *208*, 103262. https://doi.org/10.1016/j.earscirev.2020.103262

OpenAlex. (2023). *OpenAlex: A free index for the world's research*. https://openalex.org

Orduna-Malea, E., Aytac, S., & Tran, C. Y. (2019). Universities through the eyes of bibliographic databases: a retroactive growth comparison of Google Scholar, Scopus and Web of Science. *Scientometrics*, *121*(1), 433–450. https://doi.org/10.1007/s11192-019-03208-7

Organisation for Economic Co-operation and Development. (2024). Science policy | OECD. Science Policy. https://www.oecd.org/en/topics/policy-issues/science-policy.html#

Otlet, P. (1934). *Traité de documentation: le livre sur le livre, theéorie et pratique*. Editiones mundaneum. https://books.google.es/books?id=NYeCiK3NP0IC

Paez, A. (2017). Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, *10*(3), 233–240. https://doi.org/10.1111/jebm.12266

Park, H. W., Hong, H. D., & Leydesdorff, L. (2005). A comparison of the knowledge-based innovation systems in the economies of South Korea and the Netherlands using Triple Helix indicators. *Scientometrics*, *65*(1), 3–27. https://doi.org/10.1007/s11192-005-0257-4

Perkmann, M., & Walsh, K. (2009). The two faces of collaboration: impacts of university-industry relations on public research. *Industrial and Corporate Change*, *18*(6), 1033–1065. https://doi.org/10.1093/icc/dtp015

Piwowar, H., Ferguson, L. M., Schrader, A. C., & Weisweiler, N. (2022). *Open Science Factsheet No. 9 based on the 64th online seminar: OpenAlex*.

PLOS. (2021). *Announcing a new PLOS policy on inclusion in global research*. The Official PLOS Blog. https://theplosblog.plos.org/2021/09/announcing-a-new-plos-policy-on-inclusion-in-global-research/

Polman, P., & Winston, A. S. (2022). *Net positive*. Profit Editorial.

Ponds, R., van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, *86*(3), 423 – 443. https://doi.org/10.1111/j.1435-5957.2007.00126.x

Pride, D., & Knoth, P. (2018). *Peer Review and Citation Data in Predicting University Rankings, a Large-Scale Analysis BT - Digital Libraries for Open Knowledge* (E. Méndez, F. Crestani, C. Ribeiro, G. David, & J. C. Lopes, Eds.; pp. 195–207). Springer International Publishing.

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv Preprint ArXiv:2205.01833*.

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, *25*, 348.

Provençal, S., Campbell, D., & Khayat, P. (2021). *Provision and analysis of key indicators in research and innovation. Policy brief J, Research trends on the Sustainable Development Goals (SDGs) and alignment with SDG 17 through international co-publications : focusing on SDGs 12–15 plus 6 (Planet)*. https://doi.org/10.2777/03227

Purnell, P. J. (2022a). *A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case Study of SDG 13: Climate Action*. Zenodo.

Purnell, P. J. (2022b). A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case Study of SDG 13: Climate Action. *Quantitative Science Studies*, 1–56. https://doi.org/10.1162/qss_a_00215

Purnell, P. J. (2023). *Geodiversity of hunger research: country focus and its regional and international actors*. Zenodo. https://doi.org/10.5281/zenodo.7593097

QS Intelligence Unit. (2019). *QS World University Rankings*. http://www.iu.qs.com/university-rankings/world-university-rankings

QSIU. (2019). *Papers & Citations*. QS Intelligence Unit. http://www.iu.qs.com/university-rankings/indicator-papers-citations/

Rafols, I., Noyons, E., Confraria, H., & Ciarli, T. (2021). Visualising plural mappings of science for Sustainable Development Goals (SDGs). *SocArxiv*. https://doi.org/10.31235/osf.io/yfqbd

Rahim, F., Khakimova, A., Ebrahimi, A., Zolotarev, O., & Rafiei Nasab, F. (2021). Global Scientific Research on SARS-CoV-2 Vaccines: A Bibliometric Analysis. *Cell Journal (Yakhteh)*, *23*(5), 523–531. https://doi.org/10.22074/cellj.2021.7794

Reyes-García, V., García-del-Amo, D., Benyei, P., Fernández-Llamazares, Á., Gravani, K., Junqueira, A. B., Labeyrie, V., Li, X., Matias, D. M. S., McAlvay, A., Mortyn, P. G., Porcuna-Ferrer, A., Schlingmann, A., & Soleymani-Fard, R. (2019). A collaborative approach to bring insights from local observations of climate change impacts into global climate change research. *Current Opinion in Environmental Sustainability*, *39*, 1–8. https://doi.org/10.1016/j.cosust.2019.04.007

Rey-Garcia, M., Liket, K., Alvarez-Gonzalez, L. I., & Maas, K. (2017). Back to Basics. *Nonprofit Management and Leadership*, *27*(4), 493–511. https://doi.org/10.1002/nml.21259

Riley, S., Schouten, W., & Cahill, S. (2003). Exploring the Dynamics of Subjectivity and Power Between Researcher and Researched. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *4*(2). https://doi.org/10.17169/fqs-4.2.713

Rinia, E. J., van Leeuwen, Th. N., van Vuren, H. G., & van Raan, A. F. J. (1998). Comparative analysis of a set of bibliometric indicators and central peer review criteria: Evaluation of condensed matter physics in the Netherlands. *Research Policy*, *27*(1), 95–107. https://doi.org/10.1016/S0048-7333(98)00026-2

Rinia, E. J., van Leeuwen, Th. N., van Vuren, H. G., & van Raan, A. F. J. (2001). Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics research. *Research Policy*, *30*(3), 357–361. https://doi.org/10.1016/S0048-7333(00)00082-2

Rivest, M., Kashnitsky, Y., Bédard-Vallée, A., Campbell, D., Khayat, P., Labrosse, I., Pinheiro, H., Provençal, S., Roberge, G., & James, C. (2021). *Improving the Scopus and Aurora queries to identify research that supports the United Nations Sustainable Development Goals (SDGs) 2021 Version 4*. Elsevier BV. https://doi.org/10.17632/9sxdykm8s4.4

Roach, M., & Sauermann, H. (2010). A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry. *Research Policy*, *39*(3), 422–434. https://doi.org/10.1016/j.respol.2010.01.004

Rubim, I. C., & Braganholo, V. (2017). Detecting referential inconsistencies in electronic CV datasets. *Journal of the Brazilian Computer Society*, *23*(1). https://doi.org/10.1186/s13173-017-0052-0

Rushforth, A. (2023). Letter: Response to Torres-Salinas et al. on "bibliometric denialism." *Scientometrics*, *128*(12), 6781–6784. https://doi.org/10.1007/s11192-023-04842-y

Rushforth, A., & Hammarfelt, B. (2023). The rise of responsible metrics as a professional reform movement: A collective action frames account. *Quantitative Science Studies*, *4*(4), 879–897. https://doi.org/10.1162/qss_a_00280

Sabiham, S., Setiawan, B. I., Minasny, B., & Fiantis, D. (2018). More research needed for responsible peatland management in Indonesia. *The Conversation*.

Salama, P., Spiegel, P., Talley, L., & Waldman, R. (2004). Lessons learned from complex emergencies over past decade. *The Lancet*, *364*(9447), 1801–1813. https://doi.org/10.1016/S0140-6736(04)17405-9

Savage, N. (2017). Industry links boost research output. *Nature Index*.

Schotten, M., Aisati, M. El, Meester, W. J. N., Steiginga, S., & Ross, C. A. (2017). A brief history of Scopus: The world's largest abstract and citation database of scientific literature. In *Research Analytics: Boosting University Productivity and Competitiveness through Scientometrics*. https://doi.org/10.1201/9781315155890

Sciensano. (2024). *Projects*. https://www.sciensano.be/en/projects

Seidl, R., Brand, F. S., Stauffacher, M., Krütli, P., Le, Q. B., Spörri, A., Meylan, G., Moser, C., González, M. B., & Scholz, R. W. (2013). Science with Society in the Anthropocene. *AMBIO*, *42*(1), 5–12. https://doi.org/10.1007/s13280-012-0363-5

Shanghai Ranking Consultancy. (2019). *Academic Ranking of World Universities Methodology*. http://www.shanghairanking.com/ARWU-Methodology-2019.html

Singh Chawla, D. (2018, December 31). Indonesia's scientists voice concerns about the country's researcher ranking system. *Chemical & Engineering News*. https://cen.acs.org/policy/publishing/Indonesias-scientists-voice-concerns-countrys/96/web/2018/12

Singh, P., Piryani, R., Singh, V. K., & Pinto, D. (2020). Revisiting subject classification in academic databases: A comparison of the classification accuracy of Web of Science, Scopus & Dimensions. *Journal of Intelligent & Fuzzy Systems*, *39*, 2471–2476. https://doi.org/10.3233/JIFS-179906

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J. (Paul), & Wang, K. (2015). An Overview of Microsoft Academic Service (MAS) and Applications. *Proceedings of the 24th International Conference on World Wide Web*, 243–246. https://doi.org/10.1145/2740908.2742839

SIRIS Academic. (2020). *Is EU-funded research and innovation evolving towards topics related to the Sustainable Development Goals?* Visual Essay. http://science4sdgs.sirisacademic.com

Small, H. (1993). Macro-level changes in the structure of co-citation clusters: 1983–1989. *Scientometrics*, *26*(1), 5–20. https://doi.org/10.1007/BF02016789

Small, H., & Sweeney, E. (1985). Clustering thescience citation index® using co-citations. *Scientometrics*, *7*(3), 391–409. https://doi.org/10.1007/BF02017157

Small, H., Sweeney, E., & Greenlee, E. (1985). Clustering the science citation index using co-citations. II. Mapping science. *Scientometrics*, *8*(5), 321–340. https://doi.org/10.1007/BF02018057

Spier, R. (2002). The history of the peer-review process. *Trends in Biotechnology*, *20*(8), 357–358. https://doi.org/10.1016/S0167-7799(02)01985-6

Stanford University. (2024). *Industry Collaborations*. Stanford University, Engineering. https://engineering.stanford.edu/get-involved/industry-collaborations

Sutterlüty, A., Šimunović, N., Hesser, F., Stern, T., Schober, A., & Schuster, K. C. (2018). Influence of the geographical scope on the research foci of sustainable forest management: Insights from a content analysis. *Forest Policy and Economics*, *90*, 142–150. https://doi.org/10.1016/j.forpol.2018.02.003

The British Academy. (2021). *Place-Based Approaches to Research Funding*. https://doi.org/10.5871/place/9780856726576.001

The Russell Group of Universities. (2023). *Russell Group welcomes Chancellor prioritising innovation to drive growth and productivity in Autumn Statement*. Russell Group.

The World Bank. (2022). *World Bank Open Data*. World Bank Open Data.

Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, *12*(2), 430–435. https://doi.org/10.1016/j.joi.2018.03.006

Thompson, M. A., Owen, S., Lindsay, J. M., Leonard, G. S., & Cronin, S. J. (2017). Scientist and stakeholder perspectives of transdisciplinary research: Early attitudes, expectations, and tensions. *Environmental Science & Policy*, *74*, 30–39. https://doi.org/10.1016/j.envsci.2017.04.006

Tijssen, R. J. W. (2012). Co-authored research publications and strategic analysis of public–private collaboration. *Research Evaluation*, *21*(3), 204–215. https://doi.org/10.1093/reseval/rvs013

Tijssen, R. J. W., van Leeuwen, T. N., & van Wijk, E. (2009). Benchmarking university-industry research cooperation worldwide: performance measurements and indicators based on co-authorship data for the world's largest universities. *Research Evaluation*, *18*(1), 13–24. https://doi.org/10.3152/095820209X393145

Tijssen, R. J. W., Yegros-Yegros, A., & Winnink, J. J. (2016). University–industry R&D linkage metrics: validity and applicability in world university rankings. *Scientometrics*, *109*(2), 677 – 696. https://doi.org/10.1007/s11192-016-2098-8

Times Higher Education. (2021). *Impact Rankings 2021: methodology*. https://www.timeshighereducation.com/world-university-rankings/impact-rankings-2021-methodology

Torres-Salinas, D., Arroyo-Machado, W., & Robinson-Garcia, N. (2023). Bibliometric denialism. *Scientometrics*, *128*(9), 5357–5359. https://doi.org/10.1007/s11192-023-04787-2

Traag, V. A., & Waltman, L. (2019). Systematic analysis of agreement between metrics and peer review in the UK REF. Palgrave Communications, 5(1), 29. https://doi.org/10.1057/s41599-019-0233-x

Tsinghua University. (2024). *Collaborating Institutions*. Research.

United Nations. (2014). *The Road to Dignity by 2030: Ending Poverty, Transforming All Lives and Protecting the Planet, Synthesis Report of the Secretary-General on the Post-2015 Agenda*. http://www.un.org/ga/search/view_doc.asp?symbol=A/69/700&Lang=E

United Nations. (2017). *Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development*. https://unstats.un.org/sdgs/indicators/Global Indicator Framework after 2021 refinement_Eng.pdf

United Nations General Assembly. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. Resolution Adopted by the General Assembly on 25 September 2015. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/291/89/PDF/N1529189.pdf?OpenElement

University of California Berkeley. (2024). Research Policies & Administration | Research UC Berkeley. Research Policies & Administration. https://vcresearch.berkeley.edu/how/ensure-compliance/research-policies-administration

University of Pretoria. (2023). Research Policy. https://www.up.ac.za/media/shared/1/ZP_Files/s5102_23-research-policy.zp233198.pdf

University of Technology Sydney. (2024). Research Policy | University of Technology Sydney. Research Policy. https://www.uts.edu.au/about/uts-governance/policies/uts-policy/research-policy

US News & World Report LP. (2019). *How U.S. News Calculated the Best Global Universities Rankings*. https://www.usnews.com/education/best-global-universities/articles/methodology

Valderrama-Zurián, J.-C., Aguilar-Moya, R., Melero-Fuentes, D., & Aleixandre-Benavent, R. (2015). A systematic analysis of duplicate records in Scopus. *Journal of Informetrics*, *9*(3), 570–576. https://doi.org/10.1016/j.joi.2015.05.002

van den Besselaar, P., Heyman, U., & Sandstrom, U. (2017). Perverse effects of output-based research funding? Butler's Australian case revisited. *JOURNAL OF INFORMETRICS*, *11*(3), 905–918. https://doi.org/10.1016/j.joi.2017.05.016

van den Besselaar, P., & Sandström, U. (2017). Counterintuitive effects of incentives? *Research Evaluation*, *26*(4), 349–351. https://doi.org/10.1093/reseval/rvx029

van Groenigen, J. W., & Stoof, C. R. (2020). Helicopter research in soil science: A discussion. *Geoderma*, *373*, 114418. https://doi.org/10.1016/j.geoderma.2020.114418

van Leeuwen, T., Engels, T. C. E., & Kulczycki, E. (2022). Using research metrics in support of assessing social sciences research performance: a comparison of major bibliographic systems. In T. C. E. Engels & E. Kulczycki (Eds.), *Handbook on Research Assessment in the Social Sciences* (pp. 148–161). Edward Elgar Publishing. https://doi.org/10.4337/9781800372559.00017

van Nes, F., Abma, T., Jonsson, H., & Deeg, D. (2010). Language differences in qualitative research: is meaning lost in translation? *European Journal of Ageing*, *7*(4), 313–316. https://doi.org/10.1007/s10433-010-0168-y

van Raan. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, *36*(3), 397–420. https://doi.org/10.1007/BF02129602

van Raan. (2021). Coevolution of Field and Institute: The Institutionalization of Bibliometric Research Illustrated by the Emergence and Flourishing of the CWTS. In *Handbook Bibliometrics* (pp. 53–64). https://doi.org/10.1515/9783110646610-007

van Raan. (1999). Advanced bibliometric methods for the evaluation of universities. *Scientometrics*, *45*(3), 417–423. https://doi.org/10.1007/BF02457601

Van Raan, A. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, *67*(3), 491–502. https://doi.org/10.1556/Scient.67.2006.3.10

van Raan. (2019). Measuring science: Basic principles and application of advanced bibliometrics. In *Springer Handbooks* (pp. 237–280). Springer. https://doi.org/10.1007/978-3-030-02511-3_10

Van Raan, A. F. J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, *62*(1), 133–143. https://doi.org/10.1007/s11192-005-0008-6

Vanderfeesten, M., & Otten, R. (2017). *Societal Relevant Impact : Potential analysis for Aurora-Network university leaders to strengthen collaboration on societal challenges*. https://doi.org/10.5281/ZENODO.1045839

Vanderfeesten, M., Spielberg, E., & Gunes, Y. (2020). *Survey data of "Mapping Research Output to the Sustainable Development Goals (SDGs)"* https://doi.org/10.5281/ZENODO.3813230

Vanecek, J., & Pecha, O. (2020). Fast growth of the number of proceedings papers in atypical fields in the Czech Republic is a likely consequence of the national performance-based research funding system. *RESEARCH EVALUATION*. https://doi.org/10.1093/reseval/rvaa005

Visser, M., van Eck, N., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

von Grebmer, K., Bernstein, J., Alders, R., Dar, O., Kock, R., Rampa, F., Wiemers, F., Acheampong, K., Hanano, A., Higgins, B., Ní Chéilleachair, R., Foley, C., Gitter, S., Ekstrom, K., & Fritschel, H. (2020). *2020 Global Hunger Index: One Decade to Zero Hunger: Linking Health and Sustainable Food Systems*. https://www.globalhungerindex.org

von Grebmer, K., Bernstein, J., Delgado, C., Smith, D., Wiemers, M., Schiffer, T., Hanano, A., Towey, O., Ní Chéilleachair, R., Foley, C., Gitter, S., Ekstrom, K., & Fritschel, H. (2021). *Global Hunger Index: Hunger and Food Systems in Conflict Settings*. https://www.concern.net/knowledge-hub/2021-global-hunger-index

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., van Leeuwen, T. N., van Raan, A. F. J., Visser, M. S., & Wouters, P. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, *63*(12), 2419–2432. https://doi.org/10.1002/asi.22708

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392. https://doi.org/10.1002/asi.22748

Waltman, L., van Eck, N. J., van Leeuwen, T., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: an empirical analysis. *Scientometrics*, *87*(3), 467–481. https://doi.org/10.1007/s11192-011-0354-5

Wang, Shen, Huang, Wu, Dong, & Kanakia. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, *1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

Ware, M. (2008). *Peer review: benefits, perceptions and alternatives*. Citeseer.

Wastl, J., Fane, B., Draux, H., & Diwersy, M. (2021). Keywords and search strings for generating SDG training sets. *Figshare*. https://doi.org/10.6084/m9.figshare.17294255.v1

Wastl, J., Hook, D. W., Fane, B., Draux, H., & Porter, S. J. (2020). *Contextualizing Sustainable Development Research*. https://doi.org/10.6084/m9.figshare.12200081

Westerheijden, D. F. (1997). A solid base for decisions. *Higher Education*, *33*(4), 397–413. https://doi.org/10.1023/A:1002995623272

Wiesmann, D., Biesalski, H., von Grebmer, K., & Bernstein, J. (2015). *Methodological Review and Revision of the Global Hunger Index*. https://doi.org/10.2139/ssrn.2673491

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. https://doi.org/10.13140/RG.2.1.4929.1363

WIPO. (2020). *China Becomes Top Filer of International Patents in 2019 Amid Robust Growth for WIPO's IP Services, Treaties and Finances*. Press Release. https://www.wipo.int/pressroom/en/articles/2020/article_0005.html

Wouters, P. (2017). Eugene Garfield (1925–2017). *Nature*, *543*(7646), 492. https://doi.org/10.1038/543492a

Wren, J. D., Kozak, K. Z., Johnson, K. R., Deakyne, S. J., Schilling, L. M., & Dellavalle, R. P. (2007). The write position. *EMBO Reports*, *8*(11), 988–991. https://doi.org/10.1038/sj.embor.7401095

Xu, S., Hao, L., An, X., Zhai, D., & Pang, H. (2019). Types of DOI errors of cited references in Web of Science with a cleaning method. *Scientometrics*, *120*(3), 1427–1437. https://doi.org/10.1007/s11192-019-03162-4

Yakubu, A., Tindana, P., Matimba, A., Littler, K., Munung, N. S., Madden, E., Staunton, C., & De Vries, J. (2018). Model framework for governance of genomic research and biobanking in Africa ? a content description [version 2; peer review: 3 approved]. *Open Research Africa*, *1*(13). https://doi.org/10.12688/aasopenres.12844.2

Yarime, M., Trencher, G., Mino, T., Scholz, R. W., Olsson, L., Ness, B., Frantzeskaki, N., & Rotmans, J. (2012). Establishing sustainability science in higher education institutions: towards an integration of academic development, institutionalization, and stakeholder collaborations. *Sustainability Science*, *7*(1), 101–113. https://doi.org/10.1007/s11625-012-0157-5

Yi, W., & Long, C. X. (2021). Does the Chinese version of Bayh-Dole Act promote university innovation? *China Economic Quarterly International*, *1*(3), 244–257. https://doi.org/https://doi.org/10.1016/j.ceqi.2021.09.003

Yom, S. (2014). From Methodology to Practice: Inductive Iteration in Comparative Research. *Comparative Political Studies*, *48*(5), 616–644. https://doi.org/10.1177/0010414014554685

Yozwiak, N. L., Happi, C. T., Grant, D. S., Schieffelin, J. S., Garry, R. F., Sabeti, P. C., & Andersen, K. G. (2016). Roots, Not Parachutes: Research Collaborations Combat Outbreaks. *Cell*, *166*(1), 5–8. https://doi.org/10.1016/j.cell.2016.06.029

Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, *68*(10), 2511–2521. https://doi.org/10.1002/asi.23883

Zhang, Patton, D., & Kenney, M. (2013). Building global-class universities: Assessing the impact of the 985 Project. *Research Policy*, *42*(3), 765–775. https://doi.org/10.1016/j.respol.2012.10.003

Zhang, & Glanzel, W. (2012). Proceeding papers in journals versus the regular journal publications. *JOURNAL OF INFORMETRICS*, *6*(1), 88–96. https://doi.org/10.1016/j.joi.2011.06.007

Zhang, Y., & Zou, Z. (Stephen). (2022). From academy to industry: China's new trend and policies on academic technology transfer. *IAM Life Sciences*. https://www.iam-media.com/global-guide/global-life-sciences/2022/article/academy-industry-chinas-new-trend-and-policies-academic-technology-transfer

Zhou, & Wang, M. (2023). The role of government-industry-academia partnership in business incubation: Evidence from new R&D institutions in China. *Technology in Society*, *72*, 102194. https://doi.org/10.1016/j.techsoc.2022.102194

Zhou, P., Tijssen, R., & Leydesdorff, L. (2016). University-Industry Collaboration in China and the USA: A Bibliometric Comparison. *PLOS ONE*, *11*(11), e0165277. https://doi.org/10.1371/journal.pone.0165277

Zhu, J., Hu, G., & Liu, W. (2019). DOI errors and possible solutions for Web of Science. *Scientometrics*, *118*(2), 709–718. https://doi.org/10.1007/s11192-018-2980-7

Zhu, Z., Cui, S., Wang, Y., & Zhu, Z. (2022). Exploration on the collaborative relationship between government, industry, and university from the perspective of collaborative innovation. *Applied Mathematics and Nonlinear Sciences*, *7*(2), 903–912. https://doi.org/doi:10.2478/amns.2021.2.00174

Zi, A., & Blind, K. (2015). Researchers' participation in standardisation: a case study from a public research institute in Germany. *The Journal of Technology Transfer*, *40*(2), 346–360. https://doi.org/10.1007/s10961-014-9370-y

Summary

Research stakeholders including policymakers and evaluators have a demonstrated need to identify research productivity, quality, impact, networks, trends, and other aspects of researcher activity across the full range of academic disciplines and at various level of organisational aggregation. Bibliometrics is the large-scale, quantitative analysis of data about scientific publications and is a relative quick, cheap, and easy method of producing high-level answers to many questions. The main objective of this PhD dissertation is to explore the extent to which bibliometrics can be a useful tool for research policymakers and to those conducting research evaluation. These chapters also seek to find the limitations to bibliometric methods and where they can even prove harmful through their effects on the scientific community. This dissertation suggests that optimal research evaluation should incorporate a healthy balance of peer review and bibliometric analysis.

Chapter 1 frames the dissertation by presenting the two main thrusts of research evaluation as either peer review (qualitative) or bibliometrics (quantitative). Both types of evaluation are explained, and it becomes clear that each contains elements of the other which means that expert opinion and bibliometrics are inter-related to a certain extent. The chapter then provides a history of bibliometrics and expands on bibliometric data sources and methods with examples of how the techniques have been used. The final part of chapter 1 explores some limitations of bibliometric methods and data sources and sets the scene by raising five important research questions that would help research policymakers use and interpret bibliometric reports. The research questions are on data quality, classifying new subject fields, inter-stakeholder collaboration, the geodiversity of research, and the dangers of bibliometric assessment influencing academic behaviour.

Chapter 2 addresses one of the most important questions in any quantitative analysis, data quality. An old adage says ‘rubbish in, rubbish out’ meaning that any analysis that uses low quality data will produce low quality results that cannot be relied upon. The main findings show that published research papers are not always assigned to an authors’ institution and are sometimes assigned to the wrong university. In bibliometric databases, papers are linked to authors, authors are linked to affiliations, and affiliations are mapped on to organisations. There are several steps where things can go wrong. The most common reasons are where the database indexes the papers with its authors but does not link the author affiliations. Another common mistake is to inaccurately unify the affiliation to mistakenly link the affiliation to the wrong university. The implication in both cases, is that an evaluation of the university output will be incomplete. Worse still, evaluators don’t know which universities are missing papers or how many they are missing, and which universities have been credited with ‘phantom’ papers. The concluding part of the chapter adds its voice to calls for universal adoption of a community-led unique identifier for academic institutions. The Research Organization Registry (ROR) is a good example and could prove more effective than each database using its own system.

Chapter 3 develops the discussion of data discrepancies between bibliometric databases to methodological discrepancies. The 17 UN sustainable development goals (SDGs) were introduced fairly recently, and various actors have developed methods of identifying research papers related to each of the goals. The main findings presented in this chapter show there is very little agreement between four such approaches to identifying research papers related to SDG 13: Climate action. The study then delves into the reasons behind the discrepancies and discovers that each of the methods compared comprise a multi-step process. The steps include employing experts to select important

keywords from relevant documents, then development and operationalisation of complex Boolean searches with bibliometric databases. These searches create reference sets that are then calibrated through a careful balance of precision and recall, and then further enhanced through machine learning algorithms. Most of the steps are described only superficially which means that no thorough analysis of any of the methods was possible. Analysis of keywords generated during the first step in the process shows there is broad disagreement between the methods already at this early stage. These differences are likely compounded at every step thereby resulting in the broad discrepancies presented in the findings of the study. The take-home message is that such methods are in their infancy and that studies of SDG related research output and impact are not yet robust and should not be relied upon. It is preferable to allow some flexibility and time for methods to develop further and perhaps diversify to the point that different methods could measure different aspects of the sustainability research.

Chapter 4 demonstrates the ability of bibliometric analysis to contribute large-scale findings as a form of monitoring progress of academic engagement with other societal actors. We hear regular calls from governments, funding bodies, policymakers, and higher education leaders for academia to close the gap with industry, to work more closely with government, and to engage with civic society. Chapter 4 links author affiliations to organisation types which allows a 10-year comparison of academic collaboration with industry, government bodies, and nonprofit organisations. The findings reveal that academia – industry collaboration is declining relative to the overall academic output, while both academia – government collaboration and academia – nonprofit collaboration are growing at roughly the same rate as overall academic output. The implication is that academia is not closing the gap with any of the societal stakeholders on a global scale. Country level analysis reveals some regional differences such as the positive effect of new intellectual property laws in China, and proactive company outreach in Belgium on national academia – industry collaboration. The findings are interpreted in the context of conflicting pressures experienced by different stakeholders. Industry often seeks to protect its intellectual property, while government and nonprofits are looking for immediately applicable results, and academia aims to incrementally build on prior knowledge through publication. It may be that collaboration is an avenue of interest rather than a priority for some or all of these research stakeholders.

Chapter 5 examines the geodiversity of research through the lens of a case study on research publications related to SDG 2: Zero hunger. Geodiversity is studied in terms of topic focus, author location, and authorship position. The study uses the country categories described in the Global Hunger Index (GHI). The findings are that both the topic focus and the geographical location of authors decline for the most severely affected country categories. These are counterintuitive results and may be explained by the relative under-development of research infrastructure in countries most affected by hunger. These countries are also often affected by poverty and civil conflict which could deflect resources towards non-academic activity. Analysis of author position confirmed the underrepresentation of academics from countries in the more severe GHI categories in all author positions, and especially in lead author positions reported elsewhere. The chapter avoids jumping to conclusions about parachute or helicopter research, and instead includes manual analyses of sample papers in search of explanations for lack of local authors in developing regions. The findings show that some authors are indeed regionally based, but in a neighbouring country rather than the one in focus. In other cases, local contributors were mentioned in the acknowledgements rather than being listed as co-authors. Follow-on studies of acknowledgements data would help us gain more knowledge about this practice.

Chapter 6 uses bibliometric analysis of published document types across ten countries in Southeast Asia and reveals a change in scholarly publication patterns specific to Indonesia and links it to national research policy. In 2012, the Indonesian government published new guidelines on academics' promotion and pay increases that were linked to the number of published papers indexed by international bibliometric databases. Shortly afterwards, the majority share of Indonesian papers shifted from journal articles to conference proceedings papers. Conference papers are generally considered quicker and easier to publish than journal articles. Correspondence with an Indonesian conference editor confirmed this was the driving factor among the local academic community shifting its attention to publishing conference papers. The implication is that using broad bibliometric methods to evaluate individuals can backfire if the community then sees the metric as a target. In this case, the ministry set out to increase high quality research output through national policy, but inadvertently caused an unintended behavioural change among the academic community. In performance review, especially of individual academics, a healthy dose of expert opinion is advised, and bibliometric indicators should play more of a supporting role.

Chapter 7 summarises the main findings and conclusions presented in chapters 2 – 6 and then expands further on the broader implications for the uses and limitations of bibliometrics in research evaluation. The implications of the findings section (7.2) refers back to the original objective of the dissertation which was to demonstrate how bibliometrics can be a useful tool for research policymakers and research evaluators. A secondary objective was to show the limitations of bibliometrics and where complementary methods such as peer review and expert judgement can add a layer of in-depth understanding of bibliometric study results on a smaller scale. The implications are grouped into three categories. The first concerns how bibliometrics is to some extent shaping the research system, the second addresses the need to find a fine balance between bibliometrics and expert judgement, and the third identifies areas in which the broader research community can contribute to improvements in bibliometric data, methods, and systems. Chapter 7 concludes with a vision for the future comprising five suggestions. The first is for the universal incorporation of additional persistent identifiers that will improve data quality and trust in the numbers. The second is increased transparency surrounding data curation especially the opening of 'black box' methods that include contributions of subject matter experts and machine learning algorithms. The third suggests an overhaul of the contributor reward system similar to the credits at the end of a cinema film. The fourth is concerned with accepting that bibliometrics is now partially shaping research behaviour and taking steps to ensure such influences are transparent and positive. The final part of the future vision addresses the incorporation of further applications of artificial intelligence in various aspects of bibliometrics.

Nederlandse samenvatting

Belanghebbenden in wetenschappelijk onderzoek, zoals beleidsmakers en instanties die evaluaties uitvoeren, hebben een vaak uitgesproken behoefte om productiviteit, kwaliteit, impact, netwerken, trends en andere aspecten van onderzoeksactiviteiten te identificeren in het hele spectrum van academische disciplines en op verschillende organisatorische niveaus. Bibliometrie is het vakgebied dat zich bezighoudt met de grootschalige, kwantitatieve analyse van gegevens over wetenschappelijke publicaties waarmee een snelle en goed toepasbare methode tegen beheersbare kosten geleverd kan worden om relevante antwoorden te geven op vragen betreffende bovengenoemde aspecten van wetenschappelijk onderzoek. Het hoofddoel van dit proefschrift is om te onderzoeken in hoeverre bibliometrie een doelmatig hulpmiddel kan zijn in toepassing ten behoeve van wetenschapsbeleid en onderzoeksevaluatie. De hoofdstukken in het proefschrift trachten ook de beperkingen van bibliometrische methoden te vinden, en of deze methoden zelfs nadelig kunnen zijn door hun effecten op de wetenschappelijke gemeenschap. Dit proefschrift geeft aan dat een optimale onderzoeksevaluatie een gezonde balans moet bevatten van peer review en bibliometrische analyse.

Hoofdstuk 1 geeft een inkadering van het proefschrift door de twee belangrijkste pijlers van onderzoeksevaluatie te presenteren, peer review (kwalitatief) en bibliometrie (kwantitatief). Beide typen evaluatie worden besproken en het wordt duidelijk gemaakt dat elk van deze typen elementen van de ander bevat, wat betekent dat peer review en bibliometrie tot op zekere hoogte met elkaar samenhangen. Het hoofdstuk presenteert vervolgens een korte geschiedenis van het vakgebied bibliometrie waarbij ingegaan wordt op de ontwikkeling van methoden en gegevensbronnen met voorbeelden hoe deze bibliometrische methoden zijn toegepast. Het laatste deel van hoofdstuk 1 richt zich op enkele beperkingen van bibliometrische methoden en gegevensbronnen en stelt vijf belangrijke onderzoeksvragen die beleidsmakers kunnen helpen om resultaten van bibliometrisch onderzoek te gebruiken en te interpreteren. De onderzoeksvragen gaan over de kwaliteit van gegevens, het classificeren van nieuwe onderzoeksgebieden, samenwerking tussen wetenschappelijke en maatschappelijke belanghebbenden, geodiversiteit van onderzoek, en de mogelijke nadelige gevolgen van bibliometrische evaluatiemethoden op het (publiceer)gedrag van onderzoekers.

Hoofdstuk 2 behandelt een van de belangrijkste problemen in elke kwantitatieve analyse, de kwaliteit van de gebruikte gegevens. Een oud gezegde luidt 'rubbish in, rubbish out', wat betekent dat elke analyse die gebruik maakt van data van lage kwaliteit, resultaten van lage kwaliteit zal opleveren waarop niet kan worden vertrouwd. De belangrijkste bevindingen laten zien dat wetenschappelijke publicaties niet altijd correct worden toegewezen aan de instelling van de auteur(s) en bijvoorbeeld worden toegewezen aan de 'verkeerde' universiteit. In bibliometrische databestanden worden publicaties gekoppeld aan auteurs, auteurs worden gekoppeld aan affiliaties en affiliaties worden, waar relevant, gekoppeld aan organisaties. Er zijn verschillende stappen waarin deze procedure fout kan gaan. De meest voorkomende fout treedt op wanneer de database de publicaties indexeert met de auteurs, maar geen koppeling aanbrengt met de affiliaties van de auteurs. Een andere veelgemaakte fout is het onnauwkeurig unificeren van een affiliatie zodat deze affiliatie gekoppeld wordt aan de verkeerde universiteit. De implicatie in beide gevallen is dat een evaluatie van de wetenschappelijke output van de universiteit onvolledig zal zijn. Erger nog, evaluatoren weten niet welke universiteiten publicaties missen of hoeveel ze missen, en welke universiteiten voorzien zijn van verkeerd toegewezen publicaties. Het afsluitende deel van het hoofdstuk levert een ondersteuning aan oproepen voor het algemeen aanvaarden van een door de wetenschappelijke gemeenschap ontwikkelde unieke identificatie van academische instellingen. Het Research Organization Registry (ROR) is een goed voorbeeld en zou effectiever kunnen blijken dan dat elke database zijn eigen systeem gebruikt.

Hoofdstuk 3 bespreekt de verschillen in gegevens tussen de diverse bibliometrische databases en de daarmee samenhangende methodologische discrepanties. De 17 doelen voor duurzame ontwikkeling (SDG's) van de VN zijn recentelijk geïntroduceerd en verschillende actoren hebben methoden ontwikkeld om onderzoekspublicaties te identificeren die verband houden met elk van de doelen. De belangrijkste bevindingen die in dit hoofdstuk worden gepresenteerd, laten zien dat er zeer weinig overeenstemming bestaat tussen vier van dergelijke benaderingen om onderzoekspublicaties te identificeren die verband houden met SDG 13: Klimaatactie. De studie richt zich op de redenen achter deze discrepanties. Uit de analyse blijkt dat elk van de vergeleken methoden een proces met meerdere stappen omvat. Deze stappen betreffen het inzetten van experts om belangrijke trefwoorden uit relevante publicaties en andere documenten te selecteren, en vervolgens het ontwikkelen en operationaliseren van complexe Booleaanse zoekopdrachten binnen bibliometrische databases. Deze zoekopdrachten leveren referentiesets die vervolgens op elkaar worden afgestemd door een zorgvuldige balans van precisie en 'recall', en daarna verder verbeterd worden met machine-learning algoritmen. De meeste stappen worden door de actoren slechts in hoofdlijnen beschreven, wat betekent dat er geen grondige analyse van de methoden mogelijk was. Analyse van trefwoorden die zijn gegenereerd tijdens de eerste stap in het proces, toont aan dat er al in dit vroege stadium er een groot verschil optreedt tussen de methoden. Deze verschillen worden waarschijnlijk bij elke stap groter, wat resulteert in de brede discrepanties die besproken worden in de bevindingen van het onderzoek. De lering die we hieruit kunnen trekken, is dat dergelijke methoden nog in de kinderschoenen staan en dat identificatie van SDG-gerelateerde onderzoeksresultaten en -impact nog niet robuust en dus ook nog niet voldoende betrouwbaar is. Het zal tijd kosten om deze identificatie-methoden verder te ontwikkelen en mogelijk te diversifiëren waarbij de verschillende methoden verschillende aspecten kunnen meten van onderzoek dat gerelateerd is aan de VN doelen voor duurzame ontwikkeling.

Hoofdstuk 4 laat de mogelijkheden van bibliometrische methoden zien om op grote schaal bevindingen te leveren bij de monitoring van de academische betrokkenheid bij maatschappelijke actoren. We horen regelmatig oproepen van overheden, financieringsinstellingen, beleidsmakers en leiders van het hoger onderwijs dat de academische wereld de kloof met de industrie moet dichten, nauwer moet samenwerken met de overheid en zich meer moet bezighouden met problemen in de samenleving. In dit hoofdstuk worden de affiliaties van auteurs gekoppeld aan organisatietypen, wat een analyse over een periode van 10 jaar van academische samenwerking met industrie, overheidsinstanties en non-profitorganisaties mogelijk maakt. De bevindingen laten zien dat de samenwerking tussen de academische wereld en de industrie afneemt ten opzichte van de totale academische output, terwijl zowel de samenwerking tussen de academische wereld en de overheid, als de samenwerking tussen de academische wereld en non-profitorganisaties ongeveer even snel groeit als de totale academische output. De implicatie is dat de academische wereld de kloof met geen enkele maatschappelijke stakeholder op wereldschaal dicht. Analyse op landenniveau laat enkele regionale verschillen zien, zoals het positieve effect van nieuwe wetgeving op het gebied van intellectueel eigendom in China en proactieve benadering van bedrijven in België met betrekking tot de nationale samenwerking tussen de academische wereld en de industrie. Onze bevindingen worden geïnterpreteerd in de context van de 'tegenstrijdige druk' die de verschillende stakeholders ervaren. De industrie probeert vaak haar intellectuele eigendom te beschermen, terwijl de overheid en non-profitorganisaties op zoek zijn naar direct toepasbare resultaten, en de academische wereld ernaar streeft om stapsgewijs voort te bouwen op eerdere kennis door middel van publicaties. Het kan zijn dat samenwerking meer een interessegebied is in plaats van een prioriteit voor sommige of mogelijk alle belanghebbenden in het onderzoek.

Hoofdstuk 5 richt zich op de geodiversiteit van onderzoek door middel van een case study over onderzoekspublicaties met betrekking tot SDG 2: Zero hunger. Geodiversiteit wordt geanalyseerd in termen van onderwerp, auteurslocatie en auteurspositie. De studie maakt gebruik van de landencategorieën die worden beschreven in de Global Hunger Index (GHI). Onze bevindingen tonen aan dat zowel onderwerp als geografische locatie van auteurs afnemen voor de zwaarst getroffen landencategorieën. Dit zijn contra-intuïtieve resultaten welke verklaard kunnen worden door de relatieve onderontwikkeling van de onderzoeksinfrastructuur in landen die het zwaarst worden getroffen door honger. Deze landen worden ook vaak getroffen door armoede en burgerconflicten, waardoor er voor academische activiteiten weinig mogelijkheden zijn. Analyse van de auteurspositie bevestigt de ondervertegenwoordiging van academici uit landen in de ernstigere GHI-categorieën in alle auteursposities, en met name de positie van eerste auteur. We vermijden het trekken van conclusies over het zogeheten parachute- of helikopteronderzoek, en in plaats daarvan presenteren we handmatige analyses van voorbeeld-publicaties op zoek naar verklaringen voor het gebrek aan lokale auteurs in ontwikkelingsregio's. Deze bevindingen laten zien dat sommige auteurs inderdaad regionaal gevestigd zijn, maar dan in een buurland in plaats van het land waarop de focus ligt. In andere gevallen worden lokale bijdragers genoemd in de dankbetuigingen (‘acknowledgements’) in plaats van dat ze worden vermeld als coauteurs. Vervolgstudies van gegevens in dankbetuigingen kunnen helpen meer kennis te vergaren over hoe in deze context met vermelding van auteurs en toekenning van auteursposities wordt omgegaan.

Hoofdstuk 6 focust op bibliometrische analyse van specifieke typen van wetenschappelijke publicaties (met name tijdschriftartikelen versus conferentiebijdragen) in tien landen in Zuidoost-Azië. We vinden een verandering in wetenschappelijke publicatiepatronen die specifiek zijn voor Indonesië en brengen dit in verbinding met het nationale onderzoeksbeleid. In 2012 publiceerde de Indonesische overheid nieuwe richtlijnen over bevorderingen en salarisverhogingen van academici gekoppeld aan het aantal gepubliceerde artikelen geïndexeerd door internationale bibliometrische databases zoals de Web of Science of Scopus. Wat er gebeurde is dat kort daarna het grootste deel van de Indonesische wetenschappelijke artikelen verschuift van tijdschriftartikelen naar conferentiebijdragen. Conferentiebijdragen worden over het algemeen beschouwd als sneller en gemakkelijker te publiceren dan tijdschriftartikelen. Correspondentie met een Indonesische congresredacteur bevestigde dat dit de drijvende factor was onder de lokale academische gemeenschap die haar aandacht verlegde naar het publiceren van conferentiebijdragen. De implicatie is dat het gebruik van bibliometrische methoden om individuen te evalueren averechts kan werken als de academische gemeenschap vervolgens de resultaten van het meten als een doel ziet. In dit geval wilde het ministerie de hoogwaardige onderzoeksoutput verhogen via nationaal beleid, maar dit veroorzaakte een onbedoelde gedragsverandering in de academische gemeenschap. Bij prestatiebeoordelingen, met name van individuele academici, wordt een deskundige mening geadviseerd en moeten bibliometrische indicatoren een ondersteunende rol spelen.

Hoofdstuk 7 vat de belangrijkste bevindingen en conclusies samen die in hoofdstukken 2-6 zijn gepresenteerd en gaat vervolgens verder in op de bredere implicaties voor het gebruik en de beperkingen van bibliometrische methoden bij de evaluatie van onderzoek. Deze implicaties (paragraaf 7.2) verwijzen terug naar het belangrijkste doel van het proefschrift, namelijk om aan te tonen hoe het vakgebied bibliometrie nuttige hulpmiddelen kan leveren voor beleidsmakers en evaluatoren van onderzoek. Een secundair doel was om de beperkingen van bibliometrie aan te geven en waar aanvullende methoden zoals peer review en andere vormen van deskundig oordeel een beter begrip van de bibliometrische onderzoeksresultaten kunnen leveren. De implicaties zijn gegroepeerd in drie categorieën. De eerste betreft hoe bibliometrische methoden tot op zekere hoogte het onderzoekssysteem vormgeven, de tweede gaat over de noodzaak om een goede balans te vinden tussen bibliometrie en deskundig oordeel, en de derde identificeert gebieden waarop de bredere onderzoeksgemeenschap kan bijdragen aan verbeteringen in bibliometrische gegevens, methoden en classificatiesystemen. Hoofdstuk 7 sluit af met een visie voor de toekomst bestaande uit vijf suggesties. De eerste betreft universele standaardisatie en opname in wetenschappelijke bestanden van identificatiegegevens die de datakwaliteit en daarmee het vertrouwen in de data zullen verbeteren. De tweede betreft meer transparantie met betrekking tot bewerking van data, met name opening van 'black box'-methoden die bijdragen van experts uit specifieke vakgebieden en algoritmen gebruikt in machine learning omvatten. De derde gaat over een herziening van het beloningssysteem voor personen die bijdragen aan onderzoek, vergelijkbaar met de credits aan het einde van een bioscoopfilm. De vierde betreft het accepteren dat bibliometrie nu gedeeltelijk het onderzoeksgedrag vormgeeft en het nemen van stappen om ervoor te zorgen dat dergelijke invloeden transparant en positief zijn. Het laatste deel van de toekomstvisie richt zich op de opname van verdere toepassingen van kunstmatige intelligentie in verschillende aspecten van bibliometrie.

Supplementary Information

Supplementary material from

Chapter 2: The prevalence and impact of university affiliation discrepancies between four bibliographic databases – Scopus, Web of Science, Dimensions, and Microsoft Academic

Table S2.1. Data table from Figure 2.2. Scopus – Web of Science affiliation discrepancy by university

	Scopus surplus	Scopus records with affiliation discrepancy in WoS	Overlap	WoS records with affiliation discrepancy in Scopus	Web of Science surplus
Ain Shams	2,506	92	5,187	1,077	357
Alexandria	1,956	90	4,829	106	342
Assiut	1,185	56	3,367	345	218
AUB	1,263	39	3,062	1,080	278
Babylon	891	4	374	3	12
Baghdad	1,288	97	1,302	16	61
Bahrain	313	11	487	4	15
Carthage	1,398	82	3,115	757	110
Jordan	1,197	42	2,486	50	63
Khalifa	1,869	80	3,832	46	58
K. Abdulaziz	3,345	185	18,507	445	334
K. Saud	3,931	478	17,210	620	415
Kuwait	872	65	2,267	40	99

Lebanese	1,113	284	1,482	39	69
Qatar	2,067	45	3,968	45	169
Sfax	2,353	39	4,947	1,036	309
S. Qaboos	1,320	58	2,348	192	110
UAEU	1,193	84	2,422	23	95
Totals	**30,060**	**1,831**	**81,192**	**5,924**	**3,114**

Table S2.2. Data table from Figure 2.3. Scopus – Dimensions affiliation discrepancy by university

	Scopus surplus	Scopus records with affiliation discrepancy in Dimensions	Overlap	Dimensions records with affiliation discrepancy in Scopus	Dimensions surplus
Ain Shams	182	1,361	6,228	1,565	1,933
Alexandria	125	1,199	5,544	250	1,423
Assiut	77	921	3,605	75	755
AUB	96	836	3,425	153	501
Babylon	254	505	506	2	110
Baghdad	126	1,016	1,540	61	329
Bahrain	38	160	612	27	83
Carthage	86	1,052	3,453	424	216
Jordan	93	866	2,759	157	356
Khalifa	151	997	4,628	59	304
K. Abdulaziz	377	4,113	17,518	274	1,286
K. Saud	380	3,877	17,339	844	1,283
Kuwait	62	545	2,596	68	337

Lebanese	54	419	2,406	192	334
Qatar	100	884	5,094	178	572
Sfax	103	1,551	5,679	686	423
S. Qaboos	127	787	2,813	21	651
UAEU	57	688	2,951	32	417
Totals	**2,488**	**21,777**	**88,696**	**5,068**	**11,313**

Table S2.3. Data table from Figure 2.4. Scopus – Microsoft Academic affiliation discrepancy by university

	Scopus surplus	Scopus records with affiliation discrepancy in Microsoft Academic	Overlap	Microsoft Academic records with affiliation discrepancy in Scopus	Microsoft Academic surplus
Ain Shams	240	1,388	6,140	1,232	3,450
Alexandria	183	1,347	5,338	49	1,811
Assiut	110	1,084	3,411	34	1,093
AUB	154	552	3,656	1,094	855
Babylon	283	497	485	3	213
Baghdad	260	915	1,510	14	690
Bahrain	31	209	569	11	142
Carthage	107	1,620	2,862	132	260
Jordan	131	713	2,871	116	764
Khalifa	159	3,276	2,336	989	411
K. Abdulaziz	571	4,836	16,604	233	2,123
K. Saud	629	3,676	17,299	163	1,702
Kuwait	116	513	2,573	40	426

Lebanese	66	441	2,369	71	443
Qatar	186	1,015	4,871	48	1,061
Sfax	221	1,850	5,257	85	712
S. Qaboos	107	466	3,151	357	955
UAEU	109	770	2,817	26	483
Totals	**3,663**	**25,168**	**84,119**	**4,697**	**17,594**

Table S2.4. Data table from Figure 2.5. Web of Science – Dimensions affiliation discrepancy by university

	WoS surplus	WoS records with affiliation discrepancy in Dimensions	Overlap	Dimensions records with affiliation discrepancy in WoS	Dimensions surplus
Ain Shams	43	1,088	5,487	244	4,010
Alexandria	49	842	4,383	215	2,620
Assiut	19	975	2,935	47	1,457
AUB	22	1,630	2,767	100	1,213
Babylon	6	72	311	2	305
Baghdad	11	264	1,103	58	769
Bahrain	6	71	428	31	263
Carthage	53	1,291	2,638	102	1,363
Jordan	39	416	2,141	114	1,023
Khalifa	30	593	3,313	53	1,625
K. Abdulaziz	182	3,338	15,760	125	3,195
K. Saud	166	3,150	14,924	815	3,729
Kuwait	23	331	2,052	72	877

Lebanese	17	176	1,397	367	1,168
Qatar	34	503	3,645	161	2,082
Sfax	53	1,783	4,458	44	2,291
S. Qaboos	25	586	2,038	35	1,412
UAEU	15	383	2,141	55	1,197
Totals	**793**	**17,492**	**71,921**	**2,640**	**30,599**

Table S2.5. Data table from Figure 2.6. Web of Science – Microsoft Academic affiliation discrepancy by university

	WoS surplus	WoS records with affiliation discrepancy in Microsoft Academic	Overlap	Microsoft Academic records with affiliation discrepancy in WoS	Microsoft Academic surplus
Ain Shams	91	1,106	5,418	65	5,351
Alexandria	86	1,012	4,174	46	2,978
Assiut	50	1,082	2,797	26	1,715
AUB	89	638	3,693	61	1,866
Babylon	11	93	285	2	414
Baghdad	37	340	1,002	19	1,193
Bahrain	5	101	400	14	308
Carthage	63	1,787	2,130	42	1,085
Jordan	50	344	2,202	61	1,495
Khalifa	43	2,370	1,520	774	1,450
K. Abdulaziz	290	4,077	14,912	105	3,538
K. Saud	356	2,968	14,913	164	4,086
Kuwait	36	348	2,022	51	966

Lebanese	23	232	1,334	247	1,302
Qatar	56	747	3,378	26	2,576
Sfax	94	2,427	3,767	22	2,266
S. Qaboos	35	302	2,311	49	2,110
UAEU	40	410	2,089	55	1,183
Totals	**1,455**	**20,384**	**68,347**	**1,829**	**35,882**

Table S2.6. Data table from Figure 2.7. Dimensions – Microsoft Academic affiliation discrepancy by university

	Dimensions surplus	Dimensions records with affiliation discrepancy in Microsoft Academic	Overlap	Microsoft Academic records with affiliation discrepancy in Dimensions	Microsoft Academic surplus
Ain Shams	179	1,377	8,185	2,228	418
Alexandria	113	1,096	6,007	1,018	171
Assiut	55	851	3,530	904	101
AUB	120	329	3,631	1,858	120
Babylon	5	148	465	201	35
Baghdad	37	396	1,497	587	130
Bahrain	263	31	428	71	6
Carthage	46	1,339	2,715	453	87
Jordan	48	398	2,830	828	99
Khalifa	76	2,624	2,289	1,343	109
K. Abdulaziz	294	2,817	15,961	2,606	392
K. Saud	301	2,257	16,910	1,896	357
Kuwait	87	344	2,569	391	80

Lebanese	40	408	2,484	315	85
Qatar	111	871	4,901	969	108
Sfax	77	1,854	4,858	1,029	164
S. Qaboos	74	193	3,218	1,098	153
UAEU	81	517	2,793	464	65
Totals	**2,007**	**17,850**	**85,271**	**18,259**	**2,680**

Supplementary material from

Chapter 5: Geodiversity of Research: Geographical Topic Focus, Author Location, and Collaboration. A Case Study of SDG 2: Zero Hunger

Table S5.1. Countries designated or provisionally designated by category in the GHI 2021

Extremely Alarming

Somalia

Alarming

Burundi	Democratic Republic of the Congo	South Sudan
Comoros	Chad	Syrian Arab Republic
Madagascar	Central African Republic	Yemen

Serious

Afghanistan	India	Nigeria
Angola	Iraq	Pakistan
Benin	Kenya	Papua New Guinea
Botswana	North Korea	Rwanda
Burkina Faso	Lesotho	Sierra Leone
Congo (Republic of)	Liberia	Sudan
Côte d'Ivoire	Malawi	Tanzania
Djibouti	Mali	Timor-Leste
Ethiopia	Mauritania	Togo
Guinea	Mozambique	Uganda
Guinea-Bissau	Namibia	Venezuela
Haiti	Niger	Zambia

Zimbabwe

Moderate

Bangladesh
Bolivia
Cabo Verde
Cambodia
Cameroon
Ecuador
Egypt
Eswatini
Gabon
Gambia
Ghana
Guatemala
Guyana
Honduras
Indonesia
Lao PDR
Malaysia
Mauritius
Myanmar
Nepal
Nicaragua
Oman
Philippines
Senegal
Solomon Islands
South Africa
Sri Lanka
Suriname
Thailand
Viet Nam

Low

Argentina
Albania
Algeria
Armenia
Azerbaijan
Bulgaria
Colombia
Costa Rica
Dominican Republic
El Salvador
Fiji
Georgia
Iran
Jamaica
Jordan
Kazakhstan
Kyrgyzstan
Lebanon
Mexico
Mongolia
Morocco
Panama
Paraguay
Peru
Russian Federation
Saudi Arabia
Trinidad & Tobago
Tunisia
Turkmenistan
Ukraine
Uzbekistan

Countries assessed but insufficient data for categorisation

Bahrain

Bhutan

Equatorial Guinea

Eritrea

Libya

Maldives

Qatar

Countries not assessed by GHI

Andorra

Antigua and Barbuda Australia

Austria

Bahamas

Barbados

Belgium

Belize

Bermuda

Brunei

Canada

Cook Islands

Cyprus

Czech Republic

Denmark

Dominica

Finland

France

Germany

Greece

Greenland

Grenada

Hong Kong

Hungary

Iceland

Ireland

Israel

Italy

Japan

Kiribati

Liechtenstein

Luxembourg

Macao

Malta

Marshall Islands

Micronesia

Monaco

Nauru

Netherlands

New Zealand

Niue

Norway

Palau

Palestine

Poland

Portugal

Puerto Rico

South Korea

Saint Kitts and Nevis

Saint Lucia

Saint Vincent and the Grenadines

Saint-Barthélemy

Samoa

San Marino

Sao Tome and Principe

Seychelles

Singapore

Slovenia

Spain

Sweden

Switzerland

Tonga

Tuvalu

United Arab Emirates

United Kingdom

United States

Vanuatu

Vatican

Western Sahara

Table S5.2. Reasons for Mentions-only publications.

	Alarming / Extremely alarming	Serious	Moderate	Low	Not assessed
International researchers focusing on a different region	13	15	17	12	3
Researchers affiliated to neighbouring country	3	2	3	0	1
Passing mention of country but no topical focus	4	3	0	8	16

Sample data is made available in Zenodo (Purnell, 2023)

Table S5.3. Reasons for Affiliations-only publications.

	Alarming / Extremely alarming	Serious	Moderate	Low	Not assessed
Study focuses on other region, or has no regional focus	13	17	13	9	17
Local region is mentioned but not the country name	4	3	5	4	2
Local study but region not mentioned	3	0	2	7	1

Sample data is made available in Zenodo (Purnell, 2023)

Acknowledgements – A Leiden Walk

In 2018, I arranged for a cohort of university leaders from the Middle East to travel to Leiden and undertake a customised bibliometrics course at CWTS. Following the workshop, Ton van Raan led the group on a Leiden Walk – a historical tour of the city centre during which Ton pointed out interesting facts about science that are in some way related to Leiden. Later that evening, I found myself in a conversation with Ton, who had inspired me during the walk with the story of a unique publication type that was inscribed on a tombstone in Pieterskerk (St Peter's Church) over four hundred years ago. Ton knew about my interest in bibliometrics and suggested that I conduct a research-based PhD at CWTS. I quickly accepted the offer and Ton said he would co-supervise me with Ludo Waltman, an upcoming star in the field. Over the course of the PhD, Ton has continued to inspire me with his boundless energy and enthusiasm for new projects, and his limitless repertoire of fascinating stories about Leiden, science, and the study of science. Ton always takes the big picture approach that has helped me interpret my own studies the broader context of research evaluation.

I already knew Ludo from previous visits to CWTS and had been influenced by his work while preparing my own workshops. In one of our first PhD reviews, he told me there was a problem with my writing style. As a native English speaker, I had perhaps somewhat arrogantly assumed my style would easily make the grade and so I was rather taken aback. Ludo has generously helped me develop as a researcher using his unique knack of delivering direct and honest feedback with disarming charm, which throughout my PhD journey has made me feel both grateful and enlightened. Ludo insists on high standards and his astonishing attention to detail has meant that only my very best efforts are good enough.

To conduct a PhD in quantitative bibliometrics at CWTS, I had to first take steps on the path from an amateur to a professional bibliometrician. An important difference is in the sourcing and treatment of the data. CWTS hosts the underlying data for many bibliometric databases, and I was granted access to the system on the condition I would teach myself the SQL necessary to run my own queries. This was the first time I had been exposed to programming and initially it was a daunting hurdle. After the first study, I got the hang of it and improved my skills with each successive paper. By the end of the PhD, I have taken a major step towards becoming a professional bibliometrician.

Throughout this process, the CWTS team has been a constant pleasure to work with and has contributed to my enchantment with the city of Leiden. I have particularly enjoyed interactions with Paul Wouters, Ed Noyons, Martijn Visser, Robert Tijssen, Sarah de Rijcke, Mark Neijssel, Nees Jan van Eck, Vincent Traag, Thed van Leeuwen, Ismail Rafols, and Rodrigo Costas. The support team has been very friendly and super-helpful, particularly Patrick Kooij, Anita Nieuwenhuizen, Petra van der Weel, and Inge van der Weijden.

The final acknowledgement is reserved for my ever-loving wife, Mia without whose support, sacrifice, and patience, none of the above would have been possible. Mia has shared my enthusiasm and made it possible for me to spend a very substantial amount of our evenings, weekends, and holidays working on my studies over the last six years. Whenever I travelled for my studies, Mia has held the fort alone while we brought up our family. She has also endured countless hours of discussions about bibliometrics and provided the understanding and encouragement I needed to remain motivated and progress through my PhD journey.

Curriculum Vitae

Philip James Purnell was born in Brighton, United Kingdom in 1970 and graduated with A levels from the Ecclesbourne School in 1989. He was awarded a bachelor's degree in human biology from the University of Wolverhampton in 1993 and a master's degree in Neuroscience from the Institute of Psychiatry, University of London in 1994. He then joined the pharmaceutical industry holding positions in both AstraZeneca and UCB in the UK before emigrating to Spain in 2000.

Philip worked as International Project Manager for Prous Science, a Spanish scientific publishing company that was incorporated into Thomson Reuters in 2007. Philip spent five years developing the academic market for products and services related to quantitative assessment of research output, impact, and performance. In 2013, he moved to the Dubai office of Thomson Reuters and led the Thought Leadership & Customer Education unit covering the Middle East, Africa, Russia, and former Soviet countries. In 2015, Philip joined Knowledge E as a partner setting up and leading the research services division that comprised a training business, a publishing house, and a consultancy.

Philip was the ORCID Ambassador for the Middle East & Africa from 2015-18, a consultant for Times Higher Education from 2015-16, and served on the QS regional advisory board from 2020-22, and the International Advisory Committee of the Moscow Engineering Physics Institute (MEPhI) from 2014-18. He is an external PhD student under the supervision of Prof. Ludo Waltman and Prof. Ton van Raan at CWTS, Leiden University. Since 2022, Philip has been Head of institutional research and benchmarking at the Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) in Abu Dhabi and continues to play an active role in the bibliometrics community.

List of Publications

Journal articles

Purnell, Philip J. (2024). Geodiversity of research: geographical topic focus, author location, and collaboration. A case study of SDG 2: zero hunger. Scientometrics, 129(5), 2701–2727. https://doi.org/10.1007/s11192-024-04994-5

Purnell, Philip J. (2022). A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case Study of SDG 13: Climate Action. Quantitative Science Studies, 1–56. https://doi.org/10.1162/qss_a_00215

Purnell, Philip J. (2022). The prevalence and impact of university affiliation discrepancies between four bibliographic databases—Scopus, Web of Science, Dimensions, and Microsoft Academic. Quantitative Science Studies, 3(1), 99–121. https://doi.org/10.1162/qss_a_00175

Purnell, Philip J. (2021). Conference proceedings publications in bibliographic databases: a case study of countries in Southeast Asia. Scientometrics, 126(1), 355–387. https://doi.org/10.1007/s11192-020-03773-2

Gorraiz, J., Gumpenberger, C., & **Purnell, P. J.** (2014). The power of book reviews: A simple and transparent enhancement approach for book citation indexes. Scientometrics, 98(2). https://doi.org/10.1007/s11192-013-1176-4

Gorraiz, J., **Purnell, P. J.**, & Glänzel, W. (2013). Opportunities for and limitations of the Book Citation Index. Journal of the American Society for Information Science and Technology, 64(7). https://doi.org/10.1002/asi.22875

Purnell, Philip J., & Quevedo-Blasco, R. (2013). Benefits to the Spanish research community of regional content expansion in Web of Science. International Journal of Clinical and Health Psychology, 13(2), 147–154. https://doi.org/10.1016/S1697-2600(13)70018-8

Finn, P. E., **Purnell, P.**, & Pilkington, G. J. (1996). Effect of histamine and the H2 antagonist cimetidine on the growth and migration of human neoplastic glia. Neuropathology and applied neurobiology, 22(4), 317–324. https://doi.org/10.1111/j.1365-2990.1996.tb01110.x

Conference proceedings papers

Shehata, S., Santandreu Calonge, D., **Purnell, P.**, & Thompson, M. (2023). Enhancing Video-based Learning Using Knowledge Tracing: Personalizing Students' Learning Experience with ORBITS. In E. Kochmar, J. Burstein, A. Horbach, R. Laarmann-Quante, N. Madnani, A. Tack, … T. Zesch (Eds.), Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023) (pp. 100–107). https://doi.org/10.18653/v1/2023.bea-1.8

Purnell, Philip J. (2015). The Role of the Modern Library in Research Evaluation. QScience Proceedings of The Special Libraries Association-Arabian Gulf Chapter 21st Annual Conference. Mar 2015, Volume 2015, 5 https://www.qscience.com/content/papers/10.5339/qproc.2015.gsla.5

Purnell, P.J. (2014). Research output and performance in Central Asia a bibliometric study. 8th IEEE International Conference on Application of Information and Communication Technologies, AICT 2014 - Conference Proceedings. https://doi.org/10.1109/ICAICT.2014.7035995

Purnell, Philip J. (2014). Hajj and the Middle East Respiratory Syndrome (MERS) Virus. Zawya.

Datasets

Purnell, Philip J. (2023). Geodiversity of hunger research: country focus and its regional and international actors. https://doi.org/10.5281/zenodo.7593097